3Com Switch 4800G Family

Configuration Guide

Switch 4800G 24-Port

Switch 4800G 48-Port

Switch 4800G PWR 24-Port

Switch 4800G PWR 48-Port

Switch 4800G 24-Port SFP

Product Version:
Release 2202
Manual Version:
6W100-20090120
www.3com.com

3Com Corporation
350 Campus Drive, Marlborough,
MA, USA 01752 3064

ENVIRONMENTAL STATEMENT

It is the policy of 3Com Corporation to be environmentally-friendly in all operations. To uphold our policy, we are committed to:

Establishing environmental performance standards that comply with national legislation and regulations.

Conserving energy, materials and natural resources in all operations.

Reducing the waste generated by all operations. Ensuring that all waste conforms to recognized environmental standards. Maximizing the recyclable and reusable content of all products.

Ensuring that all products can be recycled, reused and disposed of safely.

Ensuring that all products are labelled according to recognized environmental standards.

Improving our environmental record on a continual basis.

End of Life Statement

3Com processes allow for the recovery, reclamation and safe disposal of all end-of-life electronic components.

Regulated Materials Statement

3Com products do not contain any hazardous or ozone-depleting material.

Environmental Statement about the Documentation

The documentation for this product is printed on paper that comes from sustainable, managed forests; it is fully biodegradable and recyclable, and is completely chlorine-free. The varnish is environmentally-friendly, and the inks are vegetable-based with a low heavy-metal content.

About This Manual

Organization

3Com Switch 4800G Family Configuration Guide is organized as follows:

Volume	Features			
00-Product Overview	Product Overview	Acronyms		
01-Access Volume	Ethernet Interface	Link Aggregation	Port Isolation	Service Loopback Group
	DLDP	LLDP	Smart Link	Monitor Link
	VLAN	GVRP	QinQ	BPDU Tunneling
	VLAN Mapping	Ethernet OAM	Connectivity Fault Detection	MSTP
	RRPP	Port Mirroring		
02-IP Services Volume	IP Addressing	ARP	DHCP	DNS
	IP Performance Optimization	UDP Helper	URPF	IPv6 Basics
	Dual Stack	Tunneling	sFlow	
03-IP Routing Volume	IP Routing Overview	Static Routing	RIP	OSPF
	IS-IS	BGP	IPv6 Static Routing	RIPng
	OSPFv3	IPv6 IS-IS	IPv6 BGP	Route Policy
	BFD	MCE		
04-Multicast Volume	Mulitcast Overview	Multicast Routing and Forwarding	IGMP	PIM
	MSDP	MBGP	IGMP Snooping	Multicast VLAN
	IPv6 Multicast Routing and Forwarding	MLD	IPv6 PIM	IPv6 MBGP
	MLD Snooping	IPv6 Multicast VLAN		
05-QoS Volume	QoS	User Profile		
06-Security Volume	AAA	802.1X	HABP	MAC Authentication
	Portal	Port Security	IP Source Guard	SSH2.0
	PKI	SSL	Public Key	ACL

Volume	Features			
07-System Volume	Login	Basic System Configuration	Device Management	File System Management
	HTTP	SNMP	RMON	MAC Address Table Management
	System Maintaining and Debugging	Information Center	PoE	Track
	NQA	NTP	VRRP	Hotfix
	Cluster Management	IRF Stack	GR Overview	Automatic Configuration
	IPC			

Conventions

The manual uses the following conventions:

Command conventions

Convention	Description
Boldface	The keywords of a command line are in **Boldface**.
italic	Command arguments are in *italic*.
[ ]	Items (keywords or arguments) in square brackets [ ] are optional.
{ x | y | ... }	Alternative items are grouped in braces and separated by vertical bars. One is selected.
[x | y | ...]	Optional alternative items are grouped in square brackets and separated by vertical bars. One or none is selected.
{ x | y | ... } *	Alternative items are grouped in braces and separated by vertical bars. A minimum of one or a maximum of all can be selected.
[x | y | ...] *	Optional alternative items are grouped in square brackets and separated by vertical bars. Many or none can be selected.
&<1-n>	The argument(s) before the ampersand (&) sign can be entered 1 to n times.
#	A line starting with the # sign is comments.

GUI conventions

Convention	Description
< >	Button names are inside angle brackets. For example, click <OK>.
[ ]	Window names, menu items, data table and field names are inside square brackets. For example, pop up the [New User] window.
/	Multi-level menus are separated by forward slashes. For example, [File/Create/Folder].

Symbols

Convention	Description
 **Warning**	Means reader be extremely careful. Improper operation may cause bodily injury.
 **Caution**	Means reader be careful. Improper operation may cause data loss or damage to equipment.
 **Note**	Means a complementary description.

Related Documentation

In addition to this manual, each 3com Switch 4800G documentation set includes the following:

Manual	Description
3Com Switch 4800G Family Command Reference Guide	Provide detailed descriptions of command line interface (CLI) commands, that you require to manage your switch.
3Com Switch 4800G Family Getting Started Guide	This guide provides all the information you need to install and use the 3Com Switch 4800G Family.

Obtaining Documentation

You can access the most up-to-date 3Com product documentation on the World Wide Web at this URL: http://www.3com.com.

Table of Contents

1 Product Features

Introduction to Product

The 3Com Switches 4800G are Gigabit Ethernet switching products and have abundant service features. They are designed as distribution and access devices for intranets and metropolitan area networks (MANs). They can also be used for connecting server groups in data centers.

The 3Com Switches 4800G support the innovative Intelligent Resilient Framework (IRF) technology. With IRF, multiple 4800G switches can be interconnected as a logical entity to form a new intelligent network featuring high availability, scalability, and manageability.

Feature Lists

The Switch 4800G supports abundant features and the related documents are divided into the volumes as listed in Table 1-1.

Table 1-1 Feature list

Volume	Features			
01-Access Volume	Ethernet Interface	Link Aggregation	Port Isolation	Service Loopback Group
	DLDP	LLDP	Smart Link	Monitor Link
	VLAN	GVRP	QinQ	BPDU Tunneling
	VLAN Mapping	Ethernet OAM	Connectivity Fault Detection	MSTP
	RRPP	Port Mirroring		
02-IP Services Volume	IP Addressing	ARP	DHCP	DNS
	IP Performance Optimization	UDP Helper	URPF	IPv6 Basics
	Dual Stack	Tunneling	sFlow	
03-IP Routing Volume	IP Routing Overview	Static Routing	RIP	OSPF
	IS-IS	BGP	IPv6 Static Routing	RIPng
	OSPFv3	IPv6 IS-IS	IPv6 BGP	Route Policy
	BFD	MCE		

Volume	Features			
04-Multicast Volume	Mulitcast Overview	Multicast Routing and Forwarding	IGMP	PIM
	MSDP	MBGP	IGMP Snooping	Multicast VLAN
	IPv6 Multicast Routing and Forwarding	MLD	IPv6 PIM	IPv6 MBGP
	MLD Snooping	IPv6 Multicast VLAN		
05-QoS Volume	QoS	User Profile		
06-Security Volume	AAA	802.1X	HABP	MAC Authentication
	Portal	Port Security	IP Source Guard	SSH2.0
	PKI	SSL	Public Key	ACL
07-System Volume	Login	Basic System Configuration	Device Management	File System Management
	HTTP	SNMP	RMON	MAC Address Table Management
	System Maintaining and Debugging	Information Center	PoE	Track
	NQA	NTP	VRRP	Hotfix
	Cluster Management	IRF Stack	GR Overview	Automatic Configuration
	IPC			

2 Features

The following sections provide an overview of the main features of each module supported by the Switch 4800G.

Access Volume

Table 2-1 Features in Access volume

Features	Description
Ethernet Interface	This document describes: • Basic Ethernet Interface Configuration • Combo Port Configuration • Configuring Flow Control on an Ethernet Interface • Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Interface • Configuring Loopback Testing on an Ethernet Interface • Configuring a Port Group • Configuring Storm Suppression • Setting the Interval for Collecting Ethernet Interface Statistics • Enabling Forwarding of Jumbo Frames • Enabling Loopback Detection on an Ethernet Interface • Configuring the MDI Mode for an Ethernet Interface • Testing the Cable on an Ethernet Interface • Configuring the Storm Constrain Function on an Ethernet Interface
Link Aggregation	Link aggregation aggregates multiple physical Ethernet ports into one logical link. This document describes: • Basic Concepts of Link Aggregation • Configuring an Aggregation Group • Configuring an Aggregate Interface • Configuring a Load Sharing Mode for Load-Sharing Link Aggregation Groups
Port Isolation	The port isolation feature allows you to isolate different ports within the same VLAN. This document describes: • Introduction to Port Isolation • Configuring the Isolation Group
Service Loopback Group	To increase service redirecting throughput, you can bundle multiple service loopback ports into a logical link, called a service loopback group. This document describes: • Introduction to Service Loopback Groups • Configuring a Service Loopback Group

Features	Description
DLDP	In the use of fibers, link errors, namely unidirectional links, are likely to occur. DLDP is designed to detect such errors. This document describes: • DLDP Introduction • Enabling DLDP • Setting DLDP Mode • Setting the Interval for Sending Advertisement Packets • Setting the DelayDown Timer • Setting the Port Shutdown Mode • Configuring DLDP Authentication • Resetting DLDP State
LLDP	LLDP enables a device to maintain and manage its own and its immediate neighbor's device information, based on which the network management system detects and determines the conditions of the communications links. This document describes: • Introduction to LLDP • Performing Basic LLDP Configuration • Configuring the Encapsulation Format for LLDPDUs • Configuring the Encapsulation Format of the Management Address • Configuring CDP Compatibility • Configuring LLDP Trapping
Smart Link	Smart Link is a solution for active-standby link redundancy backup and rapid transition in dual-uplink networking. This document describes: • Smart Link Overview • Configuring a Smart Link Device • Configuring an Associated Device
Monitor Link	Monitor link is a port collaboration function used to enable a device to be aware of the up/down state change of the ports on an indirectly connected link. This document describes: • Monitor Link Overview • Configuring Monitor Link
VLAN	Using the VLAN technology, you can partition a LAN into multiple logical LANs. This document describes: • Introduction to VLAN • Types of VLAN • Introduction and Configuration of Isolate-user-vlan • Introduction and Configuration of Voice VLAN
GVRP	GVRP is a GARP application. This document describes: • GARP overview • GVRP configuration • GARP Timers configuration
QinQ	As defined in IEEE802.1Q, 12 bits are used to identify a VLAN ID, so a device can support a maximum of 4094 VLANs. The QinQ feature extends the VLAN space by allowing Ethernet frames to travel across the service provider network with double VLAN tags. This document describes: • Introduction to QinQ • Configuring basic QinQ • Configuring Selective QinQ • Configuring the TPID Value in VLAN Tags

Features	Description
BPDU Tunneling	BPDU tunneling enables transparently transmission of customer network BPDU frames over the service provider network. This document describes: • Introduction to BPDU Tunneling • Configuring BPDU Transparent Transmission • Configuring Destination Multicast MAC Address for BPDU Tunnel Frames
VLAN Mapping	The VLAN mapping feature maps CVLAN tags to SVLAN tags. This document describes: • Configuring One-to-One VLAN Mapping • Configuring Many-to-One VLAN Mapping • Configuring Two-to-Two VLAN Mapping
Ethernet OAM	Ethernet OAM is a tool monitoring Layer-2 link status. It helps network administrators manage their networks effectively. This document describes: • Ethernet OAM overview • Configuring Basic Ethernet OAM Functions • Configuring Link Monitoring • Enabling OAM Loopback Testing
Connectivity Fault Detection	Connectivity fault detection is an end-to-end, per-VLAN link-layer OAM mechanism for link connectivity detection, fault verification, and fault location. This document describes: • Connectivity Fault Detection Overview • Basic Configuration Tasks • Configuring CC on MEPs • Configuring LB on MEPs • Configuring LT on MEPs
MSTP	MSTP is used to eliminate loops in a LAN. It is compatible with STP and RSTP. This document describes: • Introduction to MSTP • Configuring the Root Bridge • Configuring Leaf Nodes • Performing mCheck • Configuring Digest Snooping • Configuring No Agreement Check • Configuring Protection Functions
RRPP	RRPP is a link layer protocol designed for Ethernet rings. RRPP can prevent broadcast storms caused by data loops when an Ethernet ring is healthy, and rapidly restore the communication paths between the nodes after a link is disconnected on the ring. This document describes: • RRPP overview • Configuring Master Node • Configuring Transit Node • Configuring Edge Node • Configuring Assistant Edge Node • Configuring Ring Group

Features	Description
Port Mirroring	Port mirroring copies packets passing through a port to another port connected with a monitoring device for packet analysis to help implement network monitoring and troubleshooting. This document describes: • Port Mirroring overview • Local port mirroring configuration • Remote port mirroring configuration

IP Services Volume

Table 2-2 Features in the IP Services volume

Features	Description
IP Addressing	An IP address is a 32-bit address allocated to a network interface on a device that is attached to the Internet. This document describes: • Introduction to IP addresses • IP address configuration
ARP	Address Resolution Protocol (ARP) is used to resolve an IP address into a data link layer address. This document describes: • ARP Overview • Configuring ARP • Configuring Gratuitous ARP • Proxy ARP and Local Proxy ARP configuration • ARP Attack Defense configuration
DHCP	DHCP is built on a client-server model, in which the client sends a configuration request and then the server returns a reply to send configuration parameters such as an IP address to the client. This document describes: • DHCP overview • DHCP server configuration • DHCP relay agent configuration • DHCP Client configuration • DHCP Snooping configuration • BOOTP Client configuration
DNS	Used in the TCP/IP application, Domain Name System (DNS) is a distributed database which provides the translation between domain name and the IP address. This document describes: • Introduction to DNS • Configuring the DNS Client • Configuring the DNS Proxy
IP Performance Optimization	In some network environments, you need to adjust the IP parameters to achieve best network performance. This document describes: • IP performance overview • Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network • Configuring TCP Attributes • Configuring ICMP to Send Error Packets
UDP Helper	UDP Helper functions as a relay agent that converts UDP broadcast packets into unicast packets and forwards them to a specified server. This document describes: • UDP Helper overview • UDP Helper configuration

Features	Description
URPF	Unicast Reverse Path Forwarding (URPF) protects a network against source address spoofing attacks. This document describes: • URPF overview • URPF configuration
IPv6 Basics	Internet protocol version 6 (IPv6), also called IP next generation (IPng), was designed by the Internet Engineering Task Force (IETF) as the successor to Internet protocol version 4 (IPv4). This document describes: • IPv6 overview • Basic IPv6 functions configuration • IPv6 NDP configuration • PMTU discovery configuration • IPv6 TCP properties configuration • ICMPv6 packet sending configuration • IPv6 DNS Client configuration
Dual Stack	A network node that supports both IPv4 and IPv6 is called a dual stack node. A dual stack node configured with an IPv4 address and an IPv6 address can have both IPv4 and IPv6 packets transmitted. This document describes: • Dual stack overview • Dual stack configuration
Tunneling	Tunneling is an encapsulation technique, which utilizes one network transport protocol to encapsulate packets of another network transport protocol and transfer them over the network. This document describes: • Tunneling overview • IPv6 manually tunnel configuration • 6to4 tunnel configuration • ISATAP tunnel configuration
sFlow	Based on packet sampling, Sampled Flow (sFlow) is a traffic monitoring technology mainly used to collect and analyze traffic statistics. This document describes: • sFlow Overview • sFlow Configuration

IP Routing Volume

Table 2-3 Features in the IP Routing volume

Features	Description
IP Routing Overview	This document describes: • Introduction to IP routing and routing table • Routing protocol overview
Static Routing	A static route is manually configured by the administrator. The proper configuration and usage of static routes can improve network performance and ensure bandwidth for important network applications. This document describes: • Static route configuration • Detecting Reachability of the Static Route’s Nexthop

Features	Description
RIP	Routing Information Protocol (RIP) is a simple Interior Gateway Protocol (IGP), mainly used in small-sized networks. This document describes: • RIP basic functions configuration • RIP advanced functions configuration • RIP network optimization configuration
OSPF	Open Shortest Path First (OSPF) is an Interior Gateway Protocol based on the link state developed by IETF. This document describes: • Enabling OSPF • Configuring OSPF Areas • Configuring OSPF Network Types • Configuring OSPF Route Control • Configuring OSPF Sham Link • Configuring OSPF Network Optimization • Configuring OSPF Graceful Restart
IS-IS	Intermediate System-to-Intermediate System (IS-IS) is a link state protocol, which uses the shortest path first (SPF) algorithm. This document describes: • Configuring IS-IS Basic Functions • Configuring IS-IS Routing Information Control • Tuning and Optimizing IS-IS Networks • Configuring IS-IS Authentication • Configuring System ID to Host Name Mappings • Configuring IS-IS GR • Enabling the Logging of Neighbor State Changes • Enabling IS-IS SNMP Trap
BGP	Border gateway protocol (BGP) is an inter-autonomous system (inter-AS) dynamic route discovery protocol. This document describes: • Configuring BGP Basic Functions • Controlling Route Generation • Controlling Route Distribution and Reception • Configuring BGP Route Attributes • Tuning and Optimizing BGP Networks • Configuring a Large Scale BGP Network • Configuring BGP GR • Enabling Trap • Enabling Logging of Peer State Changes
IPv6 Static Routing	Static routes are special routes that are manually configured by network administrators. Similar to IPv4 static routes, IPv6 static routes work well in simple IPv6 network environments. This document describes: • IPv6 static route configuration
RIPng	RIP next generation (RIPng) is an extension of RIP-2 for IPv4. RIPng for IPv6 is IPv6 RIPng. This document describes: • Configuring RIPng Basic Functions • Configuring RIPng Route Control • Tuning and Optimizing the RIPng Network
OSPFv3	OSPFv3 is OSPF version 3 for short, supporting IPv6 and compliant with RFC2740 (OSPF for IPv6). This document describes: • Enabling OSPFv3 • Configuring OSPFv3 Area Parameters • Configuring OSPFv3 Network Types • Configuring OSPFv3 Routing Information Control • Tuning and Optimizing OSPFv3 Networks

Features	Description
IPv6 IS-IS	The IS-IS routing protocol supports multiple network protocols, including IPv6. IS-IS with IPv6 support is called IPv6 IS-IS dynamic routing protocol. This document describes: • Configuring IPv6 IS-IS Basic Functions • Configuring IPv6 IS-IS Routing Information Control
IPv6 BGP	To support multiple network layer protocols, IETF extended BGP-4 by introducing IPv6 BGP. This document describes: • Configuring IPv6 BGP Basic Functions • Controlling Route Distribution and Reception • Configuring IPv6 BGP Route Attributes • Tuning and Optimizing IPv6 BGP Networks • Configuring a Large Scale IPv6 BGP Network
Route Policy	Routing policy is used on the router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed. This document describes: • Defining Filters • Route policy configuration
BFD	Bidirectional forwarding detection (BFD) provides a single mechanism to quickly detect and monitor the connectivity of links in networks. • Configuring BFD Basic Functions • Configuring Protocol-based BFD • Enabling Trap
MCE	Multi-CE (MCE) enables a switch to function as the CEs of multiple VPN instances in a BGP/MPLS VPN network, thus reducing the investment on network equipment. • Introduction to MCE • Configuring a VPN Instance • Configuring Route Exchange between a MCE and a Site • Configuring Route Exchange between a MCE and a PE

Multicast Volume

Table 2-4 Features in Multicast volume

Features	Description
Multicast Overview	This document describes the main concepts in multicast: • Introduction to Multicast • Multicast Models • Multicast Architecture • Multicast Packets Forwarding Mechanism
Multicast Routing and Forwarding	Multicast routing and forwarding refer to some policies that filter RPF routing information for IP multicast support. This document describes: • Multicast routing and forwarding overview • Multicast routing and forwarding configuration

Features	Description
IGMP	Internet Group Management Protocol (IGMP) is a protocol in the TCP/IP suite responsible for management of IP multicast members. This document describes: • IGMP overview • Configuring basic functions of IGMP • Configuring IGMP performance parameters • Configuring IGMP SSM Mapping • Configuring IGMP Proxying
PIM	PIM leverages the unicast routing table created by any unicast routing protocol to provide routing information for IP multicast. This document describes: • Configuring PIM-DM • Configuring PIM-SM • Configuring PIM-SSM • Configuring PIM Common Features
MSDP	Multicast source discovery protocol (MSDP) describes interconnection mechanism of multiple PIM-SM domains. It is used is to discover multicast source information in other PIM-SM domains. This document describes: • MSDP configuration • Configuring an MSDP Peer Connection • Configuring SA Messages Related Parameters
MBGP	As a multicast extension of MP-BGP, MBGP enables BGP to provide routing information for multicast applications. This document describes: • Configuring MBGP Basic Functions • Configuring MBGP Route Attributes • Configuring a Large Scale MBGP Network
IGMP Snooping	Running at the data link layer, IGMP Snooping is a multicast control mechanism on the Layer 2 Ethernet switch and it is used for multicast group management and control. This document describes: • Configuring Basic Functions of IGMP Snooping • Configuring IGMP Snooping Port Functions • Configuring IGMP Snooping Querier • Configuring IGMP Snooping Policy
Multicast VLAN	Multicast VLAN configuration
IPv6 Multicast Routing and Forwarding	IPv6 multicast routing and forwarding refer to some policies that filter RPF routing information for IPv6 multicast support. This document describes: • IPv6 Multicast routing and forwarding overview • IPv6 Multicast routing and forwarding configuration
MLD	MLD is used by an IPv6 router or a Ethernet Switch to discover the presence of multicast listeners on directly-attached subnets. This document describes: • Configuring Basic Functions of MLD • Adjusting MLD Performance
IPv6 PIM	IPv6 PIM discovers multicast source and delivers information to the receivers. This document describes: • Configuring IPv6 PIM-DM • Configuring IPv6 PIM-SM • Configuring IPv6 PIM-SSM • Configuring IPv6 PIM Common Features

Features	Description
IPv6 MBGP	As an IPv6 multicast extension of MP-BGP, IPv6 MBGP enables BGP to provide routing information for IPv6 multicast applications. This document describes: • Configuring IPv6 MBGP Basic Functions • Configuring IPv6 MBGP Route Attributes • Configuring a Large Scale IPv6 MBGP Network
MLD Snooping	Multicast Listener Discovery Snooping (MLD Snooping) is an IPv6 multicast constraining mechanism that runs on Layer 2 devices to manage and control IPv6 multicast groups. This document describes: • Configuring Basic Functions of MLD Snooping • Configuring MLD Snooping Port Functions • Configuring MLD Snooping Querier • Configuring MLD Snooping Policy
IPv6 Multicast VLAN	IPv6 Multicast VLAN configuration

QoS Volume

Table 2-5 Features in the QoS ACL volume

Features	Description
QoS	This document describes: • QoS overview • Traffic classification configuration • Traffic policing Configuration • Traffic shaping Configuration • Line rate configuration • QoS policy configuration • Congestion management • Congestion avoidance configuration • Priority mapping configuration • Traffic mirroring configuration
User Profile	User profile provides a configuration template to save predefined configurations. This document describes: • Creating a User Profile • Configuring a User Profile • Enabling a User Profile

Security Volume

Table 2-6 Features in the Security volume

Features	Description
AAA	Authentication, Authorization and Accounting (AAA) provide a uniform framework used for configuring these three security functions to implement the network security management. This document describes: • Introduction to AAA, RADIUS and HWTACACS • AAA configuration • RADIUS configuration • HWTACACS configuration

Features	Description
802.1x	IEEE 802.1x (hereinafter simplified as 802.1x) is a port-based network access control protocol that is used as the standard for LAN user access authentication. This document describes: • 802.1x overview • 802.1x configuration • 802.1x Guest-VLAN configuration
HABP	On an HABP-capable switch, HABP packets can bypass 802.1x authentication and MAC authentication, allowing communication among switches in a cluster. This document describes: • Introduction to HABP • HABP configuration
MAC Authentication	MAC authentication provides a way for authenticating users based on ports and MAC addresses; it requires no client software to be installed on the hosts. This document describes: • RADIUS-Based MAC Authentication • Local MAC Authentication
Portal	Portal authentication, as its name implies, helps control access to the Internet. This document describes: • Portal overview • Portal configuration
Port Security	Port security is a MAC address-based security mechanism for network access controlling. It is an extension to the existing 802.1x authentication and MAC authentication. This document describes: • Enabling Port Security • Setting the Maximum Number of Secure MAC Addresses • Setting the Port Security Mode • Configuring Port Security Features • Configuring Secure MAC Addresses • Ignoring Authorization Information from the Server
IP Source Guard	By filtering packets on a per-port basis, IP source guard prevents illegal packets from traveling through, thus improving the network security. This document describes: • Configuring a Static Binding Entry • Configuring Dynamic Binding Function
SSH2.0	SSH ensures secure login to a remote device in a non-secure network environment. By encryption and strong authentication, it protects the device against attacks. This document describes: • Configuring Asymmetric Keys • Configuring the Device as an SSH Server • Configuring the Device as an SSH Client • Configuring an SFTP Server • Configuring an SFTP Client
PKI	The Public Key Infrastructure (PKI) is a hierarchical framework designed for providing information security through public key technologies and digital certificates and verifying the identities of the digital certificate owners. This document describes PKI related configuration.
SSL	Secure Sockets Layer (SSL) is a security protocol providing secure connection service for TCP-based application layer protocols, this document describes SSL related configuration.
Public Key	This document describes Public Key Configuration.

Features	Description
ACL	An ACL is used for identifying traffic based on a series of preset matching criteria. This document describes: • ACL overview and ACL types • ACL configuration

System Volume

Table 2-7 Features in the System volume

Features	Description
Login	Upon logging into a device, you can configure user interface properties and manage the system conveniently. This document describes: • How to log in to your Ethernet switch • Introduction to the user interface and common configurations • Logging In Through the Console Port • Logging In Through Telnet • Logging in Through Web-based Network Management System • Logging In Through NMS • Specifying Source IP address/Interface for Telnet Packets • Controlling Login Users
Basic System Configuration	Basic system configuration involves the configuration of device name, system clock, welcome message, user privilege levels and so on. This document describes: • Configuration display • Basic configurations • CLI features
Device Management	Through the device management function, you can view the current condition of your device and configure running parameters. This document describes: • Device management overview • Rebooting a device • Configuring the scheduled automatic execution function • Specifying a file for the next device boot • Upgrading Boot ROM • Configuring a detection interval • Configuring temperature alarm thresholds for a board • Clearing the 16-bit interface indexes not used in the current system • Configuring the system load sharing function • Configuring the traffic forwarding mode of SRPUs • Configuring the working mode of EA LPUs • Enabling the port down function globally • Enabling expansion memory data recovery function on a board • Identifying and diagnosing pluggable transceivers
File System Management	A major function of the file system is to manage storage devices, mainly including creating the file system, creating, deleting, modifying and renaming a file or a directory and opening a file. This document describes: • File system management • Configuration File Management • FTP configuration • TFTP configuration

Features	Description
HTTP	Hypertext Transfer Protocol (HTTP) is used for transferring web page information across the Internet. This document describes: • HTTP Configuration • HTTPS Configuration
SNMP	Simple network management protocol (SNMP) offers a framework to monitor network devices through TCP/IP protocol suite. This document describes: • SNMP overview • Basic SNMP function configuration • SNMP log configuration • Trap configuration • MIB style configuration
RMON	RMON provides an efficient means of monitoring subnets and allows SNMP to monitor remote network devices in a more proactive and effective way. This document describes: • RMON overview • RMON configuration
MAC Address Table Management	A switch maintains a MAC address table for fast forwarding packets. This document describes: • MAC address table overview • Configuring MAC Address Entries • Disabling MAC Address Learning on a VLAN • Configuring MAC Address Aging Timer • Configuring the Maximum Number of MAC Addresses an Ethernet Port or a Port Group Can Learn
System Maintaining and Debugging	For the majority of protocols and features supported, the system provides corresponding debugging information to help users diagnose errors. This document describes: • Maintenance and debugging overview • Maintenance and debugging configuration
Information Center	As the system information hub, Information Center classifies and manages all types of system information. This document describes: • Information Center Overview • Setting to Output System Information to the Console • Setting to Output System Information to a Monitor Terminal • Setting to Output System Information to a Log Host • Setting to Output System Information to the Trap Buffer • Setting to Output System Information to the Log Buffer • Setting to Output System Information to the SNMP Module • Configuring Synchronous Information Output • Disabling a Port from Generating Link Up/Down Logging Information
PoE	The Power over Ethernet (PoE) feature enables the power sourcing equipment (PSE) to feed powered devices (PDs) from Ethernet ports through twisted pair cables. This document describes: • PoE overview • Configuring the PSE • Configuring the PoE interface • Configuring PoE power management • Configuring the PoE monitoring function • Online upgrading the PSE processing software • Configuring a PD Disconnection Detection Mode • Enabling the PSE to detect nonstandard PDs

Features	Description
Track	The track module is used to implement collaboration between different modules through established collaboration objects. The detection modules trigger the application modules to perform certain operations through the track module. This document describes: • Track Overview • Configuring Collaboration Between the Track Module and the Detection Modules • Configuring Collaboration Between the Track Module and the Application Modules
NQA	NQA analyzes network performance, services and service quality by sending test packets to provide you with network performance and service quality parameters. This document describes: • NQA Overview • Configuring the NQA Server • Enabling the NQA Client • Creating an NQA Test Group • Configuring an NQA Test Group • Configuring the Collaboration Function • Configuring Trap Delivery • Configuring the NQA Statistics Function • Configuring Optional Parameters Common to an NQA Test Group • Scheduling an NQA Test Group
NTP	Network Time Protocol (NTP) is the TCP/IP that advertises the accurate time throughout the network. This document describes: • NTP overview • Configuring the Operation Modes of NTP • Configuring Optional Parameters of NTP • Configuring Access-Control Rights • Configuring NTP Authentication
VRRP	Virtual Router Redundancy Protocol (VRRP) combines a group of switches (including a master and multiple backups) on a LAN into a virtual router called VRRP group. VRRP streamlines host configuration while providing high reliability. This document describes: • VRRP overview • IPv4-Based VRRP configuration • IPv6-Based VRRP configuration
Hotfix	Hotfix is a fast, cost-effective method to fix software defects of the device without interrupting the running services. This document describes: • Hotfix Overview • One-Step Patch Installation • Step-by-Step Patch Installation • Step-by-Step Patch Uninstallation • One-Step Patch Uninstallation
Cluster Management	A cluster is a group of network devices. Cluster management is to implement management of large numbers of distributed network devices. This document describes: • Cluster Management Overview • Configuring the Management Device • Configuring the Member Devices • Configuring Access Between the Management Device and Its Member Devices • Adding a Candidate Device to a Cluster • Configuring Advanced Cluster Functions

Features	Description
IRF Stack	Intelligent Resilient Framework (IRF) allows you to build an IRF stack, namely a united device, by interconnecting multiple devices through stack ports. You can manage all the devices in the IRF stack by managing the united device. This document describes: • IRF Stack Overview • IRF Stack Working Process • Configuring IRF Stack • Logging In to an IRF Stack
GR Overview	Graceful Restart ensures the continuity of packet forwarding when a protocol restarts. This document describes: • Introduction to Graceful Restart • Basic Concepts in Graceful Restart • Graceful Restart Communication Procedure • Graceful Restart Mechanism for Several Commonly Used Protocols
Automatic Configuration	Automatic configuration enables a device to automatically obtain and execute the configuration file when it starts up without loading the configuration file. This document describes: • Introduction to Automatic Configuration • Typical Networking of Automatic Configuration • How Automatic Configuration Works
IPC	Inter-Process Communication (IPC) is a reliable communication mechanism among different nodes. This document introduces the commands for Enabling IPC Performance Statistics.

Appendix A Acronyms

#ABCDEFGHIKLMNOPQRSTUVWXZ

Acronyms	Full spelling
#	Return
10GE	Ten-GigabitEthernet
A	Return
AAA	Authentication, Authorization and Accounting
ABC	Activity Based Costing
ABR	Area Border Router
AC	Alternating Current
ACK	ACKnowledgement
ACL	Access Control List
ADSL	Asymmetric Digital Subscriber Line
AFI	Address Family Identifier
ALG	Application Layer Gateway
AM	accounting management
ANSI	American National Standard Institute
AP	Access Point
ARP	Address Resolution Protocol
AS	Autonomous System
ASBR	Autonomous System Border Router
ASCII	American Standard Code for Information Interchange
ASE	Application service element
ASIC	Application Specific Integrated Circuit
ASM	Any-Source Multicast
ASN	Auxiliary Signal Network
AT	Advanced Technology
AT	Adjacency Table
ATM	Asynchronous Transfer Mode
AUX	Auxiliary (port)
B	Return
BC	Bearer Control
BDR	Backup Designated Router
BFD	Bidirectional Forwarding Detection

Acronyms	Full spelling
BGP	Border Gateway Protocol
BIMS	Branch Intelligent Management System
BOOTP	Bootstrap Protocol
BPDU	Bridge Protocol Data Unit
BRI	Basic Rate Interface
BSR	Bootstrap Router
BT	BitTorrent
BT	Burst Tolerance
C	**Return**
CA	Call Appearance
CA	Certificate Authority
CAR	Committed Access Rate
CBS	Committed Burst Size
CBQ	Class Based Queuing
CBR	Constant Bit Rate
CBT	Core-Based Tree
CCITT	International Telephone and Telegraph Consultative Committee
CE	Customer Edge
CFD	Connectivity Fault Detection
CFM	Configuration File Management
CHAP	Challenge Handshake Authentication Protocol
CIDR	Classless Inter-Domain Routing
CIR	Committed Information Rate
CIST	Common and Internal Spanning Tree
CLNP	Connectionless Network Protocol
CPOS	Channelized POS
CPU	Central Processing Unit
CQ	Custom Queuing
CRC	Cyclic Redundancy Check
CR-LSP	Constraint-based Routing LSP
CR-LDP	Constraint-based Routing LDP
CSMA/CD	Carrier Sense Multiple Access/Collision Detect
CSNP	Complete SNP
CSPF	Constraint Shortest Path First
CST	Common Spanning Tree
CT	Call Transfer

Acronyms	Full spelling
CV	Connectivity Verification
D	Return
DAR	Deeper Application Recognition
DCE	Data Circuit-terminal Equipment
DD	Database Description
DDN	Digital Data Network
DHCP	Dynamic Host Configuration Protocol
DIS	Designated IS
DLCI	Data Link Connection Identifier
DLDP	Device Link Detection Protocol
DNS	Domain Name System
DoD	Downstream on Demand
DoS	Denial of Service
DR	Designated Router
DSCP	Differentiated Services Codepoint Priority
DSP	Digital Signal Processor
DTE	Data Terminal Equipment
DU	Downstream Unsolicited
D-V	Distance Vector Routing Algorithm
DVMRP	Distance Vector Multicast Routing Protocol
DWDM	Dense Wavelength Division Multiplexing
E	Return
EACL	Enhanced ACL
EAD	Endpoint Admission Defense
EAP	Extensible Authentication Protocol
EAPOL	Extensible Authentication Protocol over LAN
EBGP	External Border Gateway Protocol
EBS	Excess Burst Size
EGP	Exterior Gateway Protocol
ES	End System
ES-IS	End System-Intermediate System
F	Return
FCoE	Fabric Channel over Ethernet
FC	Forwarding Class
FCS	Frame Check Sequence
FDDI	Fiber Distributed Data Interface

Acronyms	Full spelling
FDI	Forward Defect Indication
FEC	Forwarding Equivalence Class
FFD	Fast Failure Detection
FG	Forwarding Group
FIB	Forwarding information base
FIFO	First In First Out
FQDN	Full Qualified Domain Name
FR	Frame Relay
FRR	Fast ReRoute
FRTT	Fairness Round Trip Time
FT	Functional Test
FTP	File Transfer Protocol
G	Return
GARP	Generic Attribute Registration Protocol
GE	Gigabit Ethernet
GR	Graceful Restart
GRE	Generic Routing Encapsulation
GTS	Generic Traffic Shaping
GVRP	GARP VLAN Registration Protocol
H	Return
HA	High Availability
HABP	HW Authentication Bypass Protocol
HDLC	High-level Data Link Control
HEC	Header Error Control
HoPE	Hiberarchy of PE
HoVPN	Hiberarchy of VPN
HQoS	Hierarchical Quality of Service
HSB	Hot Standby
HTTP	Hyper Text Transport Protocol
H-VPLS	Hiberarchy of VPLS
HVRP	Hierarchy VLAN Register Protocol
HWTACACS	HUAWEI Terminal Access Controller Access Control System
I	Return
IA	Incoming Access
IANA	Internet Assigned Number Authority
IBGP	Internal BGP

Acronyms	Full spelling
IBM	International Business Machines
ICMP	Internet Control Message Protocol
ICMPv6	Internet Control Message Protocol for IPv6
ID	IDentification/IDentity
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IGMP-Snooping	Internet Group Management Protocol Snooping
IGP	Interior Gateway Protocol
ILM	Incoming Label Map
ILS	Internet Locator Service
IN	Intelligent Network
IP	Internet Protocol
IPng	IP Next Generation
IPSec	IP Security
IPTN	IP Phone Telephony Network
IPv6	Internet protocol version 6
IPX	Internet Packet Exchange
IRF	Intelligent Resilient Framework
IS	Intermediate System
ISATAP	Intra-Site Automatic Tunnel Addressing Protocol
ISDN	Integrated Services Digital Network
IS-IS	Intermediate System-to-Intermediate System intra-domain routing information exchange protocol
ISO	International Organization for Standardization
ISP	Internet service provider
ISSU	In Service Software Upgrade
IST	Internal Spanning Tree
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
K	Return
KB	Kilobyte
KEK	Key-encrypting key
L	Return
L2TP	Layer 2 Tunneling Protocol
L2VPN	Layer 2 VPN
L3VPN	Layer 3 VPN

Acronyms	Full spelling
LACP	Link Aggregation Control Protocol
LACPDU	Link Aggregation Control Protocol Data Unit
LAN	Local Area Network
LCP	Link Control Protocol
LDAP	Lightweight Directory Access Protocol
LDP	Label Distribution Protocol
LER	Label Edge Router
LFIB	Label Forwarding Information Base
LIB	Label Information Base
LLC	Link Layer Control
LLDP	Link Layer Discovery Protocol
LOC	Loss of continuity
LOG	Call Logging
LR	Line Rate
LRTT	Loop Round Trip Time
LSA	Link State Advertisement
LSAck	Link State Acknowledgment
LSDB	Link State Database
LSP	Label Switch Path
LSPAGENT	Label Switched Path AGENT
LSPDU	Link State Protocol Data Unit
LSPM	Label Switch Path Management
LSR	Link State Request
LSR	Label Switch Router
LSR-ID	Label Switch Router Identity
LSU	Link State Update
M	Return
MAC	Media Access Control
MAN	Metropolitan Area Network
MaxBC	Max Bandwidth Constraints
MBGP	Multiprotocol Border Gateway Protocol
MD	Multicast Domain
MDI	Medium Dependent Interface
MDT	Multicast Distribution Tree
MED	multi-exit discrimination (MED)
MIB	Management Information Base

Acronyms	Full spelling
MLD	Multicast Listener Discovery Protocol
MLD-Snooping	Multicast Listener Discovery Snooping
MMC	Meet-Me Conference
MODEM	MOdulator-DEModulator
MP	Multilink PPP
MP-BGP	Multiprotocol extensions for BGP-4
MPE	Middle-level PE
MP-group	Multilink Point to Point Protocol group
MPLS	Multiprotocol Label Switching
MPLSFW	Multi-protocol Label Switch Forward
MPM	Multicast Port Management
MSC	Mobile Switching Center
MSDP	Multicast Source Discovery Protocol
MSOH	Multiplex Section Overhead
MSTI	Multi-Spanning Tree Instance
MSTP	Multiple Spanning Tree Protocol
MT	Multicast Tunnel
MTBF	Mean Time Between Failure
MTI	Multicast Tunnel Interface
MTU	Maximum Transmission Unit
MVRF	Multicast VPN Routing and Forwarding
N	**Return**
NAPT	Network Address Port Translation
NAS	Network Access Server
NAT	Net Address Translation
NBMA	Non Broadcast Multi-Access
NBT	NetBIOS over TCP/IP
NCP	Network Control Protocol
ND	Neighborhood discovery
NDA	NetStream Data Analyzer
NDC	Network Data Collector
NDP	Neighbor Discovery Protocol
NetBIOS	Network Basic Input/Output System
NHLFE	Next Hop Label Forwarding Entry
NLPID	Network Layer Protocol Identifier
NLRI	Network Layer Reachable Information

Acronyms	Full spelling
NMS	Network Management Station
NPDU	Network Protocol Data Unit
NPE	Network Provider Edge
NQA	Network Quality Analyzer
NSAP	Network Service Access Point
NSC	NetStream Collector
N-SEL	NSAP Selector
NSSA	Not-So-Stubby Area
NTDP	Neighbor Topology Discovery Protocol
NTP	Network Time Protocol
O	Return
OAM	Operation Administration and Maintenance
OAMPDU	OAM Protocol Data Units
OC-3	OC-3
OID	Object Identifier
OL	Optical Line
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
P	Return
P2MP	Point to MultiPoint
P2P	Point To Point
PAP	Password Authentication Protocol
PCB	Printed Circuit Board
PCM	Pulse Code Modulation
PD	Powered Device
PDU	Protocol Data Unit
PE	Provider Edge
PHP	Penultimate Hop Popping
PHY	Physical layer
PIM	Protocol Independent Multicast
PIM-DM	Protocol Independent Multicast-Dense Mode
PIM-SM	Protocol Independent Multicast-Sparse Mode
PIR	Peak Information Rate
PKCS	Public Key Cryptography Standards
PKI	Public Key Infrastructure
PMTU	Path MTU

Acronyms	Full spelling
PoE	Power over Ethernet
POP	Point Of Presence
POS	Packet Over SDH
PPP	Point-to-Point Protocol
PPTP	Point to Point Tunneling Protocol
PPVPN	Provider-provisioned Virtual Private Network
PQ	Priority Queuing
PRC	Primary Reference Clock
PRI	Primary Rate Interface
PS	Protection Switching
PSE	Power Sourcing Equipment
PSNP	Partial SNP
PVC	Permanent Virtual Channel
PW	Pseudo wires
Q	**Return**
QACL	QoS/ACL
QinQ	802.1Q in 802.1Q
QoS	Quality of Service
QQIC	Querier's Query Interval Code
QRV	Querier's Robustness Variable
R	**Return**
RA	Registration Authority
RADIUS	Remote Authentication Dial in User Service
RAM	random-access memory
RD	Routing Domain
RD	Router Distinguisher
RED	Random Early Detection
RFC	Request For comments
RIP	Routing Information Protocol
RIPng	RIP next generation
RM	Route management
RMON	Remote Monitoring
ROM	Read Only Memory
RP	Rendezvous Point
RPC	Remote Procedure Call
RPF	Reverse Path Forwarding

Acronyms	Full spelling
RPR	Resilient Packet Ring
RPT	Rendezvous Point Tree
RRPP	Rapid Ring Protection Protocol
RSB	Reservation State Block
RSOH	Regenerator Section Overhead
RSTP	Rapid Spanning Tree Protocol
RSVP	Resource ReserVation Protocol
RTCP	Real-time Transport Control Protocol
RTE	Route Table Entry
RTP	Real-time Transport Protocol
RTP	Real-time Transport Protocol
S	**Return**
SA	Source Active
SBM	Subnetwork Bandwidth Management
SCFF	Single Choke Fairness Frame
SD	Signal Degrade
SDH	Synchronous Digital Hierarchy
SETS	Synchronous Equipment Timing Source
SF	Sampling Frequency
SFM	Source-Filtered Multicast
SFTP	Secure FTP
Share-MDT	Share-Multicast Distribution Tree
SIP	Session Initiation Protocol
Site-of-Origin	Site-of-Origin
SLA	Service Level Agreement
SMB	Standby Main Board
SMTP	Simple Mail Transfer Protocol
SNAP	Sub Network Access Point
SNMP	Simple Network Management Protocol
SNP	Sequence Number Packet
SNPA	Subnetwork Points of Attachment
SOH	Section Overhead
SONET	Synchronous Optical NETwork
SOO	Site-of-Origin
SP	Strict Priority Queueing
SPE	Superstratum PE/Sevice Provider-end PE

Acronyms	Full spelling
SPF	Shortest Path First
SPT	Shortest Path Tree
SSH	Secure Shell
SSM	Synchronization Status Marker
SSM	Source-Specific Multicast
ST	Shared Tree
STM-1	SDH Transport Module -1
STM-16	SDH Transport Module -16
STM-16c	SDH Transport Module -16c
STM-4c	SDH Transport Module -4c
STP	Spanning Tree Protocol
SVC	Signalling Virtual Connection
Switch-MDT	Switch-Multicast Distribution Tree
T	Return
TA	Terminal Adapter
TACACS	Terminal Access Controller Access Control System
TDM	Time Division Multiplexing
TCP	Transmission Control Protocol
TE	Traffic Engineering
TEDB	TE DataBase
TFTP	Trivial File Transfer Protocol
TLS	Transparent LAN Service
TLV	Type-Length-Value
ToS	Type of Service
TPID	Tag Protocol Identifier
TRIP	Trigger RIP
TS	Traffic Shaping
TTL	Time to Live
TTY	True Type Terminal
U	Return
UDP	User Datagram Protocol
UPE	Underlayer PE or User-end PE
URL	Uniform Resource Locators
URPF	Unicast Reverse Path Forwarding
USM	User-Based Security Model

Acronyms	Full spelling
V	Return
VBR	Variable Bit Rate
VCI	Virtual Channel Identifier
VE	Virtual Ethernet
VFS	Virtual File System
VLAN	Virtual Local Area Network
VLL	Virtual Leased Lines
VOD	Video On Demand
VoIP	Voice over IP
VOS	Virtual Operate System
VPDN	Virtual Private Dial-up Network
VPDN	Virtual Private Data Network
VPI	Virtual Path Identifier
VPLS	Virtual Private Local Switch
VPN	Virtual Private Network
VRID	Virtual Router ID
VRRP	Virtual Router Redundancy Protocol
VSI	Virtual Switch Interface
VT	Virtual Tributary
VTY	Virtual Type Terminal
W	Return
WAN	Wide Area Network
WFQ	Weighted Fair Queuing
WINS	Windows Internet Naming Service
WLAN	wireless local area network
WRED	Weighted Random Early Detection
WRR	Weighted Round Robin
WTR	Wait-to-Restore
WWW	World Wide Web
X	Return
XGE	Ten-GigabitEthernet
Z	Return
ZBR	Zone Border Router

Access Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The Access Volume is organized as follows:

Features	Description
Ethernet Interface	This document describes: • Basic Ethernet Interface Configuration • Combo Port Configuration • Configuring Flow Control on an Ethernet Interface • Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Interface • Configuring Loopback Testing on an Ethernet Interface • Configuring a Port Group • Configuring Storm Suppression • Setting the Interval for Collecting Ethernet Interface Statistics • Enabling Forwarding of Jumbo Frames • Enabling Loopback Detection on an Ethernet Interface • Configuring the MDI Mode for an Ethernet Interface • Testing the Cable on an Ethernet Interface • Configuring the Storm Constrain Function on an Ethernet Interface
Link aggregation	Link aggregation aggregates multiple physical Ethernet ports into one logical link. This document describes: • Basic Concepts of Link Aggregation • Configuring a Static Aggregation Group • Configuring a Dynamic Aggregation Group • Configuring an Aggregate Interface • Configuring a Load Sharing Mode for Load-Sharing Link Aggregation Groups
Port Isolation	The port isolation feature allows you to isolate different ports within the same VLAN. This document describes: • Introduction to Port Isolation • Configuring the Isolation Group

Features	Description
Service Loopback Group	To increase service redirecting throughput, you can bundle multiple service loopback ports into a logical link, called a service loopback group. This document describes: • Introduction to Service Loopback Groups • Configuring a Service Loopback Group
DLDP	In the use of fibers, link errors, namely unidirectional links, are likely to occur. DLDP is designed to detect such errors. This document describes: • DLDP Introduction • Enabling DLDP • Setting DLDP Mode • Setting the Interval for Sending Advertisement Packets • Setting the DelayDown Timer • Setting the Port Shutdown Mode • Configuring DLDP Authentication • Resetting DLDP State
LLDP	LLDP enables a device to maintain and manage its own and its immediate neighbor's device information, based on which the network management system detects and determines the conditions of the communications links. This document describes: • Introduction to LLDP • Performing Basic LLDP Configuration • Configuring the Encapsulation Format for LLDPDUs • Configuring the Encapsulation Format of the Management Address • Configuring CDP Compatibility • Configuring LLDP Trapping
Smart Link	Smart Link is a solution for active-standby link redundancy backup and rapid transition in dual-uplink networking. This document describes: • Smart Link Overview • Configuring a Smart Link Device • Configuring an Associated Device
Monitor Link	Monitor link is a port collaboration function used to enable a device to be aware of the up/down state change of the ports on an indirectly connected link. This document describes: • Monitor Link Overview • Configuring Monitor Link
VLAN	Using the VLAN technology, you can partition a LAN into multiple logical LANs. This document describes: • Introduction to VLAN • Types of VLAN • Isolate-user-vlan configuration • Introduction and Configuration of Voice VLAN
GVRP	GVRP is a GARP application. This document describes: • GARP overview • GVRP configuration • GARP Timers configuration

Features	Description
QinQ	As defined in IEEE802.1Q, 12 bits are used to identify a VLAN ID, so a device can support a maximum of 4094 VLANs. The QinQ feature extends the VLAN space by allowing Ethernet frames to travel across the service provider network with double VLAN tags. This document describes: • Introduction to QinQ • Configuring basic QinQ • Configuring Selective QinQ • Configuring the TPID Value in VLAN Tags
BPDU Tunnel	BPDU tunneling enables transparently transmission of customer network BPDU frames over the service provider network. This document describes: • Introduction to BPDU Tunneling • Configuring BPDU Transparent Transmission • Configuring Destination Multicast MAC Address for BPDU Tunnel Frames
VLAN Mapping	The VLAN mapping feature maps CVLAN tags to SVLAN tags. This document describes: • Configuring One-to-One VLAN Mapping • Configuring Many-to-One VLAN Mapping • Configuring Two-to-Two VLAN Mapping
Ethernet OAM	Ethernet OAM is a tool monitoring Layer-2 link status. It helps network administrators manage their networks effectively. This document describes: • Ethernet OAM overview • Configuring Basic Ethernet OAM Functions • Configuring Link Monitoring • Enabling OAM Loopback Testing
Connectivity Fault Detection	Connectivity fault detection is an end-to-end, per-VLAN link-layer OAM mechanism for link connectivity detection, fault verification, and fault location. This document describes: • Connectivity Fault Detection Overview • Basic Configuration Tasks • Configuring CC on MEPs • Configuring LB on MEPs • Configuring LT on MEPs
MSTP	MSTP is used to eliminate loops in a LAN. It is compatible with STP and RSTP. This document describes: • Introduction to MSTP • Configuring the Root Bridge • Configuring Leaf Nodes • Performing mCheck • Configuring Digest Snooping • Configuring No Agreement Check • Configuring Protection Functions

Features	Description
RRPP	RRPP is a link layer protocol designed for Ethernet rings. RRPP can prevent broadcast storms caused by data loops when an Ethernet ring is healthy, and rapidly restore the communication paths between the nodes after a link is disconnected on the ring. This document describes: • RRPP overview • Configuring Master Node • Configuring Transit Node • Configuring Edge Node • Configuring Assistant Edge Node • Configuring Ring Group
Port Mirroring	Port mirroring copies packets passing through a port to another port connected with a monitoring device for packet analysis to help implement network monitoring and troubleshooting. This document describes: • Port Mirroring overview • Local port mirroring configuration • Remote port mirroring configuration

Table of Contents

1 Ethernet Interface Configuration

General Ethernet Interface Configuration

Combo Port Configuration

Introduction to Combo port

A Combo port can operate as either an optical port or an electrical port. Inside the device there is only one forwarding interface. For a Combo port, the electrical port and the corresponding optical port are TX-SFP multiplexed. You can specify a Combo port to operate as an electrical port or an optical port. That is, a Combo port cannot operate as both an electrical port and an optical port simultaneously. When one is enabled, the other is automatically disabled.

Configuring Combo port state

Follow these steps to configure the state of a Combo port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Enable a specified Combo port	**undo shutdown**	Optional By default, of the two ports in a Combo port, the one with a smaller port ID is enabled.

Note

In case of a Combo port, only one interface (either the optical port or the electrical port) is active at a time. That is, once the optical port is active, the electrical port will be inactive automatically, and vice versa.

Basic Ethernet Interface Configuration

Configuring an Ethernet interface

Three types of duplex modes are available to Ethernet interfaces:

- Full-duplex mode (full). Interfaces operating in this mode can send and receive packets simultaneously.
- Half-duplex mode (half). Interfaces operating in this mode can either send or receive packets at a given time.

- Auto-negotiation mode (auto). Interfaces operating in this mode determine their duplex mode through auto-negotiation.

Similarly, if you configure the transmission rate for an Ethernet interface by using the **speed** command with the **auto** keyword specified, the transmission rate is determined through auto-negotiation too. For a Gigabit Ethernet interface, you can specify the transmission rate by its auto-negotiation capacity.

Follow these steps to configure an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Set the description string	**description** *text*	Optional By default, the description of an interface is the interface name followed by the "interface" string, **GigabitEthernet1/0/1 Interface** for example.
Set the duplex mode	**duplex** { **auto** | **full** | **half** }	Optional **auto** by default. The optical interface of a Combo port does not support the **half** keyword.
Set the transmission rate	**speed** { **10** | **100** | **1000** | **auto** }	Optional The optical interface of a Combo port does not support the **10** or **100** keyword. By default, the port speed is in the auto-negotiation mode.
Shut down the Ethernet interface	**shutdown**	Optional By default, an Ethernet interface is in up state. To bring up an Ethernet interface, use the **undo shutdown** command.

Note

10-Gigabit Ethernet ports do not support the **duplex** command or the **speed** command.

Configuring Flow Control on an Ethernet Interface

When flow control is enabled on both sides, if traffic congestion occurs at the ingress interface, it will send a Pause frame notifying the egress interface to temporarily suspend the sending of packets. The egress interface is expected to stop sending any new packet when it receives the Pause frame. In this way, flow control helps to avoid dropping of packets. Note that this will be possible only after flow control is enabled on both the ingress and egress interfaces.

Follow these steps to enable flow control on an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Enable flow control	**flow-control**	Required Disabled by default

Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Interface

An Ethernet interface operates in one of the two physical link states: up or down. During the suppression time, physical-link-state changes will not be propagated to the system. Only after the suppression time has elapsed will the system be notified of the physical-link-state changes by the physical layer. This functionality reduces the extra overhead occurred due to frequent physical-link-state changes within a short period of time.

Follow these steps to configure the suppression time of physical-link-state changes on an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Configure the up/down suppression time of physical-link-state changes	**link-delay** *delay-time*	Required By default, the physical-link-state change suppression time is not configured.

Configuring Loopback Testing on an Ethernet Interface

You can enable loopback testing to check whether the Ethernet interface functions properly. Note that no data packets can be forwarded during the testing. Loopback testing falls into the following two categories:

- Internal loopback testing, which is performed within switching chips to test the functions related to the Ethernet interfaces.
- External loopback testing, which is used to test the hardware functions of an Ethernet interface. To perform external loopback testing on an Ethernet interface, you need to install a loopback plug on the Ethernet interface. In this case, packets sent from the interface are received by the same interface.

Follow these steps to enable Ethernet interface loopback testing:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Enable loopback testing	**loopback** { **external** | **internal** }	Optional Disabled by default.

Note

- As for the internal loopback test and external loopback test, if an interface is down, only the former is available on it; if the interface is shut down, both are unavailable.
- The **speed**, **duplex**, **mdi**, and **shutdown** commands are not applicable during loopback testing.
- With the loopback testing enabled, the Ethernet interface operates in full duplex mode. With the loopback testing disabled, the original configurations will be restored.

Configuring a Port Group

The devices allow you to configure some functions on multiple interfaces at a time by assigning the interfaces to a port group in addition to configuring them on a per-interface basis. This is helpful when you have to configure a feature in the same way on multiple interfaces.

A port group is created manually and the settings you made on it apply to all group member interfaces. Note that even though the settings are made on the port group, they are saved on an interface basis rather than on a port group basis. Thus, you can only view the settings in the view of each interface with the **display current-configuration** command or the **display this** command.

Follow these steps to configure a manual port group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a manual port group and enter manual port group view	**port-group manual** *port-group-name*	Required
Add Ethernet interfaces to the manual port group	**group-member** *interface-list*	Required

Configuring Storm Suppression

You can use the following commands to suppress the broadcast, multicast, and unknown unicast traffic. In interface configuration mode, the suppression ratio indicates the maximum broadcast, multicast or unknown unicast traffic that is allowed to pass through an interface. When the broadcast, multicast, or unknown unicast traffic over the interface exceeds the threshold, the system will discard the extra packets so that the broadcast, multicast or unknown unicast traffic ratio can drop below the limit to ensure that the network functions properly.

Note

The storm suppression ratio settings configured for an Ethernet interface may get invalid if you enable the storm constrain for the interface. For information about the storm constrain function, see Configuring the Storm Constrain Function on an Ethernet Interface.

Follow these steps to set storm suppression ratios for one or multiple Ethernet interfaces:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command. If configured in Ethernet interface view, this feature takes effect on the current port only; if configured in port group view, this feature takes effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Set the broadcast storm suppression ratio		**broadcast-suppression** { *ratio* | **pps** *max-pps* }	Optional By default, all broadcast traffic is allowed to pass through an interface, that is, broadcast traffic is not suppressed.
Set the multicast storm suppression ratio		**multicast-suppression** { *ratio* | **pps** *max-pps* }	Optional By default, all multicast traffic is allowed to pass through an interface, that is, multicast traffic is not suppressed.
Set the unknown unicast storm suppression ratio		**unicast-suppression** { *ratio* | **pps** *max-pps* }	Optional By default, all unknown unicast traffic is allowed to pass through an interface, that is, unknown unicast traffic is not suppressed.

Note

If you set storm suppression ratios in Ethernet interface view or port group view repeatedly for an Ethernet interface that belongs to a port group, only the latest settings take effect.

Setting the Interval for Collecting Ethernet Interface Statistics

Follow these steps to configure the interval for collecting interface statistics:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the interval for collecting interface statistics	**interface** *interface-type interface-number*	Optional The default interval for collecting interface statistics is 300 seconds.
	flow-interval *interval*	

Enabling Forwarding of Jumbo Frames

Due to tremendous amount of traffic occurring on an Ethernet interface, it is likely that some frames greater than the standard Ethernet frame size are received. Such frames (called jumbo frames) will be dropped. With forwarding of jumbo frames enabled, the system does not drop all the jumbo frames. Instead, it continues to process jumbo frames with a size greater than the standard Ethernet frame size and yet within the specified parameter range.

In interface configuration mode (Ethernet interface view/port-group view), you can set the length of jumbo frames that can pass through the Ethernet interface.

- If you execute the command in Ethernet interface view, the configurations take effect only on the current interface.
- If you execute the command in port-group view, the configurations take effect on all ports in the port group.

Follow these steps to enable the forwarding of jumbo frames:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable the forwarding of jumbo frames	In port-group view	**port-group manual** *port-group-name*	Use any command. By default, the device allows jumbo frames with the length of 9.216 bytes to pass through all Layer 2 Ethernet interfaces.
		jumboframe enable	
	In Ethernet interface view	**interface** *interface-type interface-number*	
		jumboframe enable	

Enabling Loopback Detection on an Ethernet Interface

If a port receives a packet that it sent out, a loop occurs. Loops may cause broadcast storms. The purpose of loopback detection is to detect loops on an interface.

When loopback detection is enabled on an Ethernet interface, the device periodically checks whether the ports have any external loopback. If it detects a loopback on a port, the device will set that port to be under loopback detection mode.

- If loops are detected on an access port, the port will be blocked. Meanwhile, trap messages will be sent to the terminal, and the corresponding MAC address forwarding entries will be removed.
- If loops are detected on a trunk port or a hybrid port, trap messages are sent to the terminal. If the loopback detection control function is also enabled on the port, the port will be blocked, trap

messages will be sent to the terminal, and the corresponding MAC address forwarding entries will be removed.

Follow these steps to configure loopback detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable global loopback detection	**loopback-detection enable**	Required Disabled by default
Configure the interval for port loopback detection	**loopback-detection interval-time** *time*	Optional 30 seconds by default
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Enable loopback detection on a port	**loopback-detection enable**	Required Disabled by default
Enable loopback detection control on a trunk port or a hybrid port	**loopback-detection control enable**	Optional Disabled by default
Enable loopback detection in all the VLANs to which trunk or hybrid ports belong	**loopback-detection per-vlan enable**	Optional Enabled only in the default VLAN(s) with trunk port or hybrid ports

Caution

- Loopback detection on a given port is enabled only after the **loopback-detection enable** command has been configured in both system view and the interface view of the port.
- Loopback detection on all ports will be disabled after the configuration of the **undo loopback-detection enable** command under system view.

Configuring the MDI Mode for an Ethernet Interface

Note

10-Gigabit Ethernet ports and combo ports operating as optical interfaces do not support this function.

Two types of Ethernet cables can be used to connect Ethernet devices: crossover cable and straight-through cable. To accommodate these two types of cables, an Ethernet interface on a device can operate in one of the following three Medium Dependent Interface (MDI) modes:

- Across mode
- Normal mode
- Auto mode

An Ethernet interface is composed of eight pins. By default, each pin has its particular role. For example, pin 1 and pin 2 are used for transmitting signals; pin 3 and pin 6 are used for receiving signals. You can change the pin roles through setting the MDI mode. For an Ethernet interface in normal mode, the pin roles are not changed. For an Ethernet interface in across mode, pin 1 and pin 2 are used for receiving signals; pin 3 and pin 6 are used for transmitting signals. To enable normal communication, you should connect the local transmit pins to the remote receive pins. Therefore, you should configure the MDI mode depending on the cable types.

- Normally, the auto mode is recommended. The other two modes are useful only when the device cannot determine the cable type.
- When straight-through cables are used, the local MDI mode must be different from the remote MDI mode.
- When crossover cables are used, the local MDI mode must be the same as the remote MDI mode, or the MDI mode of at least one end must be set to **auto**.

Follow these steps to configure the MDI mode for an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Configure the MDI mode for the Ethernet interface	**mdi** { **across** | **auto** | **normal** }	Optional Defaults to **auto**. That is, the Ethernet interface determines the physical pin roles (transmit or receive) through negotiation.

Testing the Cable on an Ethernet Interface

Note

- 10-Gigabit Ethernet ports and Combo ports operating as optical interfaces do not support this feature.
- A link in the up state goes down and then up automatically if you perform the operation described in this section on one of the Ethernet interfaces forming the link.

Follow these steps to test the current operating state of the cable connected to an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Test the cable connected to the Ethernet interface once	**virtual-cable-test**	Required

Configuring the Storm Constrain Function on an Ethernet Interface

The storm constrain function suppresses packet storms in an Ethernet. With this function enabled on an interface, the system detects the multicast traffic, or broadcast traffic passing through the interface periodically and takes corresponding actions (that is, blocking or shutting down the interface and sending trap messages and logs) when the traffic detected exceeds the threshold.

Caution

Alternatively, you can configure the storm suppression function to control a specific type of traffic. As the function and the storm constrain function are mutually exclusive, do not enable them at the same time on an Ethernet interface. For example, with broadcast storm suppression ratio set on an Ethernet interface, do not enable the storm constrain function for broadcast traffic on the interface. Refer to Configuring Storm Suppression for information about the storm suppression function.

With the storm constrain function enabled on an Ethernet interface, you can specify the system to act as follows when the traffic detected exceeds the threshold.

- Blocking the interface. In this case, the interface is blocked and thus stops forwarding the traffic of this type till the traffic detected is lower than the threshold. Note that an interface blocked by the storm constrain function can still forward other types of traffic and monitor the blocked traffic.
- Shutting down the interface. In this case, the interface is shut down and stops forwarding all types of traffic. Interfaces shut down by the storm constrain function can only be brought up by using the **undo shutdown** command or disabling the storm constrain function.

Follow these steps to configure the storm constrain function on an Ethernet interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the interval for generating traffic statistics	**storm-constrain interval** *seconds*	Optional 10 seconds by default
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Enable the storm constrain function and set the lower threshold and the upper threshold	**storm-constrain** { **broadcast** | **multicast** } { **pps** | **kbps** | **ratio** } *max-pps-values min-pps-values*	Required Disabled by default
Set the action to be taken when the traffic exceeds the upper threshold	**storm-constrain control** { **block** | **shutdown** }	Optional Disabled by default
Specify to send trap messages when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold	**storm-constrain enable trap**	Optional By default, the system sends trap messages when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold.

To do…	Use the command…	Remarks
Specify to send log when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold	**storm-constrain enable log**	Optional By default, the system sends log when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold.

Note

- For network stability sake, configure the interval for generating traffic statistics to a value that is not shorter than the default.
- The storm constrain function, after being enabled, requires a complete statistical period (specified by the storm-constrain interval command) to collect traffic data, and analyzes the data in the next period. Thus, it is normal that a period longer than one statistic period is waited for a control action to happen if you enable the function while the packet storm is present. However, the action will be taken within two periods.
- The storm constrain function is applicable to multicast packets, and broadcast packets; and you can specify the upper and lower threshold for any of the three types of packets.

Displaying and Maintaining an Ethernet Interface

To do…	Use the command…	Remarks
Display the current state of an interface/subinterface and the related information	**display interface** [*interface-type* [*interface-number*]]	Available in any view
Display the summary of an interface/subinterface	**display brief interface** [*interface-type* [*interface-number*]] [**|** { **begin** | **exclude** | **include** } *regular-expression*]	Available in any view
Display information about discarded packets on an interface	**display packet-drop interface** [*interface-type* [*interface-number*]]	Available in any view
Display summary information about discarded packets on all interfaces	**display packet-drop summary**	Available in any view
Clear the statistics of an interface/subinterface	**reset counters interface** [*interface-type* [*interface-number*]]	Available in user view
Clear the statistics of discarded packets on an interface	**reset packet-drop interface** [*interface-type* [*interface-number*]]	Available in user view
Display the Combo ports and the corresponding optical/electrical ports	**display port combo**	Available in any view

To do…	Use the command…	Remarks
Display the information about a manual port group or all the port groups	**display port-group manual** [**all** | **name** *port-group-name*]	Available in any view
Display the information about the loopback function	**display loopback-detection**	Available in any view
Display the information about storm constrain	**display storm-constrain** [**broadcast** | **multicast**] [**interface** *interface-type interface-number*]	Available in any view

Table of Contents

1 Link Aggregation Configuration

When configuring link aggregation, go to these sections for information you are interested in:

Overview

Link aggregation aggregates multiple physical Ethernet ports into one logical link, also called an aggregation group.

It allows you to increase bandwidth by distributing traffic across the member ports in the aggregation group. In addition, it provides reliable connectivity because these member ports can dynamically back up each other.

Basic Concepts of Link Aggregation

Aggregate interface

An aggregate interface is a logical Layer 2 or Layer-3 aggregate interface.

Aggregation group

An aggregation group is a collection of Ethernet interfaces. When you create an aggregate interface, an aggregation group numbered the same is created automatically depending on the type of the aggregate interface:

- If the aggregate interface is a Layer 2 interface, a Layer 2 aggregation group is created. You can assign only Layer 2 Ethernet interfaces to the group.
- If the aggregate interface is a Layer-3 interface, a Layer-3 aggregation group is created. You can assign only Layer-3 Ethernet interfaces to the group.

Note

The current device only supports Layer 2 aggregation groups.

States of the member ports in an aggregation group

A member port in an aggregation group can be in one of the following two states:

- Selected: a selected port can forward user traffic.
- Unselected: an unselected port cannot forward user traffic.

The rate of an aggregate interface is the sum of the selected member ports' rates. The duplex mode of an aggregate interface is consistent with that of the selected member ports. Note that all selected member ports use the same duplex mode.

For how the state of a member port is determined, refer to Static aggregation mode and Dynamic aggregation mode.

LACP protocol

The Link Aggregation Control Protocol (LACP) is defined in IEEE 802.3ad. It uses link aggregation control protocol data units (LACPDUs) for information exchange between LACP-enabled devices.

LACP is automatically enabled on interfaces in a dynamic aggregation group. For information about dynamic aggregation groups, refer to Dynamic aggregation mode. An LACP-enabled interface sends LACPDUs to notify the remote system (the partner) of its system LACP priority, system MAC address, LACP port priority, port number, and operational key. Upon receiving an LACPDU, the partner compares the received information with the information received on other interfaces to determine the interfaces that can operate as selected interfaces. This allows the two systems to reach an agreement on which link aggregation member ports should be placed in selected state.

Operational key

When aggregating ports, link aggregation control automatically assigns each port an operational key based on the port attributes, including the configurations of the port rate, duplex mode and link state.

In a link aggregation group, all member ports in the selected state have the same operation key.

Class-two configurations

Class-two configurations are listed in Table 1-1. In an aggregation group, if the configurations of a member port are different from the class-two configurations, that member port cannot be a selected port.

Table 1-1 Class-two configurations

Type	Considerations
Port isolation	Whether a port has joined an isolation group, and the isolation group that the port belongs to
QinQ	QinQ enable state (enable/disable), outer VLAN tags to be added, inner-to-outer VLAN priority mappings, inner-to-outer VLAN tag mappings, inner VLAN ID substitution mappings
VLAN	Permitted VLAN IDs, default VLAN, link type (trunk, hybrid, or access), IP subnet-based VLAN configuration, protocol-based VLAN configuration, tag mode
MAC address learning	MAC address learning capability, MAC address learning limit, forwarding of frames with unknown destination MAC addresses after the upper limit of the MAC address table is reached

Note

- Some configurations are called class-one configurations. Such configurations, for example, GVRP and MSTP, can be configured on aggregate interfaces and member ports but are not considered during operational key calculation.
- The change of a class-two configuration setting may affect the select state of link aggregation member ports and thus the ongoing service. To prevent unconsidered change, a message warning of the hazard will be displayed when you attempt to change a class-two setting, upon which you can decide whether to continue your change operation.

Link Aggregation Modes

Depending on the link aggregation procedure, link aggregation operates in one of the following two modes:

- Static aggregation mode
- Dynamic aggregation mode

Static aggregation mode

LACP is disabled on the member ports in a static aggregation group. In a static aggregation group, the system sets a port to selected or unselected state by the following rules:

- Select a port as the reference port from the ports that are in up state and with the same class-two configurations as the corresponding aggregate interface. These ports are selected in the order of full duplex/high speed, full duplex/low speed, half duplex/high speed, and half duplex/low speed, with full duplex/high speed being the most preferred. If two ports with the same duplex mode/speed pair are present, the one with the lower port number wins out.
- Consider the ports in up state with the same port attributes and class-two configurations as the reference port as candidate selected ports, and set all others in the unselected state.
- Static aggregation limits the number of selected ports in an aggregation group. When the number of the candidate selected ports is under the limit, all the candidate selected ports become selected ports. When the limit is exceeded, set the candidate selected ports with smaller port numbers in the selected state and those with greater port numbers in the unselected state.
- If all the member ports are down, set their states to unselected.
- Set the ports that cannot aggregate with the reference port to the unselected state.

Caution

A port that joins the aggregation group after the limit on the number of selected ports has been reached will not be placed in the selected state even if it should be in normal cases. This can prevent the ongoing traffic on the current selected ports from being interrupted. You should avoid the situation however, as this may cause the selected/unselected state of a port to change after a reboot.

Dynamic aggregation mode

LACP is enabled on member ports in a dynamic aggregation group.

In a dynamic aggregation group,

- A selected port can receive and transmit LACPDUs.
- An unselected port can receive and send LACPDUs only if it is up and with the same configurations as those on the aggregate interface.

In a dynamic aggregation group, the system sets the ports to selected or unselected state in the following steps:

1) The local system (the actor) negotiates with the remote system (the partner) to determine port state based on the port IDs on the end with the preferred system ID. The following is the detailed negotiation procedure:

- Compare the system ID (comprising the system LACP priority and the system MAC address) of the actor with that of the partner. The system with the lower LACP priority wins out. If they are the same, compare the system MAC addresses. The system with the smaller MAC address wins out.
- Compare the port IDs of the ports on the system with the smaller system ID. A port ID comprises a port LACP priority and a port number. First compare the port LACP priorities. The port with the lower LACP priority wins out. If two ports are with the same LACP priority, compare their port numbers. The port with the smaller port ID, that is, the port with smaller port number, is selected as the reference port.
- If a port (in up state) is with the same port attributes and class-two configuration as the reference port, and the peer port of the port is with the same port attributes and class-two configurations as the peer port of the reference port, consider the port as a candidate selected port; otherwise set the port to the unselected state.
- The number of selected ports that an aggregation group can contain is limited. When the number of candidate selected ports is under the limit, all the candidate selected ports are set to selected state. When the limit is exceeded, the system selects the candidate selected ports with smaller port IDs as the selected ports, and set other candidate selected ports to unselected state. At the same time, the peer device, being aware of the changes, changes the state of its ports accordingly.

2) Set the ports that cannot aggregate with the reference port to the unselected state.

Note

For static and dynamic aggregation modes:

- In an aggregation group, the port to be a selected port must be the same as the reference port in port attributes, and class-two configurations. To keep these configurations consistent, you should configure the port manually.
- Because changing a port attribute or class-two configuration setting of a port may cause the select state of the port and other member ports to change and thus affects services, you are recommended to do that with caution.

Load Sharing Mode of an Aggregation Group

A link aggregation groups operates in load sharing aggregation mode or non-load sharing mode.

The system sets the load sharing mode of an aggregation group as follows:

- When hardware resources are available, a link aggregation group with at least two selected ports operates in load sharing mode. The load sharing mode of a link aggregation group with only one selected port is non-load sharing mode.
- After hardware resources become depleted (a number of 128 link aggregation groups have been created in the system), all the link aggregation groups operate in non-load sharing mode.

 **Note**

- After you remove all ports but one selected port from a load-sharing aggregation group, the aggregation group remains to be a load sharing group.
- A load-sharing aggregation group contains at least one selected port while a non-load-sharing aggregation group can only have one selected port at most.
- After hardware resources become depleted, all new link aggregation groups operate in non-load sharing mode. They will not perform load sharing even after resources become available again for example after some aggregation groups are removed. To have them perform load sharing, you can re-enable their corresponding aggregation interfaces by shutting down and then bringing up the interfaces.

Link Aggregation Configuration Task List

Complete the following tasks to configure link aggregation:

Task		Remarks
Configuring an Aggregation Group	Configuring a Static Aggregation Group	Required Perform either of the tasks
	Configuring a Dynamic Aggregation Group	
Configuring an Aggregate Interface	Configuring the Description of an Aggregate Interface	Optional
	Enabling LinkUp/LinkDown Trap Generation for an Aggregate Interface	Optional
	Shutting Down an Aggregate Interface	Optional
Configuring a Load Sharing Mode for Load-Sharing Link Aggregation Groups		Optional

Configuring an Aggregation Group

Note

- The following ports cannot be assigned to an aggregation group: Stack ports, RRPP-enabled ports, MAC address authentication-enabled ports, port security-enabled ports, IP source guard-enabled ports, and 802.1x-enabled ports.
- You are recommended not to assign reflector ports of port mirroring to an aggregation group. For details about reflector ports, refer to *Port Mirroring Configuration* in the *Access Volume*.

Configuring a Static Aggregation Group

Follow these steps to configure a Layer 2 static aggregation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a Layer 2 aggregate interface and enter the Layer 2 aggregate interface view	**interface bridge-aggregation** *interface-number*	Required When you create a Layer 2 aggregate interface, a Layer 2 static aggregation group numbered the same is created automatically.
Exit to system view	**quit**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Repeat the two steps to assign multiple Ethernet interfaces to the aggregation group.
Assign the Ethernet interface to the aggregation group	**port link-aggregation group** *number*	

Caution

- Removing a Layer 2 aggregate interface also removes the corresponding aggregation group. At the same time, the member ports of the aggregation group, if any, leave the aggregation group.
- To guarantee a successful static aggregation, ensure that the ports at the two ends of each link to be aggregated are consistent in the selected/unselected state.

Configuring a Dynamic Aggregation Group

Follow these steps to configure a Layer 2 dynamic aggregation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Set the system LACP priority	**lacp system-priority** *system-priority*	Optional By default, the system LACP priority is 32768. Changing the system LACP priority may affect the selected/unselected state of the ports in the dynamic aggregation group.
Create a Layer 2 aggregate interface and enter the Layer 2 aggregate interface view	**interface bridge-aggregation** *interface-number*	Required When you create a Layer 2 aggregate interface, a Layer 2 static aggregation group numbered the same is created automatically.
Configure the aggregation group to work in dynamic aggregation mode	**link-aggregation mode dynamic**	Required By default, an aggregation group works in static aggregation mode.
Exit to system view	**quit**	—
Enter Layer 2 Ethernet interface view	**interface** *interface-type interface-number*	Required Repeat the two steps to assign multiple Ethernet interfaces to the aggregation group.
Assign the Ethernet interface to the aggregation group	**port link-aggregation group** *number*	
Assign the port a LACP priority	**lacp port-priority** *port-priority*	Optional By default, the LACP priority of a port is 32768. Changing the LACP priority of a port may affect the selected/unselected state of the ports in the dynamic aggregation group.

Caution

- Removing a dynamic aggregate interface also removes the corresponding aggregation group. At the same time, the member ports of the aggregation group, if any, leave the aggregation group.
- To guarantee a successful dynamic aggregation, ensure that the peer ports of the ports aggregated at one end are also aggregated. The two ends can automatically negotiate the selected state of the ports.
- When a load-sharing aggregation group becomes a non-load-sharing aggregation group because of insufficient load sharing resources, one of the following problems may occur: the number of selected ports of the actor is inconsistent with that of the partner, which may result in incorrect traffic forwarding; the peer port of a selected port is an unselected one, which may result in upper-layer protocol and traffic forwarding anomalies. You should fully consider the situation when making configuration.

Configuring an Aggregate Interface

You can perform the following configurations for an aggregate interface:

- Configuring the Description of an Aggregate Interface
- Enabling LinkUp/LinkDown Trap Generation for an Aggregate Interface
- Shutting Down an Aggregate Interface

Configuring the Description of an Aggregate Interface

Follow these steps to configure the description of an aggregate interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Layer 2 aggregate interface view	**interface bridge-aggregation** *interface-number*	—
Configure the description of the aggregate interface	**description** *text*	Optional By default, the description of an interface is *interface-name* **Interface**, such as **Bridge-Aggregation1 Interface**.

Enabling LinkUp/LinkDown Trap Generation for an Aggregate Interface

To enable an aggregate interface to generate linkUp/linkDown trap messages when the state of the interface changes, you should enable linkUp/linkDown trap generation on the aggregate interface.

Follow these steps to enable linkUp/linkDown trap generation for an aggregate interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable the trap function globally	**snmp-agent trap enable** [**standard** [**linkdown** | **linkup**] *]	Optional By default, linkUp/linkDown trap generation is enabled globally and on all interfaces.
Enter aggregate interface view	**interface bridge-aggregation** *interface-number*	—
Enable linkUp/linkDown trap generation for the aggregate interface	**enable snmp trap updown**	Optional Enabled by default

Shutting Down an Aggregate Interface

Shutting down or bringing up an aggregate interface affects the selected state of the ports in the corresponding aggregation group. When an aggregate interface is shut down, all selected ports in its aggregation group become unselected; when the aggregate interface is brought up, the selected state of the ports in the corresponding aggregation group is re-calculated.

Follow these steps to shut down an aggregate interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Layer 2 aggregate interface view	**interface bridge-aggregation** *interface-number*	—
Shut down the aggregate interface	**shutdown**	Required By default, aggregate interfaces are up.

Caution

After shutting down an aggregate interface, you are recommended not to use the **shutdown** command and then the **undo shutdown** command on the member interfaces of the corresponding link aggregation group. Otherwise, the member interfaces may be brought up.

Configuring a Load Sharing Mode for Load-Sharing Link Aggregation Groups

The hash algorithm is adopted to calculate load sharing for load-sharing link aggregation groups. Hash keys used for calculation could be service port numbers, IP addresses, MAC addresses, incoming ports, or any combinations of them. One hash key or a combination of multiple hash keys represents a load sharing mode. You can change the load sharing mode of a link aggregation group for different types of

traffic as needed. For example, for Layer 3 traffic, you can use IP addresses as hash keys for load sharing calculation.

Follow these steps to configure load sharing mode for link aggregation groups:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the link aggregation load sharing mode	**link-aggregation load-sharing mode** { **destination-ip** | **destination-mac** | **destination-port** | **ingress-port** | **source-ip** | **source-mac** | **source-port** } *	Optional By default, the hash keys for Layer 2 packets are source/destination MAC addresses, and those for Layer-3 packets are source/destination IP addresses. The setting you made applies to all load-sharing link aggregation groups.

Note

Currently, the hash keys for a switch are source IP addresses, destination IP addresses, source MAC addresses, destination MAC addresses, source ports, destination ports, or the combination of these fields carried in packets (excluding the combination of MAC addresses with IP addresses, source ports, or destination ports). The **ingress-port** parameter can only be used as a hash key when combined with a MAC address, not when combined with an IP address, source port, or destination port. The parameter alone cannot be used as a hash key either.

Displaying and Maintaining Link Aggregation

To do...	Use the command...	Remarks
Display the local system ID	**display lacp system-id**	Available in any view
Display the aggregation group-specific load sharing mode	**display link-aggregation load-sharing mode**	Available in any view
Display link aggregation details of ports	**display link-aggregation member-port** [*interface-type interface-number* [**to** *interface-type interface-number*]]	Available in any view
Display the summary information of all aggregation groups	**display link-aggregation summary**	Available in any view
Display detailed information of aggregation groups	**display link-aggregation verbose** [**bridge-aggregation** [*interface-number*]]	Available in any view
Clear the LACP statistics of ports	**reset lacp statistics** [**interface** *interface-type interface-number* [**to** *interface-type interface-number*]]	Available in user view

Link Aggregation Configuration Examples

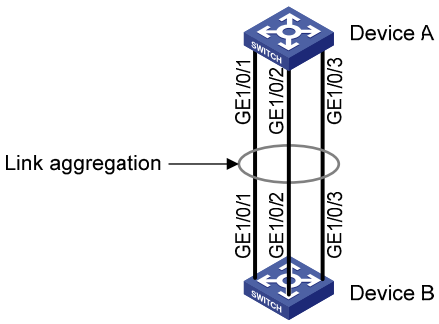
Layer 2 Static Aggregation Configuration Example

Network requirements

As shown in Figure 1-1, Device A and Device B are connected through their respective Ethernet ports GigabitEthernet1/0/1 to GigabitEthernet1/0/3.

Aggregate the ports on each device to form a static link aggregation group, thus balancing outgoing traffic across the member ports. In addition, perform load sharing based on source and destination MAC addresses.

Figure 1-1 Network diagram for Layer 2 static aggregation



Configuration procedure

1) Configure Device A

Configure the device to perform load sharing based on source and destination MAC addresses for link aggregation groups.

```
<DeviceA> system-view
[DeviceA] link-aggregation load-sharing mode source-mac destination-mac
```

Create Layer 2 aggregate interface Bridge-aggregation 1.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] quit
```

Assign Layer 2 Ethernet interfaces GigabitEthernet1/0/1 through GigabitEthernet1/0/3 to aggregation group 1.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface GigabitEthernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] interface GigabitEthernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
```

2) Configure Device B

Follow the same configuration procedure performed on Device A to configure Device B.

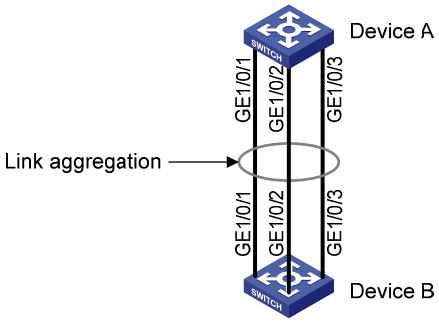
Layer 2 Dynamic Aggregation Configuration Example

Network requirements

As shown in Figure 1-2, Device A and Device B are connected through their respective Ethernet ports GigabitEthernet1/0/1 to GigabitEthernet1/0/3.

Aggregate the ports on each device to form a dynamic link aggregation group, thus balancing outgoing traffic across the member ports. In addition, perform load sharing based on source and destination MAC addresses.

Figure 1-2 Network diagram for Layer 2 dynamic aggregation



Configuration procedure

1) Configure Device A

Configure the device to perform load sharing based on source and destination MAC addresses for link aggregation groups.

```
<DeviceA> system-view
[DeviceA] link-aggregation load-sharing mode source-mac destination-mac
```

Create a Layer 2 aggregate interface Bridge-Aggregation 1 and configure the interface to work in dynamic aggregation mode.

```
[DeviceA] interface bridge-aggregation 1
[DeviceA-Bridge-Aggregation1] link-aggregation mode dynamic
[DeviceA-Bridge-Aggregation1] quit
```

Assign Layer 2 Ethernet interfaces GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to aggregation group 1.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-aggregation group 1
[DeviceA- GigabitEthernet1/0/1] quit
[DeviceA] interface GigabitEthernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-aggregation group 1
[DeviceA-GigabitEthernet1/0/2] quit
[DeviceA] interface GigabitEthernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] port link-aggregation group 1
```

2) Configure Device B

Follow the same configuration procedure performed on Device A to configure Device B.

Table of Contents

1 Port Isolation Configuration

When configuring port isolation, go to these sections for information you are interested in:

Introduction to Port Isolation

Usually, Layer 2 traffic isolation is achieved by assigning ports to different VLANs. To save VLAN resources, port isolation is introduced to isolate ports within a VLAN, allowing for great flexibility and security.

Currently:

- Some devices support only one isolation group that is created automatically by the system as isolation group 1. These devices are referred to as single-isolation-group devices. You can neither remove the isolation group nor create other isolation groups on such devices.
- There is no restriction on the number of ports assigned to an isolation group.

Configuring the Isolation Group for a Single-Isolation-Group Device

Assigning a Port to the Isolation Group

Follow these steps to add a port to the isolation group:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or, port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use one of the commands. • In Ethernet interface view, the subsequent configurations apply to the current port. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports. • In port group view, the subsequent configurations apply to all ports in the port group.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Assign the port or ports to the isolation group as an isolated port or ports		**port-isolate enable**	Required No ports are added to the isolation group by default.

Displaying and Maintaining Isolation Groups

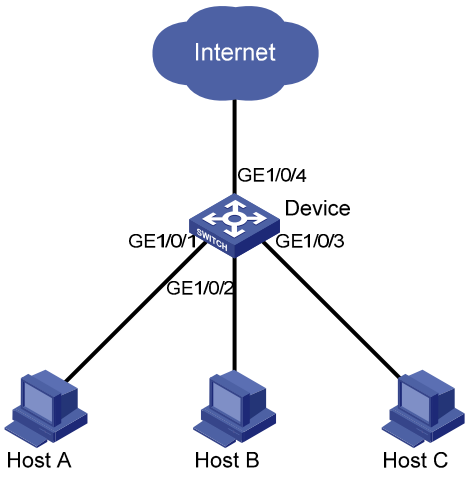
To do…	Use the command…	Remarks
Display the isolation group information on a single-isolation-group device	**display port-isolate group**	Available in any view

Port Isolation Configuration Example

Network requirements

- Users Host A, Host B, and Host C are connected to GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3 of Device.
- Device is connected to the Internet through GigabitEthernet 1/0/4.
- GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, GigabitEthernet1/0/3 and GigabitEthernet1/0/4 belong to the same VLAN. It is desired that Host A, Host B, and Host C cannot communicate with one another at Layer 2, but can access the Internet.

Figure 1-1 Networking diagram for port isolation configuration



Configuration procedure

Add ports GigabitEthernet 1/0/1, GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to the isolation group.

```
<Device> system-view
[Device] interface GigabitEthernet 1/0/1
[Device-GigabitEthernet1/0/1] port-isolate enable
[Device-GigabitEthernet1/0/1] quit
[Device] interface GigabitEthernet 1/0/2
[Device-GigabitEthernet1/0/2] port-isolate enable
[Device-GigabitEthernet1/0/2] quit
[Device] interface GigabitEthernet 1/0/3
[Device-GigabitEthernet1/0/3] port-isolate enable
```

Display the information about the isolation group.

```
<Device> display port-isolate group
```

```
Port-isolate group information:
Uplink port support: NO
Group ID: 1
Group members:
  GigabitEthernet1/0/1     GigabitEthernet1/0/2     GigabitEthernet1/0/3
```

Table of Contents

1 Service Loopback Group Configuration

When configuring a service loopback group, go to these sections for information you are interested in:

Overview

Caution

The SFP+ subcards and GE subcards of the 3Com Switch 4800G do not support service loopback groups.

Functions of Service Loopback Groups

To increase service redirecting throughput, you can bundle multiple service loopback ports into a logical link, called a service loopback group. Similar to link aggregation, a service loopback group can increase bandwidth and implement load sharing.

Service loopback groups fall into five types:

- IPv6, supporting IPv6 unicast traffic
- IPv6mc, supporting IPv6 multicast traffic
- Tunnel, supporting unicast tunnel traffic
- Multicast tunnel, supporting multicast tunnel traffic
- MPLS, supporting MPLS traffic

Note

Currently, the 3Com Switch 4800G only support the tunnel service.

Port Configuration Prerequisites of Service Loopback Groups

Before assigning a port to a service loopback group, ensure that:

- The port supports the services type or types of the service loopback group.
- The port is configured with only physical configurations such as rate and duplex mode, QoS and ACL configurations.

- The port is not configured with MSTP, 802.1x, MAC address authentication, port security mode, packet filtering, Ethernet frame filtering, or IP source guard. Additionally, the member port of a service loopback group cannot be configured with any of the above-mentioned configurations.
- The port belongs to VLAN 1.
- The port is not a member of any aggregation group or service loopback group.

States of the Ports in a Service Loopback Group

A member port in a service loopback group can be in one of the following two states:

- Selected: a selected port can forward user traffic.
- Unselected: an unselected port cannot forward user traffic.

The system sets the state of each port in a service loopback group to selected or unselected as follows:

- Select the full-duplex port with the highest rate as the reference port. If two ports with the same duplex mode/speed pair are present, the one with the lower port number wins out.
- Consider the ports the same as the reference port in rate, duplex mode, and hardware restrictions as candidate selected ports, and set the rest ports to unselected state.
- The number of selected ports is limited in a service loopback group. If the number of candidate ports exceeds the limit, those with smaller port IDs are set to selected state and the others are set to unselected state.

Note

The system follows the preemption principle when setting port state in a service loopback group. If the port you are assigning to a service loopback group can be set to selected state, the system will do that, even if this can cause an existing selected port to transit to unselected.

Configuring a Service Loopback Group

Follow these steps to configure a service loopback group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a service loopback group	**service-loopback group** *number* **type tunnel**	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Repeat the two steps to assign multiple Ethernet interfaces to the service loopback group.
Assign the Ethernet interface to the specified service loopback group	**port service-loopback group** *number*	

Caution

- You can change the service type of an existing service loopback group. For the change to be successful, you must ensure that the service group has not been referenced; the attributes of all member ports (if any) are not conflicting with the target service type; and no service loopback group has been created for the target service type, because only one service loopback group is allowed for a service type.
- You can remove any service loopback group except the referenced ones.

Displaying and Maintaining Service Loopback Groups

To do…	Use the command…	Remarks
Display information about the specified service loopback group or all service loopback groups	**display service-loopback group** [*number*]	Available in any view

Configuration Example

Network requirements

Ports of Device A support the tunnel service. Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to a service loopback group to increase bandwidth and achieve load sharing.

Configuration procedure

Create service loopback group 1 for the tunnel service.

```
<DeviceA> system-view
[DeviceA] service-loopback group 1 type tunnel
```

Assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to service loopback group 1.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] port service-loopback group 1
[DeviceA-GigabitEthernet1/0/1] interface GigabitEthernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port service-loopback group 1
[DeviceA-GigabitEthernet1/0/2] interface GigabitEthernet 1/0/3
[DeviceA-GigabitEthernet1/0/3] undo stp enable
[DeviceA-GigabitEthernet1/0/3] port service-loopback group 1
[DeviceA-GigabitEthernet1/0/3] quit
```

Create a logical interface Tunnel 1/0/0 and reference service loopback group 1 on Tunnel 1.

```
[DeviceA] interface tunnel 1/0/0
[DeviceA-Tunnel1/0/0] service-loopback-group 1
```

Table of Contents

1 DLDP Configuration

When performing DLDP configuration, go to these sections for information you are interested in:

Overview

Sometimes, unidirectional links may appear in networks. On a unidirectional link, one end can receive packets from the other end but the other end cannot. Unidirectional links result in problems such as loops in an STP-enabled network.

As for fiber links, two kinds of unidirectional links exist. One occurs when fibers are cross-connected, as shown in Figure 1-1. The other occurs when one end of a fiber is not connected or one fiber of a fiber pair gets disconnected, as illustrated by the hollow arrows in Figure 1-2.

Figure 1-1 Unidirectional fiber link: cross-connected fibers

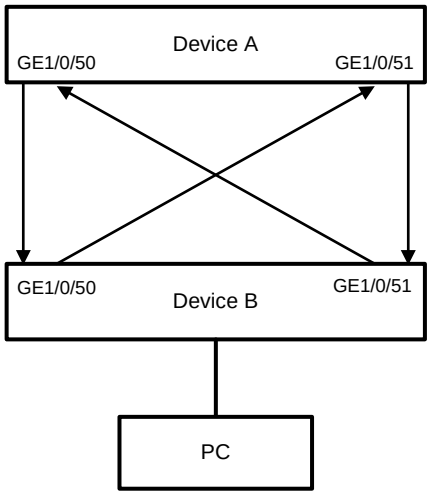
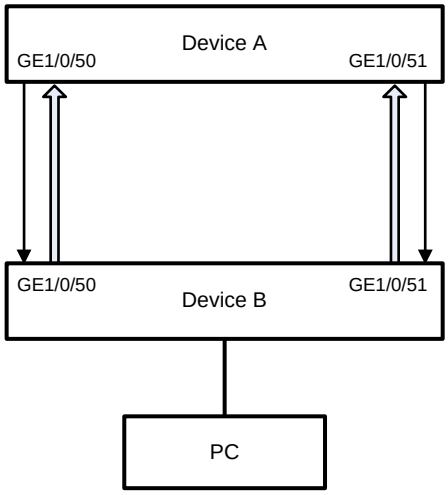


Figure 1-2 Unidirectional fiber link: a fiber not connected or disconnected



DLDP Introduction

Device Link Detection Protocol (DLDP) can detect the link status of a fiber cable or twisted pair. On detecting a unidirectional link, DLDP can shut down the related port automatically or prompt users to take measures as configured to avoid network problems.

As a data link layer protocol, DLDP cooperates with physical layer protocols to monitor the link status of a device. While the auto-negotiation mechanism provided by the physical layer detects physical signals and faults, DLDP performs operations such as identifying peer devices, detecting unidirectional links, and shutting down unreachable ports. The cooperation of physical layer protocols and DLDP ensures that physical/logical unidirectional links be detected and shut down. For a link with the devices on the both sides of it operating properly, DLDP checks to see if the cable is connected correctly and if packets can be exchanged between the two devices. Note that DLDP is not implemented through auto-negotiation.

DLDP Fundamentals

DLDP link states

A device is in one of these DLDP link states: Initial, Inactive, Active, Advertisement, Probe, Disable, and DelayDown, as described in Table 1-1.

Table 1-1 DLDP link states

State	Indicates…
Initial	DLDP is disabled.
Inactive	DLDP is enabled but the link is down.
Active	DLDP is enabled and the link is up, or the neighbor entries have been cleared.
Advertisement	All neighbors are bi-directionally reachable or DLDP has been in active state for more than five seconds. This is a relatively state where no unidirectional link has been detected.
Probe	DLDP enters this state if it receives a packet from an unknown neighbor. In this state, DLDP sends packets to check whether the link is unidirectional. As soon as DLDP transits to this state, a probe timer starts and an echo timeout timer starts for each neighbor to be probed.

State	Indicates…
Disable	A port enters this state when: • A unidirectional link is detected. • The contact with the neighbor in enhanced mode gets lost. In this state, the port does not receive or send packets other than DLDPDUs.
DelayDown	A port in the Active, Advertisement, or Probe DLDP link state transits to this state rather than removes the corresponding neighbor entry and transits to the Inactive state when it detects a port-down event. When a port transits to this state, the DelayDown timer is triggered.

DLDP timers

Table 1-2 DLDP timers

DLDP timer	Description
Active timer	Determines the Interval for sending Advertisement packets with RSY tags, which defaults to 1 second. That is, a device in the active DLDP link state sends one Advertisement packet with RSY tags every second by default. The maximum number of advertisement packets with RSY tags that can be sent successively is 5.
Advertisement timer	Determines the interval to send advertisement packets, which defaults to 5 seconds.
Probe timer	Determines the interval to send Probe packets, which defaults to 0.5 seconds. That is, a device in the probe state sends two Probe packets every second by default. The maximum number of Probe packets that can be sent successively is 10.
Echo timer	This timer is set to 10 seconds and is triggered when a device transits to the Probe state or an enhanced detect is launched. When the Echo timer expires and no Echo packet has been received from a neighbor device, the state of the link is set to unidirectional and the device transits to the Disable state. In this case, the device sends Disable packets, prompts the user to shut down the port or shuts down the port automatically (depending on the DLDP down mode configured), and removes the corresponding neighbor entries.
Entry timer	When a new neighbor joins, a neighbor entry is created and the corresponding entry timer is triggered. When a DLDP packet is received, the device updates the corresponding neighbor entry and the entry aging timer. In the normal mode, if no packet is received from a neighbor when the corresponding entry aging timer expires, DLDP sends advertisement packets with RSY tags and removes the neighbor entry. In the enhanced mode, if no packet is received from a neighbor when the Entry timer expires, DLDP triggers the enhanced timer. The setting of an Entry timer is three times that of the Advertisement timer.
Enhanced timer	In the enhanced mode, this timer is triggered if no packet is received from a neighbor when the entry aging timer expires. Enhanced timer is set to 1 second. After the Enhanced timer is triggered, the device sends up to eight probe packets to the neighbor at a frequency of one packet per second.

DLDP timer	Description
DelayDown timer	A device in the Active, Advertisement, or Probe DLDP link state transits to DelayDown state rather than removes the corresponding neighbor entry and transits to the Inactive state when it detects a port-down event. When a device transits to this state, the DelayDown timer is triggered. A device in DelayDown state only responds to port-up events. A device in the DelayDown state resumes its original DLDP state if it detects a port-up event before the DelayDown timer expires. Otherwise, it removes the corresponding DLDP neighbor information and transits to the Inactive state.
RecoverProbe timer	This timer is set to 2 seconds. That is, a port in the Disable state sends one RecoverProbe packet every two seconds to detect whether a unidirectional link has restored.

DLDP mode

DLDP can operate in two modes: normal mode and enhanced mode, as described below.

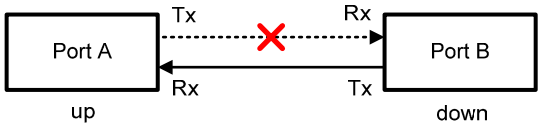
- In normal DLDP mode, when an entry timer expires, the device removes the corresponding neighbor entry and sends an Advertisement packet with RSY tag.
- In enhanced DLDP mode, when an entry timer expires, the Enhanced timer is triggered and the device sends up to eight Probe packets at a frequency of one packet per second to test the neighbor. If no Echo packet is received from the neighbor when the Echo timer expires, the device transits to the Disable state.

Table 1-3 DLDP mode and neighbor entry aging

DLDP mode	Detecting a neighbor after the corresponding neighbor entry ages out	Removing the neighbor entry immediately after the Entry timer expires	Triggering the Enhanced timer after an Entry timer expires
Normal DLDP mode	No	Yes	No
Enhanced DLDP mode	Yes	No	Yes

The enhanced DLDP mode is designed for addressing black holes. It prevents the cases where one end of a link is up and the other is down. If you configure the speed and the duplex mode by force on a device, the situation shown in Figure 1-3 may occur, where Port B is actually down but the state of Port B cannot be detected by common data link protocols, so Port A is still up. In enhanced DLDP mode, however, Port A tests Port B after the Entry timer concerning Port B expires. Port A then transits to the Disable state if it receives no Echo packet from Port A when the Echo timer expires. As Port B is physically down, it is in the Inactive DLDP state.

Figure 1-3 A case for Enhanced DLDP mode



Note

- In normal DLDP mode, only fiber cross-connected unidirectional links (as shown in Figure 1-1) can be detected.
- In enhanced DLDP mode, two types of unidirectional links can be detected. One is fiber cross-connected links (as shown in Figure 1-1). The other refers to fiber pairs with one fiber not connected or disconnected (as shown in Figure 1-2). To detect unidirectional links that are of the latter type, you need to configure the ports to operate at specific speed and in full duplex mode. Otherwise, DLDP cannot take effect. When a fiber of a fiber pair is not connected or gets disconnected, the port that can receive optical signals is in Disable state; the other port is in Inactive state.

DLDP authentication mode

You can prevent network attacks and illegal detect through DLDP authentication. Three DLDP authentication modes exist, as described below.

- Non-authentication. In this mode, the sending side sets the Authentication field and the Authentication type field of DLDP packets to 0. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.
- Plain text authentication. In this mode, before sending a DLDP packet, the sending side sets the Authentication field to the password configured in plain text and sets the Authentication type field to 1. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.
- MD5 authentication. In this mode, before sending a packet, the sending side encrypts the user configured password using MD5 algorithm, assigns the digest to the Authentication field, and sets the Authentication type field to 2. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.

DLDP implementation

1) On a DLDP-enabled link that is in up state, DLDP sends DLDP packets to the peer device and processes the DLDP packets received from the peer device. DLDP packets sent vary with DLDP states. Table 1-4 lists DLDP states and the corresponding packets.

Table 1-4 DLDP packet types and DLDP states

DLDP state	Type of DLDP packets sent
Active	Advertisement packet with RSY tag
Advertisement	Normal Advertisement packet
Probe	Probe packet
Disable	Disable packet and RecoverProbe packet

Note

When a device transits from a DLDP state other than Inactive state or Disable state to Initial state, it sends Flush packets.

2) A received DLDP packet is processed as follows.

- In any of the three authentication modes, the packet is dropped if it fails to pass the authentication.
- The packet is dropped if the setting of the interval for sending Advertisement packets it carries conflicts with the corresponding local setting.
- Other processes.

Table 1-5 Procedures for processing different types of DLDP packets

Packet type	Processing procedure	
Advertisement packet with RSY tag	Retrieving the neighbor information.	If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state.
		If the corresponding neighbor entry already exists, resets the Entry timer and transits to Probe state.
Normal Advertisement packet	Retrieves the neighbor information.	If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state.
		If the corresponding neighbor entry already exists, resets the Entry timer.
Flush packet	Determines whether or not the local port is in Disable state.	If yes, no process is performed.
		If not, removes the corresponding neighbor entry (if any).
Probe packet	Retrieves the neighbor information.	If the corresponding neighbor entry does not exist, creates the neighbor entry, transits to Probe state, and returns Echo packets.
		If the corresponding neighbor entry already exists, resets the Entry timer and returns Echo packets.

<table>
<tr><th>Packet type</th><th colspan="3">Processing procedure</th></tr>
<tr><td rowspan="3">Echo packet</td><td rowspan="3">Retrieves the neighbor information.</td><td colspan="2">If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state.</td></tr>
<tr><td rowspan="2">The corresponding neighbor entry already exists</td><td>If the neighbor information it carries conflicts with the corresponding locally maintained neighbor entry, drops the packet.</td></tr>
<tr><td>Otherwise, sets the flag of the neighbor as two-way connected. In addition, if the flags of all the neighbors are two-way connected, the device transits from Probe state to Advertisement state and disables the Echo timer.</td></tr>
<tr><td rowspan="2">Disable packet</td><td rowspan="2">Check to see if the local port is in Disable state.</td><td colspan="2">If yes, no process is performed.</td></tr>
<tr><td colspan="2">If not, the local port transits to Disable state.</td></tr>
<tr><td rowspan="2">RecoverProbe packet</td><td rowspan="2">Check to see if the local port is in Disable or Advertisement state.</td><td colspan="2">If not, no process is performed.</td></tr>
<tr><td colspan="2">If yes, returns RecoverEcho packets.</td></tr>
<tr><td rowspan="2">RecoverEcho packet</td><td rowspan="2">Check to see if the local port is in Disable state.</td><td colspan="2">If not, no process is performed.</td></tr>
<tr><td colspan="2">If yes, the local port transits to Active state if the neighbor information the packet carries is consistent with the local port information.</td></tr>
<tr><td rowspan="2">LinkDown packet</td><td rowspan="2">Check to see if the local port operates in Enhanced mode.</td><td colspan="2">If not, no process is performed.</td></tr>
<tr><td colspan="2">If yes and the local port is not in Disable state, the local transits to Disable state.</td></tr>
</table>

3) If no echo packet is received from the neighbor, DLDP performs the following processing.

Table 1-6 Processing procedure when no echo packet is received from the neighbor

No echo packet received from the neighbor	Processing procedure
In normal mode, no echo packet is received when the Echo timer expires.	DLDP transits to the Disable state, outputs log and tracking information, and sends Disable packets. In addition, depending on the user-defined DLDP down mode, DLDP shuts down the local port or prompts users to shut down the port, and removes the corresponding neighbor entry.
In enhanced mode, no echo packet is received when the enhanced timer expires.	

Link auto-recovery mechanism

If the port shutdown mode upon detection of a unidirectional link is set to **auto**, DLDP sets the state of the port where a unidirectional link is detected to DLDP down automatically. A DLDP down port cannot forward service traffic or send/receive any PDUs except DLDPDUs.

On a DLDP down port, DLDP monitors the unidirectional link. Once DLDP finds out that the state of the link has restored to bidirectional, it brings up the port. The specific process is as follows:

The DLDP down port sends out a RecoverProbe packet, which carries only information about the local port, every two seconds. Upon receiving the RecoverProbe packet, the remote end returns a RecoverEcho packet. Upon receiving the RecoverEcho packet, the local port checks whether neighbor information in the RecoverEcho packet is the same as the local port information. If they are the same, the link between the local port and the neighbor is considered to have been restored to a bidirectional link, and the port will transit from Disable state to Active state and re-establish neighborship with the neighbor.

Only DLDP down ports can send and process Recover packets, including RecoverProbe packets and RecoverEcho packets. The auto-recovery mechanism does not take effect on ports manually shut down.

DLDP neighbor state

A DLDP neighbor can be in one of the three states described in Table 1-7.

Table 1-7 Description on DLDP neighbor states

DLDP neighbor state	Description
Unknown	A neighbor is in this state when it is just detected and is being probed. No information indicating the state of the neighbor is received. A neighbor is in this state only when it is being probed. It transits to Two way state or Unidirectional state after the probe operation finishes.
Two way	A neighbor is in this state after it receives response from its peer. This state indicates the link is a two-way link.
Unidirectional	A neighbor is in this state when the link connecting it is detected to be a unidirectional link. After a device transits to this state, the corresponding neighbor entries maintained on other devices are removed.

DLDP Configuration Task List

Complete the following tasks to configure DLDP:

Task	Remarks
Enabling DLDP	Required
Setting DLDP Mode	Optional
Setting the Interval for Sending Advertisement Packets	Optional
Setting the DelayDown Timer	Optional
Setting the Port Shutdown Mode	Optional
Configuring DLDP Authentication	Optional
Resetting DLDP State	Optional

Note that:

- DLDP takes effects only on Ethernet interfaces.
- DLDP can detect unidirectional links only after all links are connected. Therefore, before enabling DLDP, make sure that optical fibers or copper twisted pairs are connected.

- To ensure unidirectional links can be detected, make sure these settings are the same on the both sides: DLDP state (enabled/disabled), the interval for sending Advertisement packets, authentication mode, and password.
- Keep the interval for sending Advertisement packets adequate to enable unidirectional links to be detected in time. If the interval is too long, unidirectional links cannot be terminated in time; if the interval is too short, network traffic may increase in vain.
- DLDP does not process any link aggregation control protocol (LACP) events. The links in an aggregation group are treated individually in DLDP.
- When connecting two DLDP-enabled devices, make sure the DLDP software version ID fields of the DLDP packets exchanged between the two devices are the same. Otherwise, DLDP may operate improperly.

Enabling DLDP

Follow these steps to enable DLDP:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable DLDP globally		**dldp enable**	Required Globally disabled by default
Enter Ethernet port view or port group view	Enter Ethernet port view	**interface** *interface-type interface-number*	Either of the two is required. The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable DLDP		**dldp enable**	Required Disabled on a port by default You can perform this operation on an optical port or an electrical port.

Caution

DLDP takes effect only when it is enabled both globally and on a port.

Setting DLDP Mode

Follow these steps to set DLDP mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set DLDP mode	**dldp work-mode** { **enhance** | **normal** }	Optional Normal by default

Setting the Interval for Sending Advertisement Packets

You can set the interval for sending Advertisement packets to enable unidirectional links to be detected in time.

Follow these steps to set the interval for sending Advertisement packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the interval for sending Advertisement packets	**dldp interval** *time*	Optional 5 seconds by default The interval for sending Advertisement packets applies to all the DLDP-enabled ports.

Caution

- Set the interval for sending Advertisement packets to a value not longer than one-third of the STP convergence time. If the interval is too long, STP loops may occur before unidirectional links are torn down, and it takes a long time for the device to detect unidirectional links, thus causing more traffic forwarding errors; if the interval is too short, unnecessary Advertisement packets can be generated to consume bandwidth. Therefore, you are recommended to use the default value.
- To enable DLDP to operate properly, make sure the intervals for sending Advertisement packets on both sides of a link are the same.

Setting the DelayDown Timer

On some ports, when the Tx line fails, the port goes down and then comes up again, causing optical signal jitters on the Rx line. When a port goes down due to a Tx failure, the device transits to the DelayDown state instead of the Inactive state to prevent the corresponding neighbor entries from being removed. In the same time, the device triggers the DelayDown timer. If the port goes up before the timer expires, the device restores the original state; if the port remains down when the timer expires, the devices transits to the Inactive state.

Follow these steps to set the DelayDown timer

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the DelayDown timer	**dldp delaydown-timer** *time*	Optional 1 second by default DelayDown timer setting applies to all the DLDP-enabled ports.

Setting the Port Shutdown Mode

On detecting a unidirectional link, the ports can be shut down in one of the following two modes.

- Manual mode. This mode applies to networks with low performance, where normal links may be treated as unidirectional links. It protects service packet transmission against false unidirectional links. In this mode, DLDP only detects unidirectional links and generates log and traps. The operations to shut down unidirectional link ports are accomplished by the administrator.
- Auto mode. In this mode, when a unidirectional link is detected, DLDP transits to Disable state, generates log and traps, and set the port as DLDP Down.

Follow these steps to set port shutdown mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set port shutdown mode	**dldp unidirectional-shutdown** { **auto** | **manual** }	Optional **auto** by default

Caution

- On a port with both remote OAM loopback and DLDP enabled, if the port shutdown mode is auto mode, the port will be shut down by DLDP when it receives a packet sent by itself, causing remote OAM loopback to operate improperly. To prevent this, you need to set the port shutdown mode to auto mode.
- If the device is busy, or the CPU utilization is high, normal links may be treated as unidirectional links. In this case, you can set the port shutdown mode to manual mode to eliminate the effects caused by false unidirectional link report.

Configuring DLDP Authentication

Follow these steps to configure DLDP authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure DLDP authentication	**dldp authentication-mode** { **md5** *md5-password* | **none** | **simple** *simple-password* }	Required **none** by default

Caution

To enable DLDP to operate properly, make sure the DLDP authentication modes and the passwords of the both sides of a link are the same.

Resetting DLDP State

After DLDP detects a unidirectional link on a port, the port enters Disable state. In this case, DLDP prompts you to shut down the port manually or shuts down the port automatically depending on the

user-defined port shutdown mode. To enable the port to perform DLDP detect again, you can reset the DLDP state of the port in one of the following methods:

- If the port is shut down with the **shutdown** command manually, use the **undo shutdown** command on the port.
- If the port is shut down by DLDP automatically, use the **dldp reset** command on the port. Alternatively, you can leave the work to DLDP, which can enable the port automatically upon detecting that the link has been restored to bidirectional. For how to reset DLDP state with the **dldp reset** command, refer to Resetting DLDP State in System View and Resetting DLDP State in Port view/Port Group View.

The DLDP state that the port transits to upon the DLDP state reset operation depends on its physical state. If the port is physically down, it transits to Inactive state; if the port is physically up, it transits to Active state.

Resetting DLDP State in System View

Resetting DLDP state in system view applies to all the ports shut down by DLDP.

Follow these steps to reset DLDP in system view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Reset DLDP state	**dldp reset**	Required

Resetting DLDP State in Port view/Port Group View

Resetting DLDP state in port view or port group view applies to the current port or all the ports in the port group shut down by DLDP.

Follow these steps to reset DLDP state in port view/port group view:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet port view/port group view	Enter Ethernet port view	**interface** *interface-type interface-number*	Either is required. The configuration performed in Ethernet port view applies to the current port only; the configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Reset DLDP state		**dldp reset**	Required

Displaying and Maintaining DLDP

To do…	Use the command…	Remarks
Display the DLDP configuration of a port	**display dldp** [*interface-type interface-number*]	Available in any view
Display the statistics on DLDP packets passing through a port	**display dldp statistics** [*interface-type interface-number*]	Available in any view

To do…	Use the command…	Remarks
Clear the statistics on DLDP packets passing through a port	**reset dldp statistics** [*interface-type interface-number*]	Available in user view

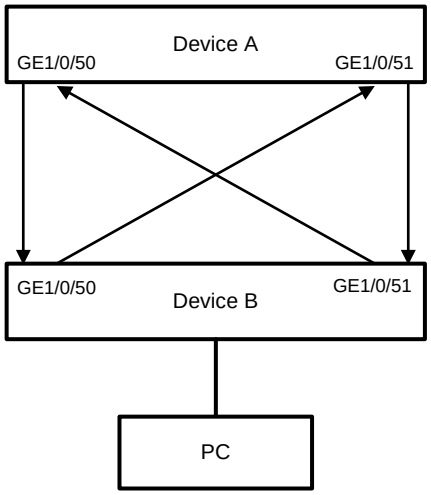
DLDP Configuration Example

DLDP Configuration Example

Network requirements

- Device A and Device B are connected through two fiber pairs, in which two fibers are cross-connected, as shown in Figure 1-4.
- It is desired that the unidirectional links can be disconnected on being detected; and the ports shut down by DLDP can be restored after the fiber connections are corrected.

Figure 1-4 Network diagram for DLDP configuration



Configuration procedure

1) Configuration on Device A

Enable DLDP on GigabitEthernet1/0/50 and GigabitEthernet 1/0/51.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/50
[DeviceA-GigabitEthernet1/0/50] dldp enable
[DeviceA-GigabitEthernet1/0/50] quit
[DeviceA] interface gigabitethernet 1/0/51
[DeviceA-GigabitEthernet1/0/51] dldp enable
[DeviceA-GigabitEthernet1/0/51] quit
```

Set the interval for sending Advertisement packets to 6 seconds.

```
[DeviceA] dldp interval 6
```

Set the DelayDown timer to 2 seconds.

```
[DeviceA] dldp delaydown-timer 2
```

Set the DLDP mode as enhanced mode.

```
[DeviceA] dldp work-mode enhance
```

Set the port shutdown mode as auto mode.

```
[DeviceA] dldp unidirectional-shutdown auto
```

Enable DLDP globally.

```
[DeviceA] dldp enable
```

Check the information about DLDP.

```
[DeviceA] display dldp
DLDP global status : enable
DLDP interval : 6s
DLDP work-mode : enhance
DLDP authentication-mode  : none
DLDP unidirectional-shutdown : auto
DLDP delaydown-timer : 2s
The number of enabled ports is 2.

Interface GigabitEthernet1/0/50
DLDP port state : disable
DLDP link state : down
The neighbor number of the port is 0.

Interface GigabitEthernet1/0/51
DLDP port state : disable
DLDP link state : down
The neighbor number of the port is 0.
```

The output information indicates that both GigabitEthernet 1/0/50 and GigabitEthernet 1/0/51 are in Disable state and the links are down, which means unidirectional links are detected and the two ports are thus shut down.

Reset DLDP state for the ports shut down by DLDP.

```
[DeviceA] dldp reset
```

2) Configuration on Device B

The configuration on Device B is the same as that on Device A and is thus omitted.

Note

If two fibers are cross-connected, all the four ports involved will be shut down by DLDP.

Troubleshooting

Symptom:

Two DLDP-enabled devices, Device A and Device B, are connected through two fiber pairs, in which two fibers are cross-connected. The unidirectional links cannot be detected; all the four ports involved are in Advertisement state.

Analysis:

The problem can be caused by the following.

- The intervals for sending Advertisement packets on Device A and Device B are not the same.
- DLDP authentication modes/passwords on Device A and Device B are not the same.

Solution:

Make sure the interval for sending Advertisement packets, the authentication mode, and the password on Device A and Device B are the same.

Table of Contents

1 LLDP Configuration

When configuring LLDP, go to these sections for information you are interested in:

Introduction to LLDP

Overview

The Link Layer Discovery Protocol (LLDP) operates on the data link layer. It stores and maintains information about the local device and the devices directly connected to it for network administrators to manage networks through NMS (network management systems). In LLDP, device information is encapsulated in LLDPDUs in the form of TLV (meaning type, length, and value) triplets and is exchanged between directly connected devices. Information in LLDPDUs received is stored in standard MIB (management information base).

LLDP Fundamental

LLDP operating mode

LLDP can operate in one of the following modes.

- TxRx mode. A port in this mode sends and receives LLDPDUs.
- Tx mode. A port in this mode only sends LLDPDUs.
- Rx mode. A port in this mode only receives LLDPDUs.
- Disable mode. A port in this mode does not send or receive LLDPDUs.

LLDP is initialized when an LLDP-enabled port changes to operate in another LLDP operating mode. To prevent LLDP from being initialized too frequently, LLDP undergoes a period before being initialized on an LLDP-enabled port when the port changes to operate in another LLDP operating mode. The period is known as initialization delay, which is determined by the re-initialization delay timer.

Sending LLDPDUs

An LLDP-enabled device operating in TxRx mode or Tx mode sends LLDPDUs to its directly connected devices periodically. It also sends LLDPDUs when the local configuration changes to inform the neighboring devices of the change timely. In any of the two cases, an interval exists between two successive operations of sending LLDPDUs. This prevents the network from being overwhelmed by LLDPDUs even if the LLDP operating mode changes frequently.

To enable the neighboring devices to be informed of the existence of a device or an LLDP operating mode change (from the disable mode to TxRx mode, or from the Rx mode to Tx mode) timely, a device can invoke the fast sending mechanism. In this case, the interval to send LLDPDUs changes to one second. After the device sends specific number of LLDPDUs, the interval restores to the normal. (A neighbor is discovered when a device receives an LLDPDU and no information about the sender is locally available.)

Receiving LLDPDUs

An LLDP-enabled device operating in TxRx mode or Rx mode checks the TLVs carried in the LLDPDUs it receives and saves the valid neighboring information. An LLDPDU also carries a TTL (time to live) setting with it. The information about a neighboring device maintained locally ages out when the corresponding TTL expires.

The TTL of the information about a neighboring device is determined by the following expression:

TTL multiplier × LLDPDU sending interval

You can set the TTL by configuring the TTL multiplier. Note that the TTL can be up to 65535 seconds. TTLs longer than it will be rounded off to 65535 seconds.

TLV Types

TLVs encapsulated in LLDPDUs fall into these categories: basic TLV, organization defined TLV, and MED (media endpoint discovery) related TLV. Basic TLVs are the base of device management. Organization specific TLVs and MED related TLVs are used for enhanced device management. They are defined in standards or by organizations and are optional to LLDPDUs.

Basic LLDP TLVs

Table 1-1 lists the basic LLDP TLV types that are currently in use.

Table 1-1 Basic LLDP TLVs

Type	Description	Remarks
End of LLDPDU TLV	Marks the end of an LLDPDU.	Required for LLDP
Chassis ID TLV	Carries the bridge MAC address of the sender	
Port ID TLV	Carries the sending port. For devices that do not send MED TLVs, port ID TLVs carry sending port name. For devices that send MED TLVs, port ID TLVs carry the MAC addresses of the sending ports or bridge MAC addresses (if the MAC addresses of the sending ports are unavailable).	
Time To Live TLV	Carries the TTL of device information	

Type	Description	Remarks
Port Description TLV	Carries Ethernet port description	Optional to LLDP
System Name TLV	Carries device name	
System Description TLV	Carries system description	
System Capabilities TLV	Carries information about system capabilities	
Management Address TLV	Carries the management address, the corresponding port number, and OID (object identifier). If the management address is not configured, it is the IP address of the interface of the VLAN with the least VLAN ID among those permitted on the port. If the IP address of the VLAN interface is not configured, IP address 127.0.0.1 is used as the management address.	

Organization defined LLDP TLVs

1) LLDP TLVs defined in IEEE802.1 include the following:

- Port VLAN ID TLV, which carries port VLAN ID.
- Port and protocol VLAN ID TLV, which carries port protocol VLAN ID.
- VLAN name TLV, which carries port VLAN name.
- Protocol identity TLV, which carries types of the supported protocols.

Note

Currently, protocol identity TLVs can only be received on the 3Com Switch 4800G.

2) IEEE 802.3 defined LLDP TLVs include the following:

- MAC/PHY configuration/status TLV, which carries port configuration, such as port speed, duplex state, whether port speed auto-negotiation is supported, the state of auto-negotiation, current speed , and current duplex state.
- Power via MDI TLV, which carries information about power supply capabilities.
- Link aggregation TLV, which carries the capability and state of link aggregation.
- Maximum frame size TLV, which carries the maximum frame size supported, namely, MTU (maximum transmission unit).

MED related LLDP TLVs

LLDP-MED TLVs provide multiple advanced applications for VoIP, such as basic configuration, network policy configuration, and address and directory management. LLDP-MED TLVs satisfy the voice device manufacturers' requirements for cost-effectiveness, easy deployment, and easy management. In addition, LLDP-MED TLVs make deploying voice devices in Ethernet easier.

- LLDP-MED capabilities TLV, which carries the MED type of the current device and the types of the LLDP MED TLVs that can be encapsulated in LLDPDUs.
- Network policy TLV, which carries port VLAN ID, supported applications (such as voice and video services), application priority, and the policy adopted.

- Extended power-via-MDI TLV, which carries the information about the power supply capability of the current device.
- Hardware revision TLV, which carries the hardware version of an MED device.
- Firmware revision TLV, which carries the firmware version of an MED device.
- Software revision TLV, which carries the software version of an MED device.
- Serial number TLV, which carries the serial number of an MED device.
- Manufacturer name TLV, which carries the manufacturer name of an MED device.
- Model name TLV, which carries the model of an MED device.
- Asset ID TLV, which carries the asset ID of an MED device. Asset ID is used for directory management and asset tracking.
- Location identification TLV, which carries the location identification of a device. Location identification can be used in location-based applications.

Note

For detailed information about LLDP TLV, refer to *IEEE 802.1AB-2005* and *ANSI/TIA-1057*.

Protocols and Standards

- IEEE 802.1AB-2005, Station and Media Access Control Connectivity Discovery
- ANSI/TIA-1057, Link Layer Discovery Protocol for Media Endpoint Devices

LLDP Configuration Task List

Complete these tasks to configure LLDP:

Task		Remarks
Basic LLDP configuration	Enabling LLDP	Required
	Setting LLDP Operating Mode	Optional
	Configuring LLDPDU TLVs	Optional
	Enable LLDP Polling	Optional
	Configuring the Parameters Concerning LLDPDU Sending	Optional
Configuring the Encapsulation Format for LLDPDUs		Optional
Configuring the Encapsulation Format of the Management Address		Optional
Configuring CDP Compatibility		Optional
Configuring LLDP Trapping		Optional

Performing Basic LLDP Configuration

Enabling LLDP

Follow these steps to enable LLDP:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable LLDP globally		**lldp enable**	Required By default, LLDP is enabled globally.
Enter Ethernet interface view/port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable LLDP		**lldp enable**	Optional By default, LLDP is enabled on a port.

Note

To make LLDP take effect, you need to enable it both globally and on the related ports.

Setting LLDP Operating Mode

Follow these steps to set LLDP operating mode:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Set the initialization delay period		**lldp timer reinit-delay** *value*	Optional 2 seconds by default.
Enter Ethernet interface view/port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Set the LLDP operating mode		**lldp admin-status** { **disable** | **rx** | **tx** | **txrx** }	Optional TxRx by default.

Configuring LLDPDU TLVs

Follow these steps to configure LLDPDU TLVs:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Set the TTL multiplier		**lldp hold-multiplier** *value*	Optional 4 by default.
Enter Ethernet interface view/port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable LLDP TLV sending for specific types of LLDP TLVs		**lldp tlv-enable** { **basic-tlv** { **all** | **port-description** | **system-capability** | **system-description** | **system-name** } | **dot1-tlv** { **all** | **port-vlan-id** | **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id*] } | **dot3-tlv** { **all** | **link-aggregation** | **mac-physic** | **max-frame-size** | **power** } | **med-tlv** { **all** | **capability** | **inventory** | **location-id** { **civic-address** *device-type country-code* { *ca-type ca-value* }*&<1–10>* | **elin-address** *tel-number* } | **network-policy** | **power-over-ethernet** } }	Optional By default, all types of LLDP TLVs except location identification TLV are sent.
Specify the management address and specify to send the management address through LLDPDUs		**lldp management-address-tlv** [*ip-address*]	Optional By default, the management address is sent through LLDPDUs, and the management address is the IP address of the interface of the VLAN with the least VLAN ID among those permitted on the port. If the IP address of the VLAN interface is not configured, IP address 127.0.0.1 is used as the management address. Refer to *VLAN Configuration* in the *Access Volume* for information about VLAN.

Note

- To enable MED related LLDP TLV sending, you need to enable LLDP-MED capabilities TLV sending first. Conversely, to disable LLDP-MED capabilities TLV sending, you need to disable the sending of other MED related LLDP TLVs.
- To disable MAC/PHY configuration/status TLV sending, you need to disable LLDP-MED capabilities TLV sending first.
- When executing the **lldp tlv-enable** command, specifying the **all** keyword for basic LLDP TLVs and organization defined LLDP TLVs (including IEEE 802.1 defined LLDP TLVs and IEEE 802.3 defined LLDP TLVs) enables sending of all the corresponding LLDP TLVs. For MED related LLDP TLVs, the **all** keyword enables sending of all the MED related LLDP TLVs except location identification TLVs.
- Enabling sending of LLDP-MED capabilities TLVs also enables sending of MAC/PHY configuration/status TLVs.

Enable LLDP Polling

With LLDP polling enabled, a device checks for the local configuration changes periodically. Upon detecting a configuration change, the device sends LLDPDUs to inform the neighboring devices of the change.

Follow these steps to enable LLDP polling:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view/port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable LLDP polling and set the polling interval		**lldp check-change-interval** *value*	Required Disabled by default

Configuring the Parameters Concerning LLDPDU Sending

Configuring time-related parameters

Follow these steps to set time-related parameters:

To do…	Use the command…	Remarks
Enter system view	**System-view**	—
Set the interval to send LLDPDUs	**lldp timer tx-interval** *value*	Optional 30 seconds by default

To do…	Use the command…	Remarks
Set the delay period to send LLDPDUs	**lldp timer tx-delay** *value*	Optional 2 seconds by default

Caution

To enable local device information to be updated on neighboring devices before being aged out, make sure the interval to send LLDPDUs is shorter than the TTL of the local device information.

Setting the number of the LLDPDUs to be sent when a new neighboring device is detected

Follow these steps to set the number of the LLDPDUs to be sent when a new neighboring device is detected

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the number of the LLDPDUs to be sent successively when a new neighboring device is detected	**lldp fast-count** *value*	Optional 3 by default

Configuring the Encapsulation Format for LLDPDUs

LLDPDUs can be encapsulated in Ethernet II or SNAP frames.

- With Ethernet II encapsulation configured, an LLDP port sends LLDPDUs in Ethernet II frames and processes only Ethernet II encapsulated incoming LLDPDUs.
- With SNAP encapsulation configured, an LLDP port sends LLDPDUs in SNAP frames and processes only SNAP encapsulated incoming LLDPDUs.

By default, LLDPDUs are encapsulated in Ethernet II frames. If the neighbor devices encapsulate LLDPDUs in SNAP frames, you can configure the encapsulation format for LLDPDUs as SNAP, thus guaranteeing communication with the other devices in the network.

Follow these steps to configure the encapsulation format for LLDPDUs:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the encapsulation format for LLDPDUs as SNAP		**lldp encapsulation snap**	Required Ethernet II encapsulation format applies by default.

Note

The configuration does not apply to LLDP-CDP packets, which use only SNAP encapsulation.

Configuring the Encapsulation Format of the Management Address

LLDP encapsulates the management address in the form of numbers or strings in management address TLVs and then advertises it.

By default, management addresses are encapsulated in the form of numbers in TLVs. If neighbors encapsulate management addresses in the form of strings in TLVs, you can configure the encapsulation format of the management address as strings, thus guaranteeing communication with the other devices in the network.

Follow these steps to configure the encapsulation format of the management address:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the encapsulation format of the management address as strings in TLVs		lldp management-address-format string	Required By default, the management address is encapsulated in the form of numbers in TLVs.

Configuring CDP Compatibility

Note

For detailed information about voice VLAN, refer to *VLAN Configuration* in the *Access Volume*.

You need to enable CDP compatibility for your device to work with Cisco IP phones.

As your LLDP-enabled device cannot recognize CDP packets, it does not respond to the requests of Cisco IP phones for the voice VLAN ID configured on the device. This can cause a requesting Cisco IP phone to send voice traffic without any tag to your device, disabling your device to differentiate the voice traffic from other types of traffic.

By configuring CDP compatibility, you can enable LLDP on your device to receive and recognize CDP packets from Cisco IP phones and respond with CDP packets carrying the voice VLAN configuration

TLV for the IP phones to configure the voice VLAN automatically. Thus, the voice traffic is confined in the configured voice VLAN to be differentiated from other types of traffic.

CDP-compatible LLDP operates in one of the follows two modes:

- TxRx where CDP packets can be transmitted and received.
- Disable where CDP packets can neither be transmitted nor be received.

Configuration Prerequisites

Before configuring CDP compatibility, make sure that:

- LLDP is enabled globally.
- LLDP is enabled on the port connected to an IP phone and is configured to operate in TxRx mode on the port.

Configuring CDP Compatibility

Follow these steps to enable LLDP to be compatible with CDP:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable CDP compatibility globally		**lldp compliance cdp**	Required Disabled by default.
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure CDP-compatible LLDP to operate in TxRx mode		**lldp compliance admin-status cdp txrx**	Required By default, CDP-compatible LLDP operates in disable mode.

Caution

As the maximum TTL allowed by CDP is 255 seconds, your TTL configuration, that is, the product of the TTL multiplier and the LLDPDU sending interval, must be less than 255 seconds for CDP-compatible LLDP to work properly with Cisco IP phones.

Configuring LLDP Trapping

LLDP trapping is used to notify NMS of the events such as new neighboring devices detected and link malfunctions.

LLDP traps are sent periodically and you can set the interval to send LLDP traps. In response to topology changes detected, a device sends LLDP traps according to the interval configured to inform the neighboring devices of the changes.

Follow these steps to configure LLDP trap:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view/port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable LLDP trap sending		**lldp notification remote-change enable**	Required Disabled by default
Quit to system view		**quit**	—
Set the interval to send LLDP traps		**lldp timer notification-interval** *value*	Optional 5 seconds by default

Displaying and Maintaining LLDP

To do…	Use the command…	Remarks
Display the global LLDP information or the information contained in the LLDP TLVs to be sent through a port	**display lldp local-information** [**global** | **interface** *interface-type interface-number*]	Available in any view
Display the information contained in the LLDP TLVs received through a port	**display lldp neighbor-information** [**interface** *interface-type interface-number*] [**brief**]	Available in any view
Display LLDP statistics	**display lldp statistics** [**global** | **interface** *interface-type interface-number*]	Available in any view
Display LLDP status of a port	**display lldp status** [**interface** *interface-type interface-number*]	Available in any view
Display the types of the LLDP TLVs that are currently sent	**display lldp tlv-config** [**interface** *interface-type interface-number*]	Available in any view

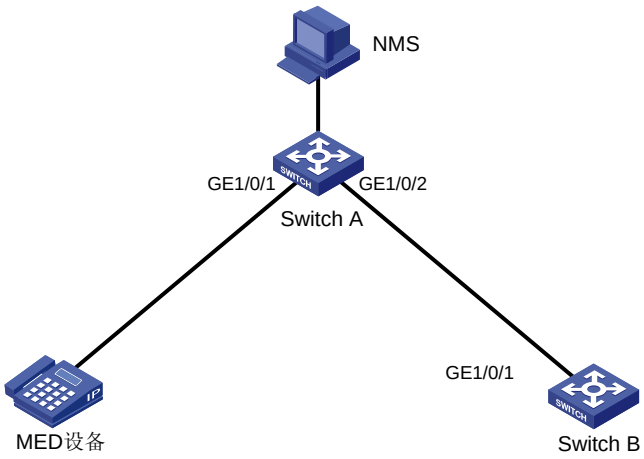
LLDP Configuration Examples

LLDP Basic Configuration Example

Network requirements

- The NMS and Switch A are located in the same Ethernet. An MED device and Switch B are connected to GigabitEthernet1/0/1 and GigabitEthernet1/0/2 of Switch A.
- Enable LLDP on the ports of Switch A and Switch B to monitor the link between Switch A and Switch B and the link between Switch A and the MED device on the NMS.

Figure 1-1 Network diagram for LLDP configuration



Configuration procedure

1) Configure Switch A.

Enable LLDP globally.

```
<SwitchA> system-view
[SwitchA] lldp enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2, setting the LLDP operating mode to Rx.

```
[SwitchA] interface gigabitethernet1/0/1
[SwitchA-GigabitEthernet1/0/1] lldp enable
[SwitchA-GigabitEthernet1/0/1] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet1/0/2
[SwitchA-GigabitEthernet1/0/2] lldp enable
[SwitchA-GigabitEthernet1/0/2] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/2] quit
```

2) Configure Switch B.

Enable LLDP globally.

```
<SwitchB> system-view
[SwitchB] lldp enable
```

Enable LLDP on GigabitEthernet1/0/1, setting the LLDP operating mode to Tx.

```
[SwitchB] interface gigabitethernet1/0/1
[SwitchB-GigabitEthernet1/0/1] lldp enable
[SwitchB-GigabitEthernet1/0/1] lldp admin-status tx
[SwitchB-GigabitEthernet1/0/1] quit
```

3) Verify the configuration.

Display the global LLDP status and port LLDP status on Switch A.

```
[SwitchA] display lldp status
Global status of LLDP : Enable
The current number of LLDP neighbors : 2
The current number of CDP neighbors : 0
LLDP neighbor information last changed time : 0 days, 0 hours, 4 minutes, 40 seconds
```

```
Transmit interval              : 30s
Hold multiplier                : 4
Reinit delay                   : 2s
Transmit delay                 : 2s
Trap interval                  : 5s
Fast start times               : 3

Port 1 [GigabitEthernet1/0/1] :
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
Roll time                      : 0s

Number of neighbors             : 1
Number of MED neighbors         : 1
Number of CDP neighbors         : 0
Number of sent optional TLV     : 0
Number of received unknown TLV  : 0

Port 2 [GigabitEthernet1/0/2] :
Port status of LLDP            : Enable
Admin status                   : Rx_Only
Trap flag                      : No
Roll time                      : 0s

Number of neighbors             : 1
Number of MED neighbors         : 0
Number of CDP neighbors         : 0
Number of sent optional TLV     : 0
Number of received unknown TLV  : 3
```

Tear down the link between Switch A and Switch B and then display the global LLDP status and port LLDP status on Switch A.

```
[SwitchA] display lldp status
Global status of LLDP : Enable
The current number of LLDP neighbors : 1
The current number of CDP neighbors : 0
LLDP neighbor information last changed time : 0 days, 0 hours, 5 minutes, 20 seconds
Transmit interval              : 30s
Hold multiplier                : 4
Reinit delay                   : 2s
Transmit delay                 : 2s
Trap interval                  : 5s
Fast start times               : 3

Port 1 [GigabitEthernet1/0/1] :
Port status of LLDP            : Enable
Admin status                   : Rx_Only
```

```
Trap flag                       : No
Roll time                       : 0s

Number of neighbors              : 1
Number of MED neighbors          : 1
Number of CDP neighbors          : 0
Number of sent optional TLV      : 0
Number of received unknown TLV   : 5

Port 2 [GigabitEthernet1/0/2] :
Port status of LLDP             : Enable
Admin status                    : Rx_Only
Trap flag                       : No
Roll time                       : 0s

Number of neighbors              : 0
Number of MED neighbors          : 0
Number of CDP neighbors          : 0
Number of sent optional TLV      : 0
Number of received unknown TLV   : 0
```

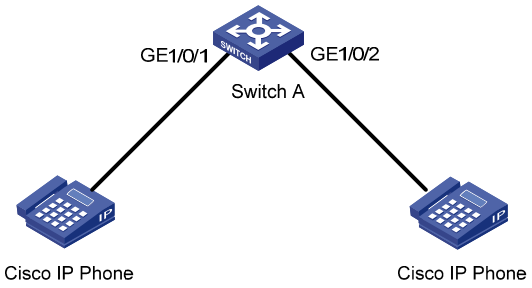
CDP-Compatible LLDP Configuration Example

Network requirements

- GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of Switch A are each connected to a Cisco IP phone.
- Configure voice VLAN 2 on Switch A. Enable CDP compatibility of LLDP on Switch A to allow the Cisco IP phones to automatically configure the voice VLAN, thus confining their voice traffic within the voice VLAN to be isolated from other types of traffic.

Network diagram

Figure 1-2 Network diagram for LLDP compatible with CDP configuration



Configuration procedure

1) Configure the voice VLAN on Switch A

Create VLAN 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] quit
```

Configure the link type of the ports to be trunk and enable the voice VLAN feature on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-type trunk
[SwitchA-GigabitEthernet1/0/1] voice vlan 2 enable
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-type trunk
[SwitchA-GigabitEthernet1/0/2] voice vlan 2 enable
[SwitchA-GigabitEthernet1/0/2] quit
```

2) Configure CDP-compatible LLDP on Switch A.

Enable LLDP globally.

```
[SwitchA] lldp enable
```

Enable LLDP to be compatible with CDP globally.

```
[SwitchA] lldp compliance cdp
```

Enable LLDP, configure LLDP to operate in TxRx mode, and configure CDP-compatible LLDP to operate in TxRx mode on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] lldp enable
[SwitchA-GigabitEthernet1/0/1] lldp admin-status txrx
[SwitchA-GigabitEthernet1/0/1] lldp compliance admin-status cdp txrx
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] lldp enable
[SwitchA-GigabitEthernet1/0/2] lldp admin-status txrx
[SwitchA-GigabitEthernet1/0/2] lldp compliance admin-status cdp txrx
[SwitchA-GigabitEthernet1/0/2] quit
```

3) Verify the configuration

Display the neighbor information on Switch A.

```
[SwitchA] display lldp neighbor-information
CDP neighbor-information of port 1[GigabitEthernet1/0/1]:
  CDP neighbor index : 1
  Chassis ID         : SEP00141CBCDBFE
  Port ID            : Port 1
  Sofrware version   : P0030301MFG2
  Platform           : Cisco IP Phone 7960
  Duplex             : Full

CDP neighbor-information of port 2[GigabitEthernet1/0/2]:
  CDP neighbor index : 2
  Chassis ID         : SEP00141CBCDBFF
  Port ID            : Port 1
  Sofrware version   : P0030301MFG2
  Platform           : Cisco IP Phone 7960
  Duplex             : Full
```

Table of Contents

1 Smart Link Configuration

When configuring Smart Link, go to these sections for information that you are interested in:

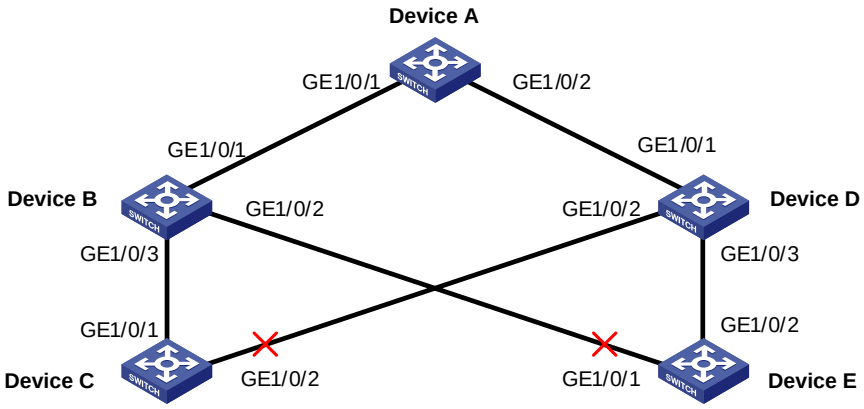
Smart Link Overview

Smart Link is a feature developed to address the slow convergence issue with the Spanning Tree Protocol (STP). (For information about STP, refer to *MSTP Configuration* in the *Access Volume*.)

Smart Link is dedicated to dual-uplink networks as shown in Figure 1-1 to provide link redundancy with subsecond convergence. It allows the backup link to take over quickly when the primary link fails. In addition to fast convergence, Smart Link is easy to configure.

Terminology

Figure 1-1 Smart Link application scenario



Smart link group

A smart link group consists of only two member ports: the master and the slave. At a time, only one port is active for forwarding, and the other port is blocked, that is, in the standby state. When link failure occurs on the active port due to port shutdown or presence of unidirectional link for example, the standby port becomes active to take over while the original active port transits to the blocked state.

As shown in Figure 1-1, GE1/0/1 and GE1/0/2 of Device C form a smart link group, with GE1/0/1 being active and GE1/0/2 being standby. GE1/0/1 and GE1/0/2 of Device E form another smart link group, with GE1/0/2 being active and GE1/0/1 being standby.

Master port

Master port is a port role in a smart link group. When both ports in a smart link group are up, the master port preferentially transits to the forwarding state. Once the master port fails, the slave port takes over to forward traffic. During this period, if the smart link group is not configured with role preemption, the master port stays in standby state until the next link switchover even if it has recovered.

As shown in Figure 1-1, you can configure GE1/0/1 of Device C and GE1/0/2 of Device E as master ports.

Slave port

Slave port is a port role in a smart link group. When both ports in a smart link group are up, the slave port is placed in the standby state. When the master port fails, the slave port takes over to forward traffic.

As shown in Figure 1-1, you can configure GE1/0/2 of Device C and GE1/0/1 of Device E as slave ports.

Flush message

Flush messages are used by a smart link group to notify other devices to refresh their MAC address forwarding entries and ARP/ND entries when link switchover occurs in the smart link group. Flush messages are common multicast data packets, and will be dropped by a blocked receiving port.

Transmit control VLAN

The transmit control VLAN is used for transmitting flush messages. When link switchover occurs, the devices (such as Device C and E in Figure 1-1) broadcast flush messages within the transmit control VLAN.

Receive control VLAN

The receive control VLAN is used for receiving and processing flush messages. When link switchover occurs, the devices (such as Device A, B, and D in Figure 1-1) receive and process flush messages in the receive control VLAN and refresh their MAC address forwarding entries and ARP/ND entries.

Protected VLAN

A smart link group controls the forwarding state of some data VLANs, which are referred to as protected VLANs. Different smart link groups on a port control different protected VLANs. The state of the port in a protected VLAN is determined by the state of the port in the smart link group.

Operating Mechanism of Smart Link

Link backup mechanism

As shown in Figure 1-1, the link on GE1/0/1 of Device C is the active link, and the link on GE1/0/2 of Device C is the standby link. Normally, GE1/0/1 is in the forwarding state, while GE1/0/2 is in the standby state. When the link on GE1/0/1 fails, GE1/0/2 takes over to forward traffic while GE1/0/1 is blocked and placed in the standby state.

When a port switches to the forwarding state, the system outputs log information to notify the user of the port state change.

As link switchover can outdate the MAC address forwarding entries and ARP/ND entries on all devices, you need a forwarding entry update mechanism to ensure proper transmission. By far, the following two update mechanisms are provided:

- Uplink traffic-triggered MAC address learning, where update is triggered by uplink traffic. This mechanism is applicable to environments with devices not supporting smart link, including devices of other vendors'.
- Flush update where a Smart Link-enabled device updates its information by transmitting flush messages over the backup link to its upstream devices. This mechanism requires the upstream devices to be capable of recognizing smart link flush messages to update its MAC address forwarding entries and ARP/ND entries.

To keep traffic forwarding stable, the master port that has been blocked due to link failure does not take over immediately upon its recovery. Instead, link switchover will occur at next link switchover.

Role preemption mechanism

As shown in Figure 1-1, the link on GE1/0/1 of Device C is the active link, and the link on GE1/0/2 of Device C is the standby link. Once GE1/0/1 fails, GE1/0/2 takes over to forward traffic. During this period, if the smart link group is configured with role preemption, GE1/0/1 takes over to forward traffic as soon as it recovers.

Load sharing mechanism

A ring network may carry traffic of multiple VLANs. Smart link can forward traffic of different VLANs in different smart link groups, thus implementing load sharing.

To implement load sharing, you can assign a port to multiple smart link groups (each configured with different protected VLANs), making sure that the state of the port is different in these smart link groups. In this way, traffic of different VLANs can be forwarded along different paths.

You can configure protected VLANs for a smart link group by referencing MSTIs.

Configuring a Smart Link Device

To use Smart Link on a device, you must configure the device with a smart link group and transmit control VLAN for flush message transmission. Device C and Device E in Figure 1-1 are two examples of Smart Link devices.

Configuration Prerequisites

- Before configuring a port as a smart link group member, shut down the port to prevent loops. You can bring up the port only after completing the smart link group configuration.
- Disable STP and RRPP on the ports you want to add to the smart link group, and make sure that the ports are not member ports of any aggregation group or service loopback group.

Configuring a Smart Link Device

Follow these steps to configure a smart link device:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a smart link group and enter smart link group view	**smart-link group** *group-id*	Required

To do…		Use the command…	Remarks
Configure protected VLANs for the smart link group		**protected-vlan reference-instance** *instance-id-list*	Required By default, no protected VLAN is configured for a smart link group.
Specify the master port for the smart link group	In smart link group view	**port** *interface-type interface-number* **master**	Required Use either approach.
	In Ethernet interface view or Layer-2 aggregate interface view	**port smart-link group** *group-id* **master**	
Specify the slave port for the smart link group	In smart link group view	**port** *interface-type interface-number* **slave**	Required Use either approach.
	In Ethernet interface view or Layer-2 aggregate interface view	**port smart-link group** *group-id* **slave**	
Enable role preemption		**preemption mode role**	Optional Disabled by default.
Configure the preemption delay		**preemption delay** *delay-time*	Optional 1 second by default.
Enable flush update in the specified control VLAN		**flush enable** [**control-vlan** *vlan-id*]	Optional By default, VLAN 1 is used for flush update.

Caution

- The **protected-vlan** command configures protected VLANs for a smart link group by referencing MSTIs. To view VLAN-to-MSTI mappings, use the **display stp region-configuration** command. For VLAN-to-MSTI mapping configuration, refer to *MSTP Configuration* in the *Access Volume*.
- The preemption delay configuration takes effect only after preemption mode is enabled.
- The protected VLANs configured for a smart link group must be different from those configured for any other smart link group.
- Make sure that the configured control VLANs are existing VLANs, and you must assign the smart link group member ports to the control VLANs.
- Do not remove the control VLANs. Otherwise, flush messages cannot be sent properly.

Smart Link Device Configuration Example

Network requirements

- Create smart link group 1.
- The protected VLANs of smart link group 1 are mapped to MSTI 0 through 8.
- Configure GigabitEthernet 1/0/1 as the master port of the smart link group, and GigabitEthernet 1/0/2 as the slave port.

- Configure VLAN 20 for flush update.

Configuration procedure

```
<Sysname> system-view
[Sysname] vlan 20
[Sysname-vlan20] quit
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] undo stp enable
[Sysname-GigabitEthernet1/0/1] port link-type trunk
[Sysname-GigabitEthernet1/0/1] port trunk permit vlan 20
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] undo stp enable
[Sysname-GigabitEthernet1/0/2] port link-type trunk
[Sysname-GigabitEthernet1/0/2] port trunk permit vlan 20
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] smart-link group 1
[Sysname-smlk-group1] protected-vlan reference-instance 0 to 8
[Sysname-smlk-group1] port gigabitethernet1/0/1 master
[Sysname-smlk-group1] port gigabitethernet1/0/2 slave
[Sysname-smlk-group1] flush enable control-vlan 20
```

Configuring an Associated Device

The active and standby links in a smart link group may traverse multiple devices between the Smart Link device and the destination device. For Smart Link to work, you need to enable all the ports on the way to the destination to process the flush messages sent from the smart link device.

For example, as all the numbered ports on Device A, B, and D in Figure 1-1 are on the way of the active and standby links from Device C and E to Device A, you need to enable the ports to process flush messages received from the control VLAN configured on Device C and E.

Configuring an Associated Device

Follow these steps to configure an associated device:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	—
Configure the control VLAN for receiving flush messages	**smart-link flush enable** [**control-vlan** *vlan-id-list*]	Required No control VLAN exists for receiving flush messages of Smart Link by default.

Caution

- Configure all the control VLANs to receive flush messages.
- If no control VLAN is specified for processing flush messages, the device forwards the received flush messages directly without processing them.
- Make sure that the receive control VLAN is the same as the transmit control VLAN configured on the Smart Link device. If they are not the same, the associated device will forward the received flush messages directly without any processing.
- Do not remove the control VLANs. Otherwise, flush messages cannot be sent properly.
- Make sure that the control VLANs are existing VLANs, and you must assign the port capable of receiving flush messages to the control VLANs.

Associated Device Configuration Example

Network requirements

Configure GigabitEthernet 1/0/1 to receive and process flush messages in VLAN 20.

Configuration procedure

```
<Sysname> system-view
[Sysname] vlan 20
[Sysname-vlan20] quit
[Sysname] interface gigabitethernet1/0/1
[Sysname-GigabitEthernet1/0/1] port link-type trunk
[Sysname-GigabitEthernet1/0/1] port trunk permit vlan 20
[Sysname-GigabitEthernet1/0/1] smart-link flush enable control-vlan 20
```

Displaying and Maintaining Smart Link

To do...	Use the command…	Remarks
Display smart link group information	**display smart-link group** { *group-id* | **all** }	Available in any view
Display information about the received flush messages	**display smart-link flush**	Available in any view
Clear the statistics about flush messages	**reset smart-link statistics**	Available in user view

Smart Link Configuration Examples

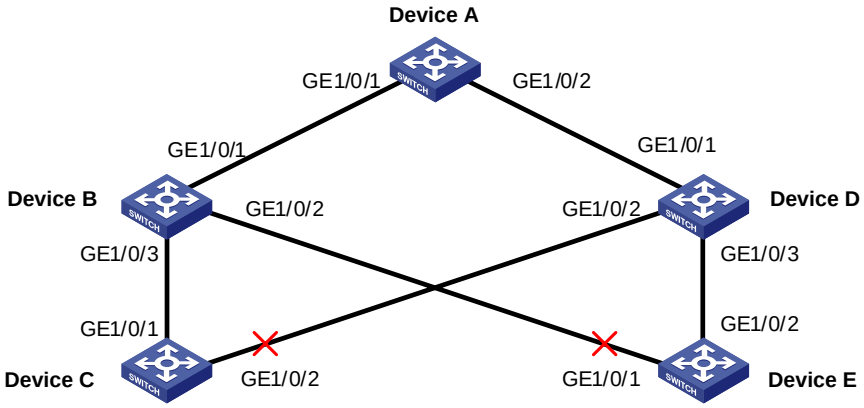
Single Smart Link Group Configuration Example

Network requirements

As shown in Figure 1-2, both Device C and Device E are dually uplinked to Device A.

Configure Smart Link on the devices for uplink backup, adopting VLAN 1 (the default) for flush update.

Figure 1-2 Network diagram for single smart link group configuration



Configuration procedure

1) Configuration on Device C

Create smart link group 1.

```
<DeviceC> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo stp enable
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] smart-link group 1
```

Configure all VLANs mapped to MSTIs 0 through 32 as the protected VLANs.

```
[DeviceC-smlk-group1] protected-vlan reference-instance 0 to 32
```

Configure GigabitEthernet 1/0/1 as the master port and GigabitEthernet 1/0/2 as the slave port.

```
[DeviceC-smlk-group1] port gigabitethernet1/0/1 master
[DeviceC-smlk-group1] port gigabitethernet1/0/2 slave
```

Configure VLAN 1 as the transmit control VLAN.

```
[DeviceC-smlk-group1] flush enable
```

2) Configuration on Device E

Create smart link group 1.

```
<DeviceE> system-view
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] undo stp enable
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] undo stp enable
[DeviceE-GigabitEthernet1/0/2] quit
[DeviceE] smart-link group 1
```

Configure all VLANs mapped to MSTIs 0 through 32 as the protected VLANs.

```
[DeviceC-smlk-group1] protected-vlan reference-instance 0 to 32
```

Configure GigabitEthernet 1/0/2 as the master port and GigabitEthernet 1/0/1 as the slave port.

```
[DeviceE-smlk-group1] port gigabitethernet1/0/2 master
[DeviceE-smlk-group1] port gigabitethernet1/0/1 slave
```

Configure VLAN 1 as the transmit control VLAN.

```
[DeviceE-smlk-group1] flush enable
```

3) Configuration on Device B

Configure VLAN 1 as the receive control VLAN for GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] smart-link flush enable
```

4) Configuration on Device D

Configure VLAN 1 as the receive control VLAN for GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3.

```
<DeviceD> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] smart-link flush enable
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] smart-link flush enable
[DeviceD-GigabitEthernet1/0/2] quit
[DeviceD] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] smart-link flush enable
```

5) Configuration on Device A

Configure VLAN 1 as the receive control VLAN for GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, and GigabitEthernet 1/0/3.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable
```

After completing the configuration, you can use the **display** command to verify the smart link configuration and view flush message statistics.

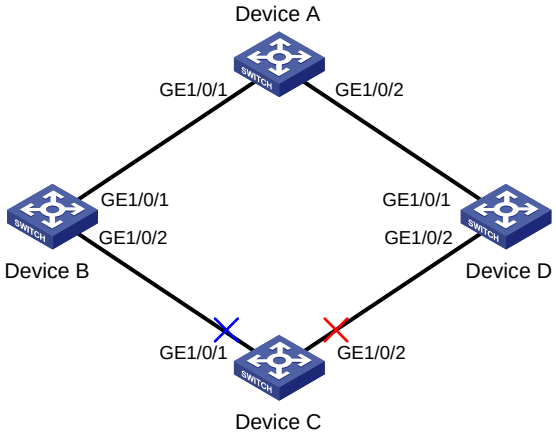
Multiple Smart Link Groups Load Sharing Configuration Example

Network requirements

As shown in Figure 1-3:

- The traffic of VLAN 1 through VLAN 200 on Device C are dually uplinked to Device A by Device B and Device D. Implement load sharing to uplink the traffic of VLAN 1 through VLAN 100 and the traffic of VLAN 101 through VLAN 200 over different links to Device A.
- Implement dual link backup on Device C: the traffic of VLANs 1 through 100 (mapped to MSTI 0) is uplinked to Device A by Device B; the traffic of VLANs 101 through 200 (mapped to MSTI 2) is uplinked to Device A by Device D. Smart link group 1 references MSTI 0, and smart link group 2 references MSTI 2.
- The control VLAN of smart link group 1 is VLAN 10 and that of smart link group 2 is VLAN 101.

Figure 1-3 Network diagram for multiple smart link groups load sharing configuration



Configuration procedure

1) Configuration on Device C

Create VLANs and configure VLAN-to-MSTI mappings.

```
<DeviceC> system-view
[DeviceC] vlan 1 to 200
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 0 vlan 1 to 100
[DeviceC-mst-region] instance 2 vlan 101 to 200
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Disable STP on the ports, configure the ports as trunk ports, and configure the ports to allow packets from VLAN 1 through 200 to pass through.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo stp enable
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceC-GigabitEthernet1/0/2] quit
```

Create smart link group 1.

```
[DeviceC] smart-link group 1
```

Configure protected VLANs for smart link group 1.

```
[DeviceC-smlk-group1] protected-vlan reference-instance 0
```

Configure GigabitEthernet 1/0/1 as the master port and GigabitEthernet 1/0/2 as the slave port.

```
[DeviceC-smlk-group1] port gigabitethernet1/0/1 master
[DeviceC-smlk-group1] port gigabitethernet1/0/2 slave
```

Enable role preemption.

```
[DeviceC-smlk-group1] preemption mode role
```

Configure VLAN 10 as the transmit control VLAN of smart link group 1.

```
[DeviceC-smlk-group-1] flush enable control-vlan 10
[DeviceC-smlk-group-1] quit
```

Create smart link group 2.

```
[DeviceC] smart-link group 2
```

Configure protected VLANs for smart link group 2.

```
[DeviceC-smlk-group2] protected-vlan reference-instance 2
```

Configure GigabitEthernet 1/0/1 as the slave port and GigabitEthernet 1/0/2 as the master port.

```
[DeviceC-smlk-group2] port gigabitethernet1/0/1 slave
[DeviceC-smlk-group2] port gigabitethernet1/0/2 master
```

Enable role preemption.

```
[DeviceC-smlk-group2] preemption mode role
```

Configure VLAN 101 as the transmit control VLAN of smart link group 2.

```
[DeviceC-smlk-group2] flush enable control-vlan 101
```

2) Configuration on Device B

Configure VLAN 10 and VLAN 101 as the receive control VLANs of GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
<DeviceB> system-view
[DeviceB] vlan 1 to 200
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 101
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 101
```

3) Configuration on Device D

Configure VLAN 10 and VLAN 101 as the receive control VLANs of GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
<DeviceD> system-view
[DeviceD] vlan 1 to 200
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
```

```
[DeviceD-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 101
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceD-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 101
```

4) Configuration on Device A

Configure VLAN 10 and VLAN 101 as the receive control VLANs of GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
<DeviceA> system-view
[DeviceA] vlan 1 to 200
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 1 to 200
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable control-vlan 10 101
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 1 to 200
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable control-vlan 10 101
```

After completing the configuration, you can use the **display** command to verify the smart link configuration and view flush message statistics.

Table of Contents

1 Monitor Link Configuration

When configuring monitor link, go to these sections for information you are interested in:

Overview

Monitor link is a port collaboration function used to enable a device to be aware of the up/down state change of the ports on an indirectly connected link. Monitor link is usually used in conjunction with Layer-2 topology protocols. The idea is to adapt the up/down state of downlink ports to the up/down state of uplink ports, triggering link switchover on the downlink device in time.

Terminology

Monitor link group

A monitor link group is a set of uplink ports and downlink ports. For the purpose of monitor link, uplink ports refer to the monitored ports while downlink ports refer to the ports adapted to the up/down state of the monitored ports. A port can be assigned to only one monitor link group. Both Layer-2 Ethernet ports and Layer-2 aggregate interfaces can be assigned to a monitor link group.

Uplink

The uplink is the link monitored by the monitor link group. The monitor link group is down when the group has no uplink ports or all uplink ports are down. The monitor link group is up when any uplink port is up.

Downlink

The downlink is the state-adaptive link in the monitor link group. The state of the downlink ports is always consistent with the up/down state of the monitor link group.

How Monitor Link Works

A monitor link group works independently of other monitor link groups. When a monitor link group contains no uplink ports or all its uplink ports go down, the monitor link group goes down and forces all downlink ports down at the same time. When any uplink port goes up, the monitor link group goes up and brings up all the downlink ports.

Caution

Do not manually shut down or bring up the downlink ports in a monitor link group.

Configuring Monitor Link

Configuration Prerequisites

Before assigning a port to a monitor link group, make sure the port is not the member port of any aggregation group or service loopback group.

Configuration Procedure

Follow these steps to configure monitor link:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a monitor link group and enter monitor link group view		**monitor-link group** *group-id*	Required
Configure the uplink for the monitor link group	In monitor link group view	**port** *interface-type interface-number* **uplink**	Use either approach Repeat this step to add more uplink ports
	In Ethernet port view or Layer-2 aggregate interface view	**port monitor-link group** *group-id* **uplink**	
Configure the downlink for the monitor link group	In monitor link group view	**port** *interface-type interface-number* **downlink**	Use either approach Repeat this step to add more downlink ports
	In Ethernet port view or Layer-2 aggregate interface view	**port monitor-link group** *group-id* **downlink**	

Caution

- A port can be assigned to only one monitor link group.
- You are recommended to configure uplink ports prior to downlink ports, thus avoiding undesired down/up state changes on the downlink ports.

Monitor Link Configuration Example

Network requirements

GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of a device are up. Configure GigabitEthernet 1/0/2 to change its up/down state as GigabitEthernet 1/0/1.

Configuration procedure

```
<Sysname> system-view
[Sysname] monitor-link group 1
[Sysname-mtlk-group1] port gigabitethernet 1/0/1 uplink
[Sysname-mtlk-group1] port gigabitethernet 1/0/2 downlink
```

Displaying and Maintaining Monitor Link

To do…	Use the command…	Remarks
Display monitor link group information	**display monitor-link group** { *group-id* **| all** }	Available in any view

Monitor Link Configuration Example

Network requirements

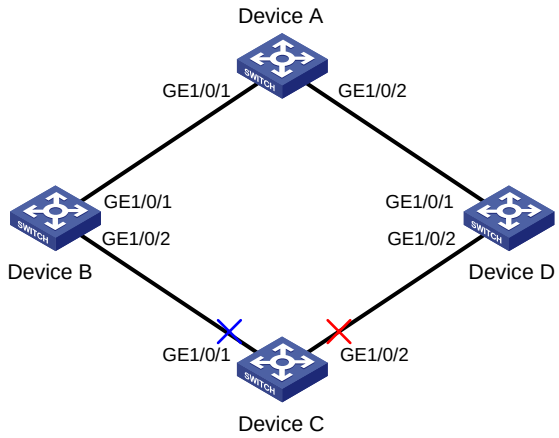
As shown in Figure 1-1:

- Device C is dually uplinked to Device A through a smart link group.
- It is required that when GigabitEthernet 1/0/1 or GigabitEthernet 1/0/2 of Device A fails, Device C can sense the link failure and perform link switchover in the smart link group.

Note

For detailed information about smart link, refer to *Smart Link Configuration* in the *Access Volume*.

Figure 1-1 Network diagram for smart link in combination with monitor link configuration



Configuration procedure

1) Configuration on Device C

Create smart link group 1.

```
<DeviceC> system-view
```

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo stp enable
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] smart-link group 1
```

Configure the smart link group to protect all the VLANs mapped to MSTIs 0 through 32.

```
[DeviceC-smlk-group1] protected-vlan reference-instance 0 to 32
```

Configure GigabitEthernet 1/0/1 as the master port and GigabitEthernet 1/0/2 as the slave port.

```
[DeviceC-smlk-group1] port gigabitethernet 1/0/1 master
[DeviceC-smlk-group1] port gigabitethernet 1/0/2 slave
```

Enable the smart link group to transmit flush messages in VLAN 1.

```
[DeviceC-smlk-group1] flush enable
```

2) Configuration on Device A

Configure VLAN 1 as the control VLAN for receiving flush messages on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] smart-link flush enable
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] smart-link flush enable
```

3) Configuration on Device B

Create monitor link group 1.

```
<DeviceB> system-view
[DeviceB] monitor-link group 1
```

Configure GigabitEthernet 1/0/1 as an uplink port and GigabitEthernet 1/0/2 as a downlink port.

```
[DeviceB-mtlk-group1] port gigabitethernet 1/0/1 uplink
[DeviceB-mtlk-group1] port gigabitethernet 1/0/2 downlink
```

Configure VLAN 1 as the control VLAN for receiving flush messages on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceB-mtlk-group-1] quit
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] smart-link flush enable
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] smart-link flush enable
```

4) Configuration on Device D

Create monitor link group 1.

```
<DeviceD> system-view
[DeviceD] monitor-link group 1
```

Configure GigabitEthernet 1/0/1 as the uplink port and GigabitEthernet 1/0/2 as the downlink port.

```
[DeviceD-mtlk-group1] port gigabitethernet 1/0/1 uplink
[DeviceD-mtlk-group1] port gigabitethernet 1/0/2 downlink
```

Configure VLAN 1 as the control VLAN for receiving flush messages on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceD-mtlk-group1] quit
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] smart-link flush enable
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] smart-link flush enable
```

Table of Contents

1 VLAN Configuration

When configuring VLAN, go to these sections for information you are interested in:

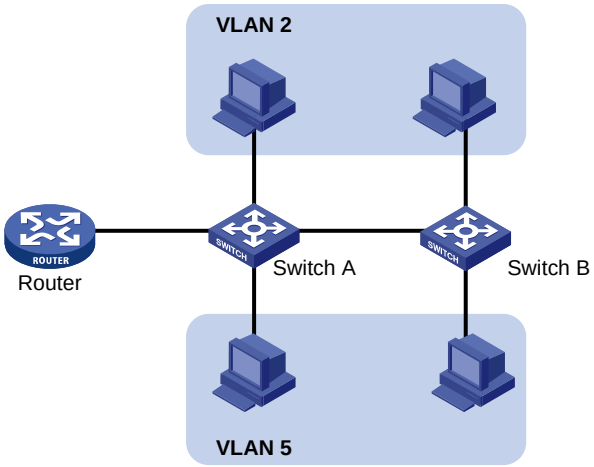
Introduction to VLAN

VLAN Overview

Ethernet is a network technology based on the Carrier Sense Multiple Access/Collision Detect (CSMA/CD) mechanism. As the medium is shared, collisions and excessive broadcasts cannot be avoided on an Ethernet. To address the issue, virtual LAN (VLAN) was introduced.

The idea is to break a LAN down into separate VLANs, that is, Layer 2 broadcast domains whereby frames are switched between ports assigned to the same VLAN. VLANs are isolated from each other at Layer 2. A VLAN is a bridging domain, and all broadcast traffic is contained within it, as shown in Figure 1-1.

Figure 1-1 A VLAN diagram



A VLAN is logically divided on an organizational basis rather than on a physical basis. For example, all workstations and servers used by a particular workgroup can be connected to the same LAN, regardless of their physical locations.

VLAN technology delivers the following benefits:

1) Confining broadcast traffic within individual VLANs. This reduces bandwidth waste and improves network performance.
2) Improving LAN security. By assigning user groups to different VLANs, you can isolate them at Layer 2. To enable communication between VLANs, routers or Layer 3 switches are required.
3) Flexible virtual workgroup creation. As users from the same workgroup can be assigned to the same VLAN regardless of their physical locations, network construction and maintenance is much easier and more flexible.

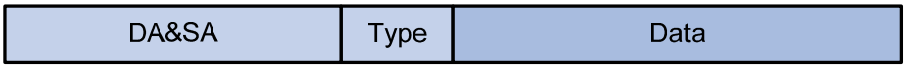
VLAN Fundamentals

To enable a network device to identify frames of different VLANs, a VLAN tag field is inserted into the data link layer encapsulation.

The format of VLAN-tagged frames is defined in IEEE 802.1Q issued by IEEE in 1999.

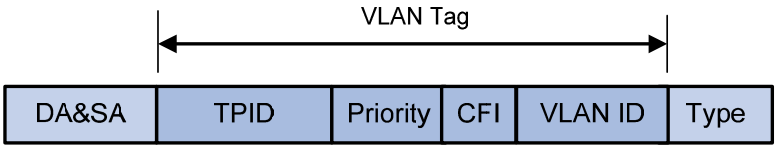
In the header of a traditional Ethernet data frame, the field after the destination MAC address and the source MAC address is the Type field indicating the upper layer protocol type, as shown in Figure 1-2.

Figure 1-2 The format of a traditional Ethernet frame



IEEE 802.1Q inserts a four-byte VLAN tag after the DA&SA field, as shown in Figure 1-3.

Figure 1-3 The position and format of VLAN tag



A VLAN tag comprises four fields: tag protocol identifier (TPID), priority, canonical format indicator (CFI), and VLAN ID.

- The 16-bit TPID field with a value of 0x8100 indicates that the frame is VLAN-tagged.
- The 3-bit priority field indicates the 802.1p priority of the frame. For information about frame priority, refer to *QoS Configuration* in the *QoS Volume*.
- The 1-bit CFI field specifies whether the MAC addresses are encapsulated in the standard format when packets are transmitted across different media. Value 0 indicates that MAC addresses are encapsulated in the standard format; value 1 indicates that MAC addresses are encapsulated in a non-standard format. The filed is 0 by default.
- The 12-bit VLAN ID field identifies the VLAN the frame belongs to. The VLAN ID range is 0 to 4095. As 0 and 4095 are reserved by the protocol, a VLAN ID actually ranges from 1 to 4094.

When receiving a frame, a network device handles the frame depending on whether the frame is VLAN tagged and the value of the VLAN tag, if any. For more information, refer to section Introduction to Port-Based VLAN.

Note

- The Ethernet II encapsulation format is used here. Besides the Ethernet II encapsulation format, other encapsulation formats, including 802.2 LLC, 802.2 SNAP, and 802.3 raw, are also supported by Ethernet. The VLAN tag fields are also added to frames encapsulated in these formats for VLAN identification.
- For a frame with multiple VLAN tags, the device handles it according to its outer VLAN tag, while transmits its inner VLAN tags as payload.

Types of VLAN

You can implement VLAN based on:

- Port
- MAC address
- Protocol
- IP subnet
- Policy
- Other criteria

This chapter covers port-based VLAN, MAC-based VLAN, protocol-based VLAN, and IP-based VLAN. You can configure the four types of VLANs on a port at the same time. When determining to which VLAN a packet passing through the port should be assigned, the device looks up the VLANs in the default order of MAC-based VLANs, IP-based VLANs, protocol-based VLANs, and port-based VLANs.

Configuring Basic VLAN Settings

Follow these steps to configure basic VLAN settings:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create VLANs	**vlan** { *vlan-id1* [**to** *vlan-id2*] | **all** }	Optional Using this command can create multiple VLANs in bulk.
Enter VLAN view	**vlan** *vlan-id*	Required If the specified VLAN does not exist, this command creates the VLAN first. By default, only the default VLAN (that is, VLAN 1) exists in the system.
Configure a name for the current VLAN	**name** *text*	Optional By default, the name of a VLAN is its VLAN ID, **VLAN 0001** for example.
Configure the description of the current VLAN	**description** *text*	Optional VLAN ID is used by default, for example, **VLAN 0001**.

Note

- As the default VLAN, VLAN 1 cannot be created or removed.
- You cannot manually create or remove VLANs reserved for special purposes.
- Dynamic VLANs cannot be removed with the **undo vlan** command.
- A VLAN with a QoS policy applied cannot be removed.
- For isolate-user-VLANs or secondary VLANs, if you have used the **isolate-user-vlan** command to create mappings between them, you cannot remove them until you remove the mappings between them first.
- A VLAN operating as a probe VLAN for remote port mirroring cannot be removed with the undo vlan command. To do that, remove the remote mirroring VLAN configuration from it first.

Configuring Basic Settings of a VLAN Interface

For hosts of different VLANs to communicate, you must use a router or Layer 3 switch to perform layer 3 forwarding. To achieve this, VLAN interfaces are used.

VLAN interfaces are virtual interfaces used for Layer 3 communication between different VLANs. They do not exist as physical entities on devices. For each VLAN, you can create one VLAN interface. You can assign the VLAN interface an IP address and specify it as the gateway of the VLAN to forward traffic destined for an IP network segment different from that of the VLAN.

Follow these steps to configure basic settings of a VLAN interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a VLAN interface and enter VLAN interface view	**interface vlan-interface** *vlan-interface-id*	Required If the VLAN interface already exists, you enter its view directly.
Assign an IP address to the VLAN interface	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Optional No IP address is assigned to any VLAN interface by default.
Configure the description of the VLAN interface	**description** *text*	Optional VLAN interface name is used by default, for example, **Vlan-interface1 Interface.**
Bring up the VLAN interface	**undo shutdown**	Optional By default, a VLAN interface is in the up state. In this case, the VLAN interface is up so long as one port in the VLAN is up and goes down if all ports in the VLAN go down. An administratively shut down VLAN interface however will be in the down state until you bring it up, regardless of how the state of the ports in the VLAN changes.

Note

Before creating a VLAN interface for a VLAN, create the VLAN first.

Port-Based VLAN Configuration

Introduction to Port-Based VLAN

Port-based VLANs group VLAN members by port. A port forwards traffic for a VLAN only after it is assigned to the VLAN.

Port link type

You can configure the link type of a port as access, trunk, or hybrid. The three link types use different VLAN tag handling methods. When configuring the link type of a port, note that:

- An access port can belong to only one VLAN. Usually, ports directly connected to PCs are configured as access ports.
- A trunk port can carry multiple VLANs to receive and send traffic for them. Except traffic of the default VLAN, traffic passes through a trunk port will be VLAN tagged. Usually, ports connecting network devices are configured as trunk ports to allow members of the same VLAN to communicate with each other across multiple network devices.
- Like a trunk port, a hybrid port can carry multiple VLANs to receive and send traffic for them. Unlike a trunk port, a hybrid port allows traffic of all VLANs to pass through VLAN untagged. You can configure a port connected to a network device or user terminal as a hybrid port for access link connectivity or trunk connectivity.

Default VLAN

By default, VLAN 1 is the default VLAN for all ports. You can configure the default VLAN for a port as required.

Use the following guidelines when configuring the default VLAN on a port:

- Because an access port can join only one VLAN, its default VLAN is the VLAN to which it belongs and cannot be configured.
- Because a trunk or hybrid port can join multiple VLANs, you can configure a default VLAN for the port.
- You can use a nonexistent VLAN as the default VLAN for a hybrid or trunk port but not for an access port. Therefore, after you remove the VLAN that an access port resides in with the **undo vlan** command, the default VLAN of the port changes to VLAN 1. The removal of the VLAN specified as the default VLAN of a trunk or hybrid port, however, does not affect the default VLAN setting on the port.

Note

- Do not set the voice VLAN as the default VLAN of a port in automatic voice VLAN assignment mode. Otherwise, the system prompts error information. For information about voice VLAN, refer to Voice VLAN Configuration.
- The local and remote ports must use the same default VLAN ID for the traffic of the default VLAN to be transmitted properly.

A port configured with the default VLAN handles a frame as follows:

Port type	Actions (in the inbound direction)		Actions (in the outbound direction)
	Untagged frame	Tagged frame	
Access	Tag the frame with the default VLAN tag.	• Receive the frame if its VLAN ID is the same as the default VLAN ID. • Drop the frame if its VLAN ID is different from the default VLAN ID.	Remove the default VLAN tag and send the frame.
Trunk	Check whether the default VLAN is permitted on the port: • If yes, tag the frame with the default VLAN tag. • If not, drop the frame.	• Receive the frame if its VLAN is carried on the port. • Drop the frame if its VLAN is not carried on the port.	• Remove the tag and send the frame if the frame carries the default VLAN tag. • Send the frame without removing the tag if its VLAN is carried on the port but is different from the default one.
Hybrid			Send the frame if its VLAN is carried on the port. The frame is sent with the VLAN tag removed or intact depending on your configuration with the **port hybrid vlan** command. This is true of the default VLAN.

Assigning an Access Port to a VLAN

You can assign an access port to a VLAN in VLAN view, interface view, or port group view.

1) In VLAN view

Follow these steps to assign one or multiple access ports to a VLAN in VLAN view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	Required If the specified VLAN does not exist, this command creates the VLAN first.

To do…	Use the command…	Remarks
Assign one or a group of access ports to the current VLAN	**port** *interface-list*	Required By default, all ports belong to VLAN 1.

2) In interface or port group view

Follow these steps to assign an access port (in interface view) or multiple access ports (in port group view) to a VLAN:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. • In Ethernet interface view, the subsequent configurations apply to the current port. • In port group view, the subsequent configurations apply to all ports in the port group. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Configure the link type of the port or ports as access		**port link-type access**	Optional The link type of a port is access by default.
Assign the current access port(s) to a VLAN		**port access vlan** *vlan-id*	Optional By default, all access ports belong to VLAN 1.

Note

- Before assigning an access port to a VLAN, create the VLAN first.
- After you configure a command on a Layer-2 aggregate interface, the system starts applying the configuration to the aggregate interface and its aggregation member ports. If the system fails to do that on the aggregate interface, it stops applying the configuration to the aggregation member ports. If it fails to do that on an aggregation member port, it simply skips the port and moves to the next port.

Assigning a Trunk Port to a VLAN

A trunk port can carry multiple VLANs. You can assign it to a VLAN in interface view or port group view.

Follow these steps to assign a trunk port to one or multiple VLANs:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. • In Ethernet interface view, the subsequent configurations apply to the current port. • In port group view, the subsequent configurations apply to all ports in the port group. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Configure the link type of the port or ports as trunk		**port link-type trunk**	Required
Assign the trunk port(s) to the specified VLAN(s)		**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port carries only VLAN 1.
Configure the default VLAN of the trunk port(s)		**port trunk pvid vlan** *vlan-id*	Optional VLAN 1 is the default VLAN by default.

Note

- To change the link type of a port from trunk to hybrid or vice versa, you must set the link type to access first.
- The local and remote hybrid ports must use the same default VLAN ID for the traffic of the default VLAN to be transmitted properly.
- After configuring the default VLAN for a trunk port, you must use the **port trunk permit vlan** command to configure the trunk port to allow packets from the default VLAN to pass through, so that the egress port can forward packets from the default VLAN.
- After you configure a command on a Layer-2 aggregate interface, the system starts applying the configuration to the aggregate interface and its aggregation member ports. If the system fails to do that on the aggregate interface, it stops applying the configuration to the aggregation member ports. If it fails to do that on an aggregation member port, it simply skips the port and moves to the next port.

Assigning a Hybrid Port to a VLAN

A hybrid port can carry multiple VLANs. You can assign it to a VLAN in interface view or port group view.

Follow these steps to assign a hybrid port to one or multiple VLANs:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. • In Ethernet interface view, the subsequent configurations apply to the current port. • In port group view, the subsequent configurations apply to all ports in the port group. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Configure the link type of the port(s) as hybrid		**port link-type hybrid**	Required
Assign the hybrid port(s) to the specified VLAN(s)		**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required By default, a hybrid port allows only packets of VLAN 1 to pass through untagged.
Configure the default VLAN of the hybrid port		**port hybrid pvid vlan** *vlan-id*	Optional VLAN 1 is the default by default.

Note

- To change the link type of a port from trunk to hybrid or vice versa, you must set the link type to access first.
- Before assigning a hybrid port to a VLAN, create the VLAN first.
- The local and remote hybrid ports must use the same default VLAN ID for the traffic of the default VLAN to be transmitted properly.
- After configuring the default VLAN for a hybrid port, you must use the **port hybrid vlan** command to configure the hybrid port to allow packets from the default VLAN to pass through, so that the egress port can forward packets from the default VLAN.
- After you configure a command on a Layer-2 aggregate interface, the system starts applying the configuration to the aggregate interface and its aggregation member ports. If the system fails to do that on the aggregate interface, it stops applying the configuration to the aggregation member ports. If it fails to do that on an aggregation member port, it simply skips the port and moves to the next port.

MAC-Based VLAN Configuration

Introduction to MAC-Based VLAN

MAC-based VLANs group VLAN members by MAC address. They only apply to untagged frames.

When receiving an untagged frame, the device looks up the list of MAC-to-VLAN mappings based on the MAC address of the frame for a match. If a match is found, the system forwards the frame in the corresponding VLAN. If no match is found, the system looks up other types of VLANs to make the forwarding decision.

MAC-based VLANs are mostly used in conjunction with security technologies such as 802.1X to provide secure, flexible network access for terminal devices.

Approaches to Creating MAC Address-to-VLAN Mappings

In addition to creating MAC address-to-VLAN mappings at the CLI, you can use an authentication server to automatically issue MAC address-to-VLAN mappings.

- Manually Static configuration (through CLI)

You can associate MAC addresses with VLANs by using corresponding commands.

- Automatic configuration through the authentication server (that is, VLAN issuing)

The device associates MAC addresses with VLANs dynamically based on the information provided by the authentication server. If a user goes offline, the corresponding MAC address-to-VLAN association is removed automatically. Automatic configuration requires MAC address-to–VLAN mapping be configured on the authentication server. For detailed information, refer to *802.1X Configuration* in the *Security Volume*.

The two configuration approaches can be used at the same time, that is, you can configure a MAC address-to-VLAN entry on both the local device and the authentication server at the same time. Note that the MAC address-to-VLAN entry configuration takes effect only when the configuration on the local device is consistent with that on the authentication server. Otherwise, the previous configuration takes effect.

Configuring a MAC Address-Based VLAN

Note

MAC-based VLANs are available only on hybrid ports.

Follow these steps to configure a MAC-based VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Associate MAC addresses with a VLAN	**mac-vlan mac-address** *mac-address* [**mask** *mac-mask*] **vlan** *vlan-id* [**priority** *priority*]	Required Support for the **mask** keyword in this command depends on the device model.

To do...		Use the command...	Remarks
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command. In Ethernet interface view, the subsequent configurations apply only to the current port; in port group view, the subsequent configurations apply to all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the link type of the port(s) as hybrid		**port link-type hybrid**	Required
Configure the current hybrid port(s) to permit packets of specific MAC-based VLANs to pass through		**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required By default, a hybrid port only permits the packets of VLAN 1 to pass through.
Enable MAC-based VLAN		**mac-vlan enable**	Required Disabled by default
Configure VLAN matching precedence		**vlan precedence** { **mac-vlan** | **ip-subnet-vlan** }	Optional By default, VLANs are preferentially matched based on MAC addresses.

Protocol-Based VLAN Configuration

Introduction to Protocol-Based VLAN

Note

Protocol-based VLANs are only applicable on hybrid ports.

In this approach, inbound packets are assigned to different VLANs based on their protocol types and encapsulation formats. The protocols that can be used for VLAN assignment include IP, IPX, and AppleTalk (AT). The encapsulation formats include Ethernet II, 802.3 raw, 802.2 LLC, and 802.2 SNAP.

A protocol-based VLAN is defined by a protocol template comprised of encapsulation format and protocol type. A port can be associated with multiple protocol templates. An untagged packet reaching a port associated with protocol-based VLANs will be processed as follows.

- If the packet matches a protocol template, the packet will be tagged with the VLAN tag corresponding to the protocol template.
- If the packet matches no protocol template, the packet will be tagged with the default VLAN ID of the port.

The port processes a tagged packet as it processes tagged packets of a port-based VLAN.

- If the port permits the VLAN ID of the packet to pass through, the port forwards the packet.
- If the port does not permit the VLAN ID of the packet to pass through, the port drops the packet.

This feature is mainly used to assign packets of the specific service type to a specific VLAN.

Configuring a Protocol-Based VLAN

Follow these steps to configure a protocol-based VLAN:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter VLAN view		**vlan** *vlan-id*	Required If the specified VLAN does not exist, this command creates the VLAN first.
Create a protocol template for the VLAN		**protocol-vlan** [*protocol-index*] { **at** | **ipv4** | **ipv6** | **ipx** { **ethernetii** | **llc** | **raw** | **snap** } | **mode** { **ethernetii etype** *etype-id* | **llc** { **dsap** *dsap-id* [**ssap** *ssap-id*] | **ssap** *ssap-id* } | **snap etype** *etype-id* } }	Required
Exit VLAN view		**quit**	Required
Enter interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. • In Ethernet interface view, the subsequent configurations apply to the current port. • In port group view, the subsequent configurations apply to all ports in the port group. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Configure the port link type as hybrid		**port link-type hybrid**	Required
Configure the port(s) to permit the packets of the specified protocol-based VLANs to pass through untagged		**port hybrid vlan** *vlan-id-list* **untagged**	Required
Associate the hybrid port(s) with the specified protocol-based VLAN		**port hybrid protocol-vlan vlan** *vlan-id* { *protocol-index* [**to** *protocol-end*] | **all** }	Required

Caution

- Do not configure both the *dsap-id* and *ssap-id* arguments in the **protocol-vlan** command as 0xe0 or 0xff when configuring the user-defined template for **llc** encapsulation. Otherwise, the encapsulation format of the matching packets will be the same as that of the **ipx llc** or **ipx raw** packets respectively.
- When you use the **mode** keyword to configure a user-defined protocol template, do not set *etype-id* in **ethernetii etype** *etype-id* to 0x0800, 0x8137, 0x809b, or 0x86dd. Otherwise, the encapsulation format of the matching packets will be the same as that of the IPv4, IPX, AppleTalk, and IPv6 packets respectively.
- A protocol-based VLAN on a hybrid port can process only untagged inbound packets, whereas the voice VLAN in automatic mode on a hybrid port can process only tagged voice traffic. Therefore, do not configure a VLAN as both a protocol-based VLAN and a voice VLAN. For more information, refer to Voice VLAN Configuration.
- After you configure a command on a Layer-2 aggregate interface, the system starts applying the configuration to the aggregate interface and its aggregation member ports. If the system fails to do that on the aggregate interface, it stops applying the configuration to the aggregation member ports. If it fails to do that on an aggregation member port, it simply skips the port and moves to the next port.

IP Subnet-Based VLAN Configuration

Introduction

In this approach, packets are assigned to VLANs based on their source IP addresses and subnet masks. A port configured with IP subnet-based VLANs assigns a received untagged packet to a VLAN based on the source address of the packet.

This feature is used to assign packets from the specified network segment or IP address to a specific VLAN.

Configuring an IP Subnet-Based VLAN

Note

This feature is only applicable on hybrid ports.

Follow these steps to configure an IP subnet-based VLAN:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—

To do…		Use the command…	Remarks
Associate an IP subnet with the current VLAN		**ip-subnet-vlan** [*ip-subnet-index*] **ip** *ip-address* [*mask*]	Required The IP network segment or IP address to be associated with a VLAN cannot be a multicast network segment or a multicast address.
Return to system view		**quit**	—
Enter interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use either command. • In Ethernet interface view, the subsequent configurations apply to the current port. • In port group view, the subsequent configurations apply to all ports in the port group. • In Layer-2 aggregate interface view, the subsequent configurations apply to the Layer-2 aggregate interface and all its member ports.
	Enter Layer-2 aggregate interface view	**interface bridge-aggregation** *interface-number*	
	Enter port group view	**port-group manual** *port-group-name*	
Configure port link type as hybrid		**port link-type hybrid**	Required
Configure the hybrid port(s) to permit the specified IP subnet-based VLANs to pass through		**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required
Associate the hybrid port(s) with the specified IP subnet-based VLAN		**port hybrid ip-subnet-vlan vlan** *vlan-id*	Required

Note

After you configure a command on a Layer-2 aggregate interface, the system starts applying the configuration to the aggregate interface and its aggregation member ports. If the system fails to do that on the aggregate interface, it stops applying the configuration to the aggregation member ports. If it fails to do that on an aggregation member port, it simply skips the port and moves to the next port.

Displaying and Maintaining VLAN

To do...	Use the command…	Remarks
Display VLAN information	**display vlan** [*vlan-id1* [**to** *vlan-id2*] | **all** | **dynamic** | **interface** *interface-type interface-number.subnumber* | **reserved** | **static**]	Available in any view

To do...	Use the command…	Remarks
Display VLAN interface information	**display interface vlan-interface** [*vlan-interface-id*]	Available in any view
Display hybrid ports or trunk ports on the device	**display port** { **hybrid** | **trunk** }	Available in any view
Display MAC address-to-VLAN entries	**display mac-vlan { all** | **dynamic** | **mac-address** *mac-address* [**mask** *mac-mask*] | **static** | **vlan vlan-id }**	Available in any view
Display all interfaces with MAC-based VLAN enabled	**display mac-vlan interface**	Available in any view
Display protocol information and protocol indexes of the specified VLANs	**display protocol-vlan vlan** { *vlan-id* [**to** *vlan-id*] | **all** }	Available in any view
Display protocol-based VLAN information on specified interfaces	**display protocol-vlan interface** { *interface-type interface-number* [**to** *interface-type interface-number*] | **all** }	Available in any view
Display IP subnet-based VLAN information and IP subnet indexes of specified VLANs	**display ip-subnet-vlan vlan** { *vlan-id* [**to** *vlan-id*] | **all** }	Available in any view
Display the IP subnet-based VLAN information and IP subnet indexes of specified ports	**display ip-subnet-vlan interface** { *interface-list* | **all** }	Available in any view
Clear statistics on a port	**reset counters interface** [*interface-type* [*interface-number*]]	Available in user view

Note

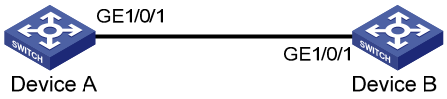
The **reset counters interface** command can be used to clear statistics on a VLAN interface. For more information, refer to *Ethernet Interface Commands* in the *Access Volume*.

VLAN Configuration Example

Network requirements

- Device A connects to Device B through a trunk port GigabitEthernet 1/0/1;
- The default VLAN ID of GigabitEthernet 1/0/1 is 100;
- GigabitEthernet 1/0/1 allows packets from VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass through.

Figure 1-4 Network diagram for port-based VLAN configuration



Configuration procedure

1) Configure Device A

Create VLAN 2, VLAN 6 through VLAN 50, and VLAN 100.

```
<DeviceA> system-view
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] vlan 100
[DeviceA-vlan100] vlan 6 to 50
Please wait... Done.
```

Enter GigabitEthernet 1/0/1 interface view.

```
[DeviceA] interface GigabitEthernet 1/0/1
```

Configure GigabitEthernet 1/0/1 as a trunk port and configure its default VLAN ID as 100.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk pvid vlan 100
```

Configure GigabitEthernet 1/0/1 to deny the packets of VLAN 1 (by default, the packets of VLAN 1 are permitted to pass through on all the ports).

```
[DeviceA-GigabitEthernet1/0/1] undo port trunk permit vlan 1
```

Configure GigabitEthernet 1/0/1 to permit packets from VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass through.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 2 6 to 50 100
Please wait... Done.
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] quit
```

2) Configure Device B as you configure Device A.

Verification

Verifying the configuration on Device A is similar to that of Device B. So only Device A is taken for example here.

Display the information about GigabitEthernet 1/0/1 of Device A to verify the above configurations.

```
<DeviceA> display interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 current state: UP
 IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 001e-c16f-ae68
 Description: GigabitEthernet1/0/1 Interface
 Loopback is not set
 Media type is twisted pair
 Port hardware type is  1000_BASE_T
 Unknown-speed mode, unknown-duplex mode
 Link speed type is autonegotiation, link duplex type is autonegotiation
```

```
Flow-control is not enabled
The Maximum Frame Length is 9216
Broadcast MAX-ratio: 100%
Unicast MAX-ratio: 100%
Multicast MAX-ratio: 100%
Allow jumbo frame to pass
PVID: 100
Mdi type: auto
Link delay is 0(sec)
Port link-type: trunk
 VLAN passing  : 2, 6-50, 100
 VLAN permitted: 2, 6-50, 100
 Trunk port encapsulation: IEEE 802.1q
Port priority: 0
Peak value of input: 0 bytes/sec, at 2000-04-26 12:01:40
Peak value of output: 0 bytes/sec, at 2000-04-26 12:01:40
Last 300 seconds input:  0 packets/sec 0 bytes/sec      -%
Last 300 seconds output:  0 packets/sec 0 bytes/sec     -%
Input (total):  0 packets, 0 bytes
         0 unicasts, 0 broadcasts, 0 multicasts
Input (normal):  0 packets, - bytes
         0 unicasts, 0 broadcasts, 0 multicasts
Input:  0 input errors, 0 runts, 0 giants, 0 throttles
         0 CRC, 0 frame, - overruns, 0 aborts
         - ignored, - parity errors
Output (total): 0 packets, 0 bytes
         0 unicasts, 0 broadcasts, 0 multicasts, 0 pauses
Output (normal): 0 packets, - bytes
         0 unicasts, 0 broadcasts, 0 multicasts, 0 pauses
Output: 0 output errors, - underruns, - buffer failures
         0 aborts, 0 deferred, 0 collisions, 0 late collisions
         0 lost carrier, - no carrier
```

The output above shows that:

- The port (GigabitEthernet 1/0/1) is a trunk port.
- The default VLAN of the port is VLAN 100.
- The port permits packets of VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass through.

Therefore, the configuration is successful.

2 Isolate-User-VLAN Configuration

When configuring an isolate-user VLAN, go to these sections for information you are interested in:

Overview

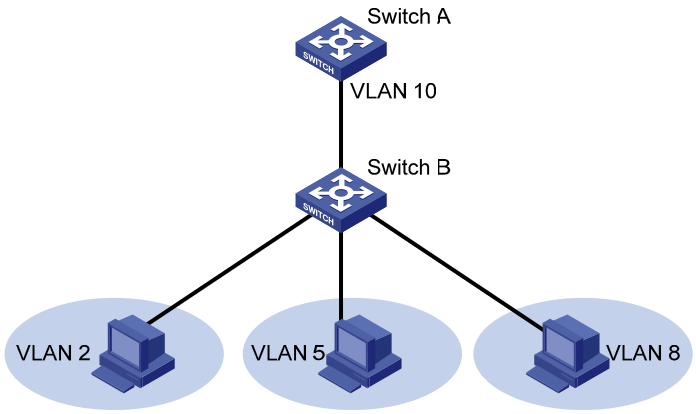
An isolate-user-VLAN adopts a two-tier VLAN structure. In this approach, two types of VLANs, isolate-user-VLAN and secondary VLAN, are configured on the same device.

The following are the characteristics of the isolate-user-VLAN implementation:

- Isolate-user-VLANs are mainly used for upstream data exchange. An isolate-user-VLAN can be associated with multiple secondary VLANs. As the upstream device is aware of only the isolate-user-VLAN but not the secondary VLANs, network configuration is simplified and VLAN resources are saved.
- You can isolate the Layer 2 traffic of different users by assigning the ports connected to them to different secondary VLANs. To enable communication between secondary VLANs associated with the same isolate-user-VLAN, you can enable local proxy ARP on the upstream device to realize Layer 3 communication between the secondary VLANs.

As illustrated in the following figure, the isolate-user-VLAN function is enabled on Switch B. VLAN 10 is the isolate-user-VLAN, and VLAN 2, VLAN 5, and VLAN 8 are secondary VLANs associated with VLAN 10 and are invisible to Switch A.

Figure 2-1 An isolate-user-VLAN example



Configuring Isolate-User-VLAN

Configure the isolate-user-VLAN through the following steps:

1) Configure the isolate-user-VLAN;
2) Configure the secondary VLANs;

3) Assign non-trunk ports to the isolate-user-VLAN and ensure that at least one port takes the isolate-user-VLAN as its default VLAN;
4) Assign non-trunk ports to each secondary VLAN and ensure that at least one port in a secondary VLAN takes the secondary VLAN as its default VLAN;
5) Associate the isolate-user-VLAN with the specified secondary VLANs.

Follow these steps to configure an isolate-user-VLAN:

To do...		Use the command	Remarks
Enter system view		**system-view**	—
Create a VLAN and enter VLAN view		**vlan** *vlan-id*	—
Configure the VLAN as an isolate-user-VLAN		**isolate-user-vlan enable**	Required
Return to system view		**quit**	—
Assign ports to the isolate-user-VLAN and ensure that at least one port takes the isolate-user-VLAN as its default VLAN	Access port	Refer to Assigning an Access Port to a VLAN	Use either approach.
	Hybrid port	Refer to Assigning a Hybrid Port to a VLAN	
Return to system view		**quit**	—
Create secondary VLANs		**vlan** { *vlan-id1* [**to** *vlan-id2*] | **all** }	Required
Quit to system view		**quit**	—
Assign ports to each secondary VLAN and ensure that at least one port in a secondary VLAN takes the secondary VLAN as its default VLAN	Access port	Refer to Assigning an Access Port to a VLAN	Required to choose either
	Hybrid port	Refer to Assigning a Hybrid Port to a VLAN	
Return to system view		**quit**	—
Associate the isolate-user-VLAN with the specified secondary VLANs		**isolate-user-vlan** *isolate-user-vlan-id* **secondary** *secondary-vlan-id* [**to** *secondary-vlan-id*]	Required

Note

After associating an isolate-user-VLAN with the specified secondary VLANs, you cannot add/remove a port to/from each involved VLAN or remove each involved VLAN. To do that, you must cancel the association first.

Displaying and Maintaining Isolate-User-VLAN

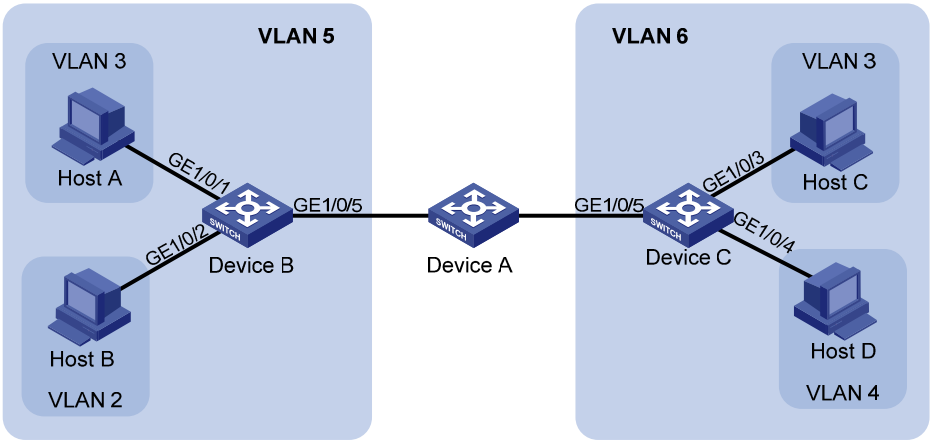
To do...	Use the command...	Remarks
Display the mapping between an isolate-user-VLAN and its secondary VLAN(s)	**display isolate-user-vlan** [*isolate-user-vlan-id*]	Available in any view

Isolate-User-VLAN Configuration Example

Network requirements

- Connect Device A to downstream devices Device B and Device C;
- Configure VLAN 5 on Device B as an isolate-user-VLAN, assign the uplink port GigabitEthernet 1/0/5 to VLAN 5, and associate VLAN 5 with secondary VLANs VLAN 2 and VLAN 3. Assign GigabitEthernet 1/0/2 to VLAN 2 and GigabitEthernet 1/0/1 to VLAN 3.
- Configure VLAN 6 on Device C as an isolate-user-VLAN, assign the uplink port GigabitEthernet 1/0/5 to VLAN 6, and associate VLAN 6 with secondary VLANs VLAN 3 and VLAN 4. Assign GigabitEthernet 1/0/3 to VLAN 3 and GigabitEthernet 1/0/4 to VLAN 4.
- For Device A, Device B only has VLAN 5 and Device C only has VLAN 6.

Figure 2-2 Network diagram for isolate-user-VLAN configuration



Configuration procedure

The following part provides only the configuration on Device B and Device C.

1) Configure Device B

Configure the isolate-user-VLAN.

```
<DeviceB> system-view
[DeviceB] vlan 5
[DeviceB-vlan5] isolate-user-vlan enable
[DeviceB-vlan5] port gigabitethernet 1/0/5
[DeviceB-vlan5] quit
```

Configure the secondary VLANs.

```
[DeviceB] vlan 3
[DeviceB-vlan3] port gigabitethernet 1/0/1
[DeviceB-vlan3] quit
```

```
[DeviceB] vlan 2
[DeviceB-vlan2] port gigabitethernet 1/0/2
[DeviceB-vlan2] quit
```

Associate the isolate-user-VLAN with the secondary VLANs.

```
[DeviceB] isolate-user-vlan 5 secondary 2 to 3
```

2) Configure Device C

Configure the isolate-user-VLAN.

```
<DeviceC> system-view
[DeviceC] vlan 6
[DeviceC-vlan6] isolate-user-vlan enable
[DeviceC-vlan6] port gigabitethernet 1/0/5
[DeviceC-vlan6] quit
```

Configure the secondary VLANs.

```
[DeviceC] vlan 3
[DeviceC-vlan3] port gigabitethernet 1/0/3
[DeviceC-vlan3] quit
[DeviceC] vlan 4
[DeviceC-vlan4] port gigabitethernet 1/0/4
```

Associate the isolate-user-VLAN with the secondary VLANs.

```
[DeviceC-vlan4] quit
[DeviceC] isolate-user-vlan 6 secondary 3 to 4
```

Verification

Display the isolate-user-VLAN configuration on Device B.

```
[DeviceB] display isolate-user-vlan
 Isolate-user-VLAN VLAN ID : 5
 Secondary VLAN ID : 2-3

 VLAN ID: 5
 VLAN Type: static
 Isolate-user-VLAN type : isolate-user-VLAN
 Route Interface: not configured
 Description: VLAN 0005
 Name: VLAN 0005
 Tagged   Ports: none
 Untagged Ports:
    gigabitethernet 1/0/1          gigabitethernet 1/0/2          gigabitethernet 1/0/5

 VLAN ID: 2
 VLAN Type: static
 Isolate-user-VLAN type : secondary
Route Interface: not configured
 Description: VLAN 0002
 Name: VLAN 0002
 Tagged   Ports: none
 Untagged Ports:
```

```
   gigabitethernet 1/0/2            gigabitethernet 1/0/5

VLAN ID: 3
VLAN Type: static
Isolate-user-VLAN type : secondary
Route Interface: not configured
Description: VLAN 0003
Name: VLAN 0003
Tagged   Ports: none
Untagged Ports:
   gigabitethernet 1/0/1            gigabitethernet 1/0/5
```

3 Voice VLAN Configuration

When configuring a voice VLAN, go to these sections for information you are interested in:

Overview

A voice VLAN is configured specially for voice traffic. After assigning the ports connecting to voice devices to a voice VLAN, you can configure quality of service (QoS) parameters for the voice traffic, thus improving transmission priority and ensuring voice quality.

A device determines whether a received packet is a voice packet by checking its source MAC address. A packet whose source MAC address complies with the voice device Organizationally Unique Identifier (OUI) address is regarded as voice traffic and assigned to the voice VLAN.

You can configure the OUI addresses in advance or use the default OUI addresses. Table 3-1 lists the default OUI address for each vendor's devices.

Table 3-1 The default OUI addresses of different vendors

Number	OUI address	Vendor
1	0001-e300-0000	Siemens phone
2	0003-6b00-0000	Cisco phone
3	0004-0d00-0000	Avaya phone
4	00d0-1e00-0000	Pingtel phone
5	0060-b900-0000	Philips/NEC phone
6	00e0-7500-0000	Polycom phone
7	00e0-bb00-0000	3Com phone

Note

- In general, as the first 24 bits of a MAC address (in binary format), an OUI address is a globally unique identifier assigned to a vendor by IEEE. OUI addresses mentioned in this document, however, are different from those in common sense. OUI addresses in this document are used by the system to determine whether a received packet is a voice packet. They are the results of the AND operation of the two arguments *mac-address* and *oui-mask* in the **voice vlan mac-address** command.
- You can remove the default OUI address of a device manually and then add new ones manually.

Voice VLAN Assignment Modes

A port can be assigned to a voice VLAN in one of the following two modes:

- In automatic mode, the system matches the source MAC addresses in the untagged packets sent when the IP phone is powered on against the OUI addresses. If a match is found, the system automatically assigns the port to the voice VLAN, issues ACL rules and configures the packet precedence. You can configure voice VLAN aging time on the device. The system will remove a port from the voice VLAN if no packet is received from the port after the aging time expires. Assigning/removing ports to/from a voice VLAN are automatically performed by the system.
- In manual mode, you should assign an IP phone connecting port to a voice VLAN manually. Then, the system matches the source MAC addresses in the packets against the OUI addresses. If a match is found, the system issues ACL rules and configures the packet precedence. In this mode, assigning/removing ports to/from a voice VLAN are performed manually.
- Both modes forward tagged packets according to their tags.

The following table lists the co-relation between the port voice VLAN mode, the voice traffic type of an IP phone, and the port link type.

Table 3-2 Co-relation

Voice VLAN assignment mode	Voice traffic type	Port link type
Automatic mode	Tagged voice traffic	Access: not supported
		Trunk: supported if the default VLAN of the connecting port exists and is not the voice VLAN and the connecting port belongs to the default VLAN
		Hybrid: supported if the default VLAN of the connecting port exists and is not the voice VLAN and the traffic of the default VLAN is permitted to pass through the connecting port
	Untagged voice traffic	Access, Trunk, hybrid: not supported
Manual mode	Tagged voice traffic	Access: not supported
		Trunk: supported if the default VLAN of the connecting port exists and is not the voice VLAN and the connecting port belongs to the default VLAN
		Hybrid: supported if the default VLAN of the connecting port exists and is not the voice VLAN, the traffice of the default VLAN is permitted to pass through the port, and the traffic of the Voice VLAN is permitted to pass through the connecting port tagged
	Untagged voice traffic	Access: supported if the default VLAN of the connecting port is the voice VLAN
		Trunk: supported if the default VLAN of the connecting port is the voice VLAN and that the voice VLAN is permitted to pass through the connecting port
		Hybrid port: supported if the default VLAN of the connecting port is the voice VLAN and is permitted to pass through the connecting port untagged

Caution

If an IP phone sends tagged voice traffic and its connecting port is configured with 802.1X authentication and guest VLAN, you should assign different VLAN IDs for the voice VLAN, the default VLAN of the connecting port, and the 802.1X guest VLAN.

Note

- The default VLANs for all ports are VLAN 1. You can configure the default VLAN of a port and configure a port to permit a certain VLAN to pass through with commands. For more information, refer to Port-Based VLAN Configuration.
- Use the **display interface** command to display the default VLAN of a port and the VLANs permitted to pass through the port.

Security Mode and Normal Mode of Voice VLANs

Voice VLAN-enabled ports can operate in security mode or normal mode based on their inbound packet filtering mechanisms.

- Security mode: only voice packets whose source MAC addresses comply with the recognizable OUI addresses can pass through the voice VLAN-enabled inbound port, while other non-voice packets are dropped, including authentication packets, such as 802.1X authentication packets.
- Normal mode: both voice packets and non-voice packets are allowed to pass through a voice VLAN-enabled inbound port. Voice packets are forwarded according to the voice VLAN forwarding mechanism whereas the non-voice packets are forwarded according to the normal VLAN forwarding mechanism.

It is recommended not to transmit both voice packets and non-voice packets in a voice VLAN. If necessary, please ensure that the voice VLAN security mode is disabled.

Configuring a Voice VLAN

Configuration Prerequisites

Before configuring a VLAN as a voice VLAN, create the VLAN first. Note that you cannot configure VLAN 1 (the system-default VLAN) as a voice VLAN.

Setting a Port to Operate in Automatic Voice VLAN Assignment Mode

Follow these steps to set a port to operate in automatic voice VLAN assignment mode:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Set the voice VLAN aging time	**voice vlan aging** *minutes*	Optional 1440 minutes by default. The voice VLAN aging time configuration is only applicable on ports in automatic voice VLAN assignment mode.
Enable the voice VLAN security mode	**voice vlan security enable**	Optional Enabled by default.
Add a recognizable OUI address	**voice vlan mac-address** *oui* **mask** *oui-mask* [**description** *text*]	Optional By default, each voice VLAN has default OUI addresses configured. Refer to Table 3-1 for the default OUI addresses of different vendors.
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Configure the port to operate in automatic voice VLAN assignment mode	**voice vlan mode auto**	Optional Automatic voice VLAN assignment mode is enabled by default. The voice VLAN assignment modes on different ports are independent of one another.
Enable voice VLAN on the port	**voice vlan** *vlan-id* **enable**	Required Not enabled by default

Note

- An S4800G switch supports up to eight voice VLANs globally.
- A protocol-based VLAN on a hybrid port can process only untagged inbound packets, whereas the voice VLAN in automatic mode on a hybrid port can process only tagged voice traffic. Therefore, do not configure a VLAN as both a protocol-based VLAN and a voice VLAN. For more information, refer to Protocol-Based VLAN Configuration.
- Do not configure the default VLAN of a port in automatic voice VLAN assignment mode as the voice VLAN.

Setting a Port to Operate in Manual Voice VLAN Assignment Mode

Follow these steps to set a port to operate in manual voice VLAN assignment mode:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enable the voice VLAN security mode		**voice vlan security enable**	Optional Enabled by default.
Add a recognizable OUI address		**voice vlan mac-address** *oui* **mask** *oui-mask* [**description** *text*]	Optional By default, each voice VLAN has default OUI addresses configured. Refer to Table 3-1 for the default OUI addresses of different vendors.
Enter interface view		**interface** *interface-type interface-number*	—
Configure the port to operate in manual voice VLAN assignment mode		**undo voice vlan mode auto**	Required Disabled by default
Assign the port in manual voice VLAN assignment mode to the voice VLAN	Access port	Refer to Assigning an Access Port to a VLAN.	Use one of the three approaches. After you assign an access port to the voice VLAN, the voice VLAN becomes the default VLAN of the port automatically.
	Trunk port	Refer to Assigning a Trunk Port to a VLAN.	
	Hybrid port	Refer to Assigning a Hybrid Port to a VLAN.	
Configure the voice VLAN as the default VLAN of the port	Trunk port	Refer to section Assigning a Trunk Port to a VLAN.	Optional This operation is required for untagged inbound voice traffic and prohibited for tagged inbound voice traffic.
	Hybrid port	Refer to Assigning a Hybrid Port to a VLAN.	
Enable voice VLAN on the port		**voice vlan enable**	Required

Note

- An S4800G switch supports up to eight voice VLANs globally.
- You can configure different voice VLANs on different ports at the same time. However, one port can be configured with only one voice VLAN, and this voice VLAN must be a static VLAN that already exists on the device.
- Voice VLAN is mutually exclusive with Link Aggregation Control Protocol (LACP) on a port.
- To make voice VLAN take effect on a port which is enabled with voice VLAN and operates in manual voice VLAN assignment mode, you need to assign the port to the voice VLAN manually.

Displaying and Maintaining Voice VLAN

To do...	Use the command...	Remarks
Display the voice VLAN state	**display voice vlan state**	Available in any view
Display the OUI addresses currently supported by system	**display voice vlan oui**	Available in any view

Voice VLAN Configuration Examples

Automatic Voice VLAN Mode Configuration Example

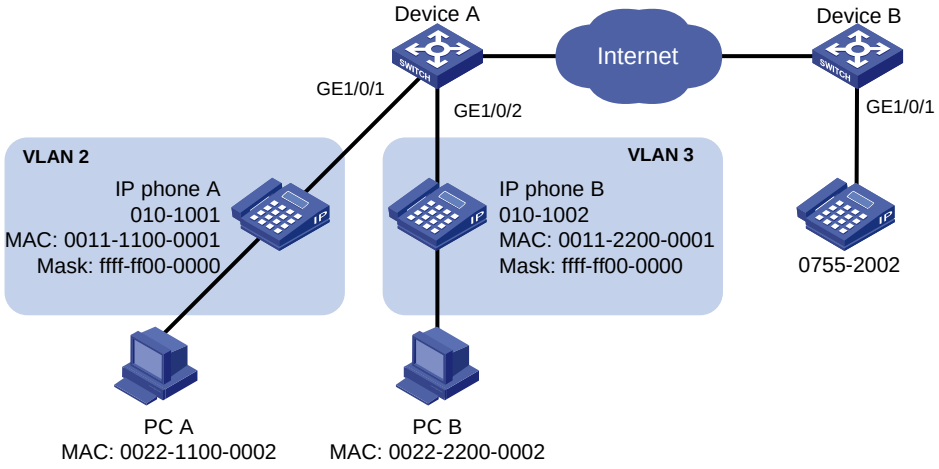
Network requirements

As shown in Figure 3-1,

- The MAC address of IP phone A is 0011-1100-0001. The phone connects to a downstream device named PC A whose MAC address is 0022-1100-0002 and to GigabitEthernet 1/0/1 on an upstream device named Device A.
- The MAC address of IP phone B is 0011-2200-0001. The phone connects to a downstream device named PC B whose MAC address is 0022-2200-0002 and to GigabitEthernet 1/0/2 on Device A.
- Device A uses voice VLAN 2 to transmit voice packets for IP phone A and voice VLAN 3 to transmit voice packets for IP phone B.

Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to work in automatic voice VLAN assignment mode. In addition, if one of them has not received any voice packet in 30 minutes, the port is removed from the corresponding voice VLAN automatically.

Figure 3-1 Network diagram for automatic voice VLAN assignment mode configuration



Configuration procedure

Create VLAN 2 and VLAN 3.

```
<DeviceA> system-view
[DeviceA] vlan 2 to 3
```

Set the voice VLAN aging time to 30 minutes.

```
[DeviceA] voice vlan aging 30
```

Since GigabitEthernet 1/0/1 may receive both voice traffic and data traffic at the same time, to ensure the quality of voice packets and effective bandwidth use, configure voice VLANs to work in security mode, that is, configure the voice VLANs to transmit only voice packets. (Optional. By default, voice VLANs work in security mode.)

```
[DeviceA] voice vlan security enable
```

Configure the allowed OUI addresses as MAC addresses prefixed by 0011-1100-0000 or 0011-2200-0000. In this way, Device A identifies packets whose MAC addresses match any of the configured OUI addresses as voice packets.

```
[DeviceA] voice vlan mac-address 0011-1100-0001 mask ffff-ff00-0000 description IP phone A
[DeviceA] voice vlan mac-address 0011-2200-0001 mask ffff-ff00-0000 description IP phone B
```

Configure GigabitEthernet 1/0/1 to operate in automatic voice VLAN assignment mode. (Optional. By default, a port operates in automatic voice VLAN assignment mode.)

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] voice vlan mode auto
```

Configure GigabitEthernet 1/0/1 as a hybrid port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type access
 Please wait... Done.
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
```

Configure VLAN 2 as the voice VLAN for GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] voice vlan 2 enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] voice vlan mode auto
[DeviceA-GigabitEthernet1/0/2] port link-type access
 Please wait... Done.
[DeviceA-GigabitEthernet1/0/2] port link-type hybrid
[DeviceA-GigabitEthernet1/0/2] voice vlan 3 enable
```

Verification

Display the OUI addresses, OUI address masks, and description strings supported currently.

```
<DeviceA> display voice vlan oui
Oui Address     Mask            Description
0001-e300-0000  ffff-ff00-0000  Siemens phone
0003-6b00-0000  ffff-ff00-0000  Cisco phone
0004-0d00-0000  ffff-ff00-0000  Avaya phone
0011-1100-0000  ffff-ff00-0000  IP phone A
0011-2200-0000  ffff-ff00-0000  IP phone B
00d0-1e00-0000  ffff-ff00-0000  Pingtel phone
0060-b900-0000  ffff-ff00-0000  Philips/NEC phone
00e0-7500-0000  ffff-ff00-0000  Polycom phone
00e0-bb00-0000  ffff-ff00-0000  3com phone
```

Display the current states of voice VLANs.

```
<DeviceA> display voice vlan state
```

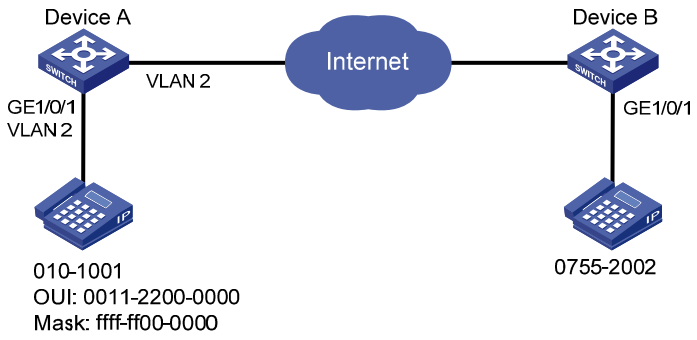
```
Maximum of Voice VLANs: 16
 Current Voice VLANs: 2
 Voice VLAN security mode: Security
 Voice VLAN aging time: 30 minutes
 Voice VLAN enabled port and its mode:
 PORT                        VLAN        MODE
 ---------------------------------------------
 GigabitEthernet1/0/1        2           AUTO
 GigabitEthernet1/0/2        3           AUTO
```

Manual Voice VLAN Assignment Mode Configuration Example

Network requirements

- Create VLAN 2 and configure it as a voice VLAN permitting only voice traffic to pass through.
- The IP phones send untagged voice traffic. Configure GigabitEthernet 1/0/1 as a hybrid port.
- Configure GigabitEthernet 1/0/1 to operate in manual voice VLAN assignment mode. Configure GigabitEthernet 1/0/1 to allow voice traffic with an OUI address of 0011-2200-0000, a mask of ffff-ff00-0000, and a description string **test** to be forwarded through the voice VLAN.

Figure 3-2 Network diagram for manual voice VLAN assignment mode configuration



Configuration procedure

Configure the voice VLAN to operate in security mode. (Optional. A voice VLAN operates in security mode by default.)

```
<DeviceA> system-view
[DeviceA] voice vlan security enable
```

Add a recognizable OUI address 0011-2200-0000.

```
[DeviceA] voice vlan mac-address 0011-2200-0000 mask ffff-ff00-0000 description test
```

Create VLAN 2 and configure it as the voice VLAN.

```
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] voice vlan 2 enable
```

Configure GigabitEthernet 1/0/1 to operate in manual voice VLAN assignment mode.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo voice vlan mode auto
```

Configure GigabitEthernet 1/0/1 as a hybrid port.

```
[DeviceA-GigabitEthernet1/0/1]port link-type access
 Please wait... Done.
[DeviceA-GigabitEthernet1/0/1]port link-type hybrid
```

Configure the voice VLAN (VLAN 2) as the default VLAN of GigabitEthernet 1/0/1 and configure GigabitEthernet 1/0/1 to permit the voice traffic of VLAN 2 to pass through untagged.

```
[DeviceA-GigabitEthernet1/0/1] port hybrid pvid vlan 2
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 2 untagged
```

Enable voice VLAN on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] voice vlan enable
```

Verification

Display the OUI addresses, OUI address masks, and description strings supported currently.

```
<DeviceA> display voice vlan oui
Oui Address     Mask            Description
0001-e300-0000  ffff-ff00-0000  Siemens phone
0003-6b00-0000  ffff-ff00-0000  Cisco phone
0004-0d00-0000  ffff-ff00-0000  Avaya phone
0011-2200-0000  ffff-ff00-0000  test
00d0-1e00-0000  ffff-ff00-0000  Pingtel phone
0060-b900-0000  ffff-ff00-0000  Philips/NEC phone
00e0-7500-0000  ffff-ff00-0000  Polycom phone
00e0-bb00-0000  ffff-ff00-0000  3com phone
```

Display the current voice VLAN state.

```
<DeviceA> display voice vlan state
Maximum of Voice VLANs: 16
 Current Voice VLANs: 2
 Voice VLAN security mode: Security
 Voice VLAN aging time: 100 minutes
 Voice VLAN enabled port and its mode:
 PORT                        VLAN        MODE
 ------------------------------------------------
 GigabitEthernet1/0/1        2           MANUAL
```

Table of Contents

1 GVRP Configuration

The GARP VLAN Registration Protocol (GVRP) is a GARP application. It functions based on the operating mechanism of GARP to maintain and propagate dynamic VLAN registration information for the GVRP devices on the network.

When configuring GVRP, go to these sections for information you are interested in:

Introduction to GVRP

GARP

The Generic Attribute Registration Protocol (GARP) provides a mechanism that allows participants in a GARP application to distribute, propagate, and register with other participants in a LAN the attributes specific to the GARP application, such as the VLAN or multicast address attribute.

GARP itself does not exist on a device as an entity. GARP-compliant participants are known as GARP applications. One example is GVRP. When a GARP participant is present on a port on your device, the port is regarded as a GARP participant.

GARP messages and timers

1) GARP messages

A GARP application entity exchanges information with other GARP application entities by:

- Sending Join messages to register with other entities its attributes, the attributes received from other GARP application entities, and the attributes manually configured on it.
- Sending Leave messages to have its attributes deregistered on other devices. A GARP participant also sends Leave messages when it receives Leave messages from other GARP participants or when attributes are manually deregistered on it.
- Sending LeaveAll messages to deregister all the attributes so that all GARP participants can re-register all attributes with each other. A LeaveAll message is sent upon expiration of a LeaveAll timer, which starts upon the startup of a GARP application entity.

Join messages, Leave messages, and LeaveAll message make sure the reregistration and deregistration of GARP attributes are performed in an orderly way.

Through message exchange, all attribute information that needs registration propagates to all GARP participants on the LAN.

2) GARP timers

GARP uses the following four timers to set the interval for sending GARP messages:

- Hold timer — When a GARP application entity receives the first registration request, it starts a Hold timer and collects succeeding requests. When the timer expires, the entity sends all these requests in one Join message. This helps you save bandwidth.
- Join timer — A GARP participant sends a Join message at most twice for reliability sake and uses a Join timer to set the sending interval. If the first Join message has not been acknowledged before the Join timer expires, the GARP participant sends the second Join message.
- Leave timer — Starts upon receipt of a Leave message sent for deregistering some attribute information. If no Join message is received before this timer expires, the GARP participant removes the attribute information as requested.
- LeaveAll timer — Starts when a GARP participant starts. When this timer expires, the entity sends a LeaveAll message so that other participants can re-register its attribute information. Then, a LeaveAll timer starts again.

Note

- The settings of GARP timers apply to all GARP applications, such as GVRP, on a LAN.
- On a GARP-enabled network, a device may send LeaveAll messages at the interval set by its LeaveAll timer or the LeaveAll timer on another device on the network, whichever is smaller. This is because each time a device on the network receives a LeaveAll message it resets its LeaveAll timer.

Operating mechanism of GARP

The GARP mechanism allows the configuration of a GARP application entity to propagate throughout a LAN quickly. In GARP, a GARP application entity registers or deregisters its attributes with other entities by making or withdrawing declarations of attributes and at the same time, based on received declarations or withdrawals, handles attributes of other entities. When a port receives an attribute declaration, it registers the attribute; when a port receives an attribute withdrawal, it deregisters the attribute.

GARP application entities send protocol data units (PDUs) with a particular multicast MAC address as destination. Based on this address, a device can identify to which GVRP application (GVRP for example) a GARP PDU will be delivered.

GARP message format

Figure 1-1 GARP message format

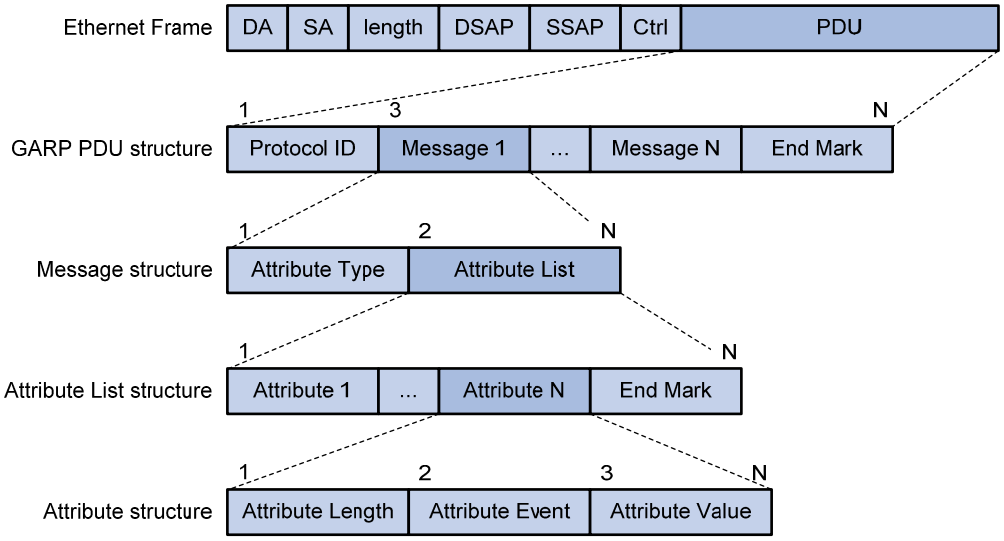


Figure 1-1 illustrates the GARP message format. Table 1-1 describes the GARP message fields.

Table 1-1 Description on the GARP message fields

Field	Description	Value
Protocol ID	Protocol identifier for GARP	1
Message	One or multiple messages, each containing an attribute type and an attribute list	—
Attribute Type	Defined by the concerned GARP application	0x01 for GVRP, indicating the VLAN ID attribute
Attribute List	Contains one or multiple attributes	—
Attribute	Consists of an Attribute Length, an Attribute Event, and an Attribute Value	—
Attribute Length	Number of octets occupied by an attribute, inclusive of the attribute length field	2 to 255 (in bytes)
Attribute Event	Event described by the attribute	• 0: LeaveAll event • 1: JoinEmpty event • 2: JoinIn event • 3: LeaveEmpty event • 4: LeaveIn event • 5: Empty event
Attribute Value	Attribute value	VLAN ID for GVRP If the Attribute Event is LeaveAll, Attribute Value is omitted.
End Mark	Indicates the end of a GARP PDU	0x00

GVRP

GVRP enables a device to propagate local VLAN registration information to other participant devices and dynamically update the VLAN registration information from other devices to its local database

about active VLAN members and through which port they can be reached. It thus ensures that all GVRP participants on a bridged LAN maintain the same VLAN registration information. The VLAN registration information propagated by GVRP includes both manually configured local static entries and dynamic entries from other devices.

GVRP provides the following three registration types on a port:

- Normal — Enables the port to dynamically register and deregister VLANs, and to propagate both dynamic and static VLAN information.
- Fixed — Disables the port to dynamically register and deregister VLANs or propagate information about dynamic VLANs, but allows the port to propagate information about static VLANs. A trunk port with fixed registration type thus allows only manually configured VLANs to pass through even though it is configured to carry all VLANs.
- Forbidden — Disables the port to dynamically register and deregister VLANs and to propagate VLAN information except information about VLAN 1. A trunk port with forbidden registration type thus allows only VLAN 1 to pass through even though it is configured to carry all VLANs.

Protocols and Standards

GVRP is described in IEEE 802.1Q.

GVRP Configuration Task List

Complete these tasks to configure GVRP:

Task	Remarks
Configuring GVRP Functions	Required
Configuring GARP Timers	Optional

Note

- GVRP configuration made in Ethernet interface view or Layer-2 aggregate interface view takes effect on the current interface only; .GVRP configuration made in port group view takes effect on all the member ports in the group.
- GVRP configuration made on a member port in an aggregation group takes effect only after the port is removed from the aggregation group.

Configuring GVRP Functions

Before enabling GVRP on a port, you must enable GVRP globally.

Follow these steps to configure GVRP functions on a trunk port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable GVRP globally	**gvrp**	Required Globally disabled by default.

To do…		Use the command…	Remarks
Enter Ethernet interface view, Layer 2 aggregate interface view, or port-group view	Enter Ethernet interface view or Layer 2 aggregate interface view	**interface** *interface-type interface-number*	Required Perform either of the commands.
	Enter port-group view	**port-group manual** *port-group-name*	
Enable GVRP on the port or port group		**gvrp**	Required Disabled by default.
Configure the GVRP registration mode on the port or port group		**gvrp registration** { **fixed** | **forbidden** | **normal** }	Optional The default is **normal**.

Note

- GVRP can be configured only on trunk ports.
- GVRP is mutually exclusive with service loopback.
- In an MSTP network, GVRP can run on only the CIST. In addition, blocked ports on the CIST cannot receive/send GVRP packets.
- If both GVRP and remote port mirroring are used, GVRP may register the remote probe VLAN to unexpected ports, resulting in undesired duplicates to be received by the monitor port. For more information about port mirroring, refer to Port Mirroring Configuration in the Access Volume.
- Enabling GVRP on a Layer 2 aggregate interface enables both the aggregate interface and all selected member ports in the corresponding link aggregation group to participate in dynamic VLAN registration and deregistration.
- On a GVRP-enabled trunk port, you need to configure the port trunk permit vlan all command on the port to ensure that the traffic of all dynamically registered VLANs can pass through the port.

Configuring GARP Timers

Among the four GARP timers, the LeaveAll timer is configured in system view and takes effect on all ports, while the other three are configured on a port basis.

Follow these steps to configure GARP timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the GARP LeaveAll timer	**garp timer leaveall** *timer-value*	Optional The default is 1000 centiseconds.

To do…		Use the command…	Remarks
Enter Ethernet interface view, Layer 2 aggregate interface view, or port-group view	Enter Ethernet or Layer 2 aggregate interface view	**interface** *interface-type interface-number*	Required Perform either of the commands. Depending on the view you accessed, the subsequent configuration takes effect on a port or all ports in a port-group.
	Enter port-group view	**port-group manual** *port-group-name*	
Configure the Hold timer		**garp timer hold** *timer-value*	Optional 10 centiseconds by default.
Configure the Join timer		**garp timer join** *timer-value*	Optional 20 centiseconds by default.
Configure the Leave timer		**garp timer leave** *timer-value*	Optional 60 centiseconds by default.

Note

As shown in Table 1-2, the values of GARP timers are dependent on each other:

- If you want to set a value beyond the value range for a timer, you may change the value range by tuning the value of another related timer.
- If you want to restore the default settings of the timers, restore the Hold timer first, and then the Join, Leave, and LeaveAll timers.

Table 1-2 Dependencies of GARP timers

Timer	Lower limit	Upper limit
Hold	10 centiseconds	No greater than half of the Join timer setting
Join	No less than two times the Hold timer setting	Less than half of the leave timer setting
Leave	Greater than two times the Join timer setting	Less than the LeaveAll timer setting
LeaveAll	Greater than the Leave timer setting	32765 centiseconds

Displaying and Maintaining GVRP

To do…	Use the command…	Remarks
Display statistics about GARP	**display garp statistics** [**interface** *interface-list*]	Available in any view
Display GARP timers for specified or all ports	**display garp timer** [**interface** *interface-list*]	Available in any view
Display the local VLAN information maintained by GVRP	**display gvrp local-vlan interface** *interface-type interface-number*	Available in any view

To do…	Use the command…	Remarks
Display the current GVRP state	**display gvrp state interface** *interface-type interface-number* **vlan** *vlan-id*	Available in any view
Display statistics about GVRP	**display gvrp statistics** [**interface** *interface-list*]	Available in any view
Display the global GVRP state	**display gvrp status**	Available in any view
Display the information about dynamic VLAN operations performed on a port	**display gvrp vlan-operation interface** *interface-type interface-number*	Available in any view
Clear the GARP statistics	**reset garp statistics** [**interface** *interface-list*]	Available in user view

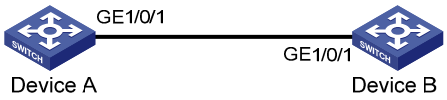
GVRP Configuration Examples

GVRP Configuration Example I

Network requirements

Configure GVRP for dynamic VLAN information registration and update among devices, adopting the normal registration mode on ports.

Figure 1-2 Network diagram for GVRP configuration



Configuration procedure

1) Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on trunk port GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2) Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
```

```
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on trunk port GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[DeviceB] vlan 3
```

3) Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
 Now, the following dynamic VLAN exist(s):
  3
```

Display dynamic VLAN information on Device B.

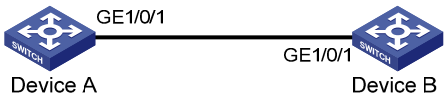
```
[DeviceB] display vlan dynamic
 Now, the following dynamic VLAN exist(s):
  2
```

GVRP Configuration Example II

Network requirements

Configure GVRP for dynamic VLAN information registration and update among devices. Specify fixed GVRP registration on Device A and normal GVRP registration on Device B.

Figure 1-3 Network diagram for GVRP configuration



Configuration procedure

1) Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1 and set the GVRP registration type to fixed on the port.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
[DeviceA-GigabitEthernet1/0/1] gvrp registration fixed
```

```
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2) Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[Sysname] vlan 3
```

3) Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
 No dynamic vlans exist!
```

Display dynamic VLAN information on Device B.

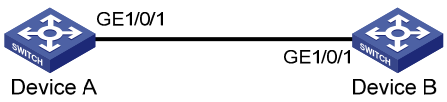
```
[DeviceB] display vlan dynamic
 Now, the following dynamic VLAN exist(s):
  2
```

GVRP Configuration Example III

Network requirements

To prevent dynamic VLAN information registration and update among devices, set the GVRP registration mode to **forbidden** on Device A and **normal** on Device B.

Figure 1-4 Network diagram for GVRP configuration



Configuration procedure

1) Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1 and set the GVRP registration type to forbidden on the port.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
[DeviceA-GigabitEthernet1/0/1] gvrp registration forbidden
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2) Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a trunk port, allowing all VLANs to pass through.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[DeviceB] vlan 3
```

3) Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
 No dynamic vlans exist!
```

Display dynamic VLAN information on Device B.

```
[DeviceB] display vlan dynamic
 No dynamic vlans exist!
```

Table of Contents

1 QinQ Configuration

When configuring QinQ, go to these sections for information you are interested in:

Note

Throughout this document, customer network VLANs (CVLANs), also called inner VLANs, refer to the VLANs that a customer uses on the private network; and service provider network VLANs (SVLANs), also called outer VLANs, refer to the VLANs that a service provider uses to carry VLAN tagged traffic for customers.

Introduction to QinQ

Background

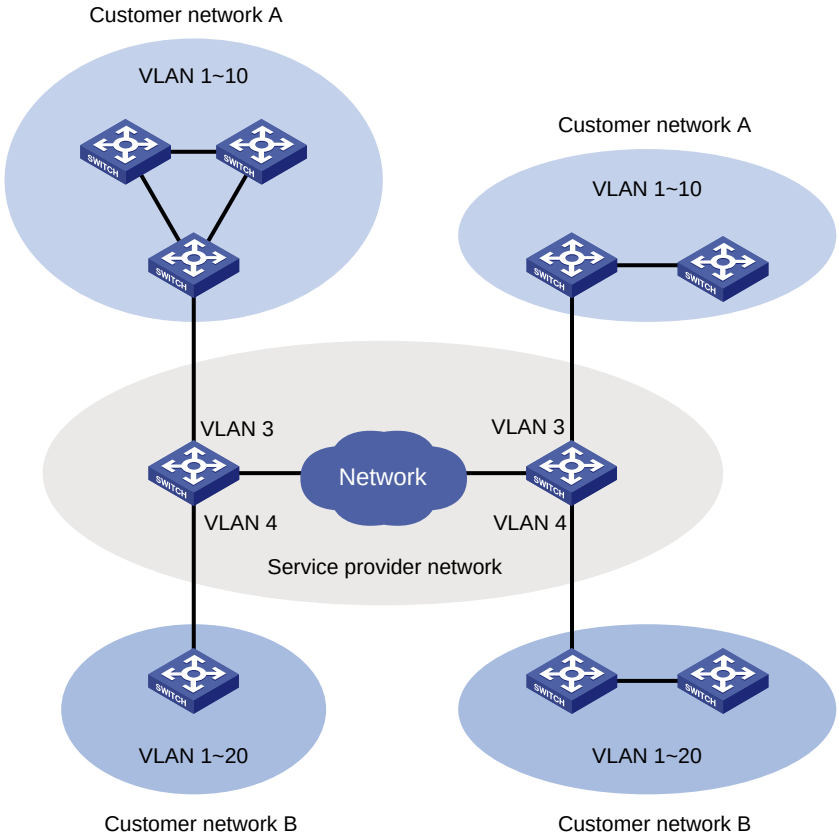
In the VLAN tag field defined in IEEE 802.1Q, only 12 bits are used for VLAN IDs, so a device can support a maximum of 4094 VLANs. In actual applications, however, a large number of VLANs are required to isolate users, especially in metropolitan area networks (MANs), and 4094 VLANs are far from satisfying such requirements.

QinQ Mechanism and Benefits

The QinQ feature is a flexible, easy-to-implement Layer 2 VPN technique. It enables the edge device on the service provider network to encapsulate an outer VLAN tag in Ethernet frames from customer networks (private networks), so that the Ethernet frames will travel across the service provider network (public network) with double VLAN tags. QinQ enables a service provider to use a single SVLAN to serve customers who have multiple CVLANs.

The devices in the public network forward a frame only according to its outer VLAN tag and learn its source MAC address into the MAC address table of the outer VLAN. The inner VLAN tag of the frame is transmitted as the payload.

Figure 1-1 Schematic diagram of the QinQ feature



As shown in Figure 1-1, customer network A has CVLANs 1 through 10, while customer network B has CVLANs 1 through 20. The SVLAN allocated by the service provider for customer network A is SVLAN 3, and that for customer network B is SVLAN 4. When a tagged Ethernet frame of customer network A enters the service provider network, it is tagged with outer VLAN 3; when a tagged Ethernet frame of customer network B enters the service provider network, it is tagged with outer VLAN 4. In this way, there is no overlap of VLAN IDs among customers, and traffic from different customers does not become mixed.

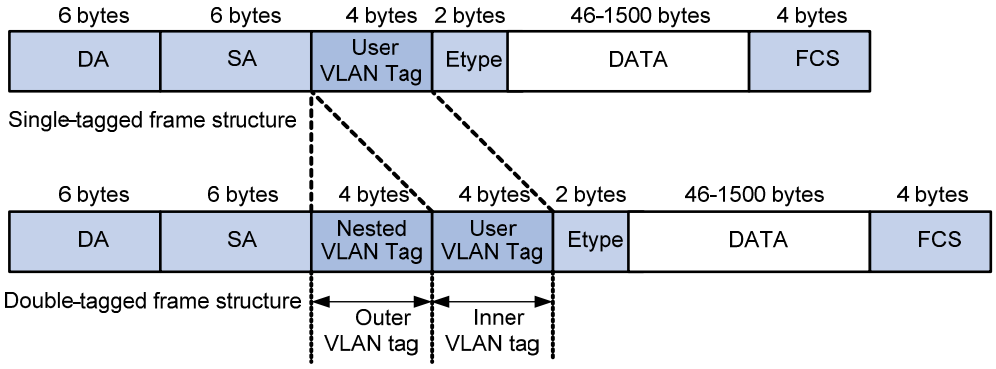
By tagging tagged frames, QinQ expands the available VLAN space from 4094 to 4094 × 4094 and thus satisfies the requirement for VLAN space in MAN. It mainly addresses the following issues:

- Releases the stress on the SVLAN resource.
- Enables customers to plan their CVLANs without conflicting with SVLANs.
- Provides an easy-to-implement Layer 2 VPN solution for small-sized MANs or intranets.

QinQ Frame Structure

A QinQ frame is transmitted double-tagged over the service provider network. The inner VLAN tag is the CVLAN tag while the outer one is the SVLAN tag that the service provider has allocated to the customer. Figure 1-2 shows the structure of single-tagged and double-tagged Ethernet frames.

Figure 1-2 Single-tagged frame structure vs. double-tagged Ethernet frame structure



Note

The default maximum transmission unit (MTU) of an interface is 1500 bytes. The size of an outer VLAN tag is 4 bytes. Therefore, you are recommended to increase the MTU of each interface on the service provider network. The recommended minimum MTU is 1504 bytes. For how to configure the MTU of an interface, refer to *Ethernet Interface Configuration* in the *Access Volume*.

Implementations of QinQ

There are two types of QinQ implementations: basic QinQ and selective QinQ.

1) Basic QinQ

Basic QinQ is a port-based feature. When a frame arrives at a basic QinQ-enabled port, the port tags it with the port's default VLAN tag, regardless of whether the frame is tagged or untagged. If the received frame is already tagged, it becomes a double-tagged frame; if it is untagged, it becomes a frame tagged with the port's default VLAN tag.

2) Selective QinQ

Selective QinQ is a more flexible, VLAN-based implementation of QinQ. In addition to all the functions of basic QinQ, selective QinQ provides per-CVLAN actions for frames received on the same port:

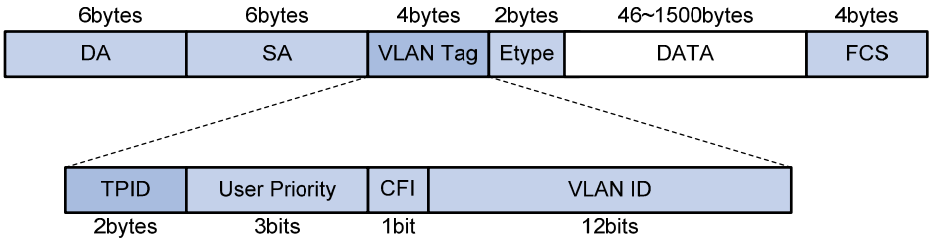
- Tagging frames with different outer VLAN tags based on different inner VLAN IDs.
- Marking the outer VLAN 802.1p priority based on the existing inner VLAN 802.1p priority.
- Modifying the inner VLAN IDs while tagging the frames with outer VLAN tags.

Modifying the TPID in a VLAN Tag

A VLAN tag uses the tag protocol identifier (TPID) field to identify the protocol type of the tag. The value of this field, as defined in IEEE 802.1Q, is 0x8100.

0 shows the 802.1Q-defined tag structure of an Ethernet frame.

Figure 1-3 VLAN tag structure of an Ethernet frame



The device determines whether a received frame carries a SVLAN tag or a CVLAN tag by checking the corresponding TPID value. Upon receiving a frame, the device compares the configured TPID value with the value of the TPID field in the frame. If the two match, the devices considers that the frame carries the corresponding VLAN tag. For example, if a frame carries a SVLAN tag with the TPID value 0x9100 and a CVLAN tag with the TPID value 0x8100 while the configured TPID value of the SVLAN tag is 0x9100 and that of the CVLAN tag is 0x8200, the device considers that the frame carries only the SVLAN tag but not the CVLAN tag.

In addition, the systems of different vendors may set the TPID of the outer VLAN tag of QinQ frames to different values. For compatibility with these systems, you can modify the TPID value so that the QinQ frames, when sent to the public network, carry the TPID value identical to the value of a particular vendor to allow interoperability with the devices of that vendor.

The TPID in an Ethernet frame has the same position with the protocol type field in a frame without a VLAN tag. To avoid problems in packet forwarding and handling in the network, you cannot set the TPID value to any of the values in the table below.

Table 1-1 Reserved protocol type values

Protocol type	Value
ARP	0x0806
PUP	0x0200
RARP	0x8035
IP	0x0800
IPv6	0x86DD
PPPoE	0x8863/0x8864
MPLS	0x8847/0x8848
IPX/SPX	0x8137
IS-IS	0x8000
LACP	0x8809
802.1x	0x888E
Cluster	0x88A7
Reserved	0xFFFD/0xFFFE/0xFFFF

QinQ Configuration Task List

Table 1-2 QinQ configuration task list

Configuration task		Remarks
Configuring Basic QinQ		Optional
Configuring Selective QinQ	Configuring an Outer VLAN Tagging Policy	Optional
Configuring the TPID Value in VLAN Tags		Optional

Note

- QinQ requires configurations only on the service provider network, not on the customer network.
- QinQ configurations made in Ethernet interface view take effect on the current interface only; those made in Layer-2 aggregate interface view take effect on the current aggregate interface and all the member ports in the aggregation group; those made in port group view take effect on all member ports in the current port group.
- Basic and selective QinQ should both be configured on the ports connecting customer networks.
- Do not configure QinQ on a reflector port. For information about reflector ports, refer to *Port Mirroring Configuration* in the *Access Volume*.

Configuring Basic QinQ

Enabling Basic QinQ

Follow these steps to enable basic QinQ:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command.
	Enter port group view	**port-group manual** *port-group-name*	
Enable QinQ on the port(s)		**qinq enable**	Required Disabled by default.

Configuring Selective QinQ

Configuring an Outer VLAN Tagging Policy

Basic QinQ can only tag received frames with the default VLAN tag of the receiving port, while selective QinQ allows adding different outer VLAN tags based on different inner VLAN tags.

3Com Switch 4800G support the configuration of basic QinQ and selective QinQ at the same time on a port and when the two features are both enabled on the port, frames that meet the selective QinQ

condition are handled with selective QinQ on this port first, and the left frames are handled with basic QinQ.

Follow these steps to configure an outer VLAN tagging policy:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command
	Enter port group view	**port-group manual** *port-group-name*	
Enter QinQ view and configure the SVLAN tag for the port to add		**qinq vid** *vlan-id*	Required By default, the SVLAN tag to be added is the default VLAN tag of the receiving port.
Tag frames of the specified CVLANs with the current SVLAN		**raw-vlan-id inbound** { **all** | *vlan-list* }	Required

Caution

- An inner VLAN tag corresponds to only one outer VLAN tag.
- If you want to change an outer VLAN tag, you must delete the old outer VLAN tag configuration and configure a new outer VLAN tag.

Configuring the TPID Value in VLAN Tags

You can configure the TPID value in VLAN tags in system view, where the configuration takes effect on all ports of the device.

Follow these steps to configure a TPID value globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the TPID value in the CVLAN tag or the SVLAN tag	**qinq ethernet-type** [**customer-tag** | **service-tag**] *hex-value*	Optional Both 0x8100 by default

QinQ Configuration Examples

Basic QinQ Configuration Example

Network requirements

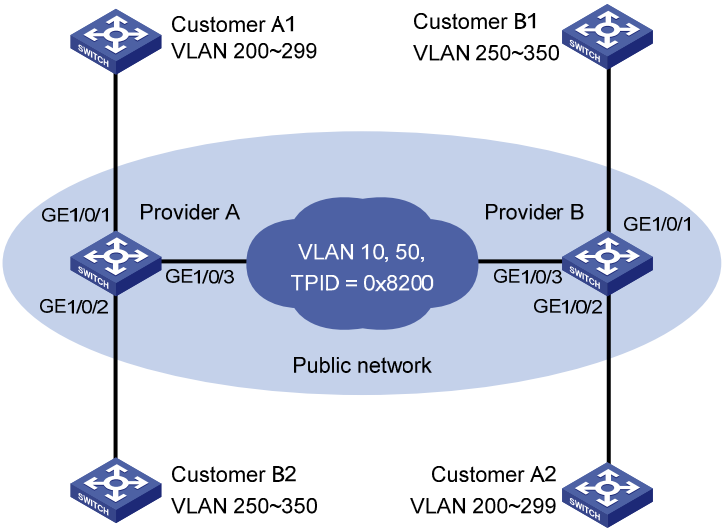
- Provider A and Provider B are edge devices on the service provider network and are interconnected through trunk ports. They belong to SVLAN 10 and 50.

- Customer A1, Customer A2, Customer B1 and Customer B2 are edge devices on the customer network.
- Third-party devices with a TPID value of 0x8200 are deployed between Provider A and Provider B.

Make configuration to achieve the following:

- Frames of VLAN 200 through VLAN 299 can be exchanged between Customer A1and Customer A2 through VLAN 10 of the service provider network.
- Frames of VLAN 250 through VLAN 350 can be exchanged between Customer B1 and Customer B2 through VLAN 50 of the service provider network.

Figure 1-4 Network diagram for VLAN transparent transmission configuration



Configuration procedure

Note

Make sure that the devices in the service provider network have been configured to allow QinQ packets to pass through.

1) Configuration on Provider A

- Configure GigabitEthernet 1/0/1

Configure VLAN 10 as the default VLAN of GigabitEthernet 1/0/1.

```
<ProviderA> system-view
[ProviderA] interface gigabitethernet 1/0/1
[ProviderA-GigabitEthernet1/0/1] port access vlan 10
```

Enable basic QinQ on GigabitEthernet 1/0/1.

```
[ProviderA-GigabitEthernet1/0/1] qinq enable
[ProviderA-GigabitEthernet1/0/1] quit
```

- Configure GigabitEthernet 1/0/2

Configure GigabitEthernet 1/0/2 as a hybrid port and configure VLAN 50 as the default VLAN of the port.

```
[ProviderA] interface gigabitethernet 1/0/2
[ProviderA-GigabitEthernet1/0/2] port link-type hybrid
[ProviderA-GigabitEthernet1/0/2] port hybrid pvid vlan 50
[ProviderA-GigabitEthernet1/0/2] port hybrid vlan 50 untagged
```

Enable basic QinQ on GigabitEthernet 1/0/2.

```
[ProviderA-GigabitEthernet1/0/2] qinq enable
[ProviderA-GigabitEthernet1/0/2] quit
```

- Configure GigabitEthernet 1/0/3

Configure GigabitEthernet 1/0/3 as a trunk port to permit frames of VLAN 10 and 50 to pass through.

```
[ProviderA] interface gigabitethernet 1/0/3
[ProviderA-GigabitEthernet1/0/3] port link-type trunk
[ProviderA-GigabitEthernet1/0/3] port trunk permit vlan 10 50
```

Set the TPID value in the outer tag to 0x8200.

```
[ProviderA-GigabitEthernet1/0/3] quit
[ProviderA] qinq ethernet-type service-tag 8200
```

2) Configuration on Provider B

- Configure GigabitEthernet 1/0/1

Configure VLAN 50 as the default VLAN of GigabitEthernet 1/0/1.

```
<ProviderB> system-view
[ProviderB] interface gigabitethernet 1/0/1
[ProviderB-GigabitEthernet1/0/1] port access vlan 50
```

Enable basic QinQ on GigabitEthernet 1/0/1.

```
[ProviderB-GigabitEthernet1/0/1] qinq enable
[ProviderB-GigabitEthernet1/0/1] quit
```

- Configure GigabitEthernet 1/0/2

Configure GigabitEthernet 1/0/2 as a hybrid port and configure VLAN 10 as the default VLAN of the port.

```
[ProviderB] interface gigabitethernet 1/0/2
[ProviderB-GigabitEthernet1/0/2] port link-type hybrid
[ProviderB-GigabitEthernet1/0/2] port hybrid pvid vlan 10
[ProviderB-GigabitEthernet1/0/2] port hybrid vlan 10 untagged
```

Enable basic QinQ on GigabitEthernet 1/0/2.

```
[ProviderB-GigabitEthernet1/0/2] qinq enable
[ProviderB-GigabitEthernet1/0/2] quit
```

- Configure GigabitEthernet 1/0/3

Configure GigabitEthernet 1/0/3 as a trunk port to permit frames of VLAN 10 and 50 to pass through.

```
[ProviderB] interface gigabitethernet 1/0/3
[ProviderA-GigabitEthernet1/0/B] port link-type trunk
[ProviderA-GigabitEthernet1/0/B] port trunk permit vlan 10 50
```

Set the TPID value in the outer tag to 0x8200.

```
[ProviderB-GigabitEthernet1/0/3] quit
[ProviderB] qinq ethernet-type service-tag 8200
```

3) Configuration on third-party devices

Configure the third-party devices between Provider A and Provider B as follows: configure the port connecting GigabitEthernet 1/0/3 of Provider A and that connecting GigabitEthernet 1/0/3 of Provider B to allow tagged frames of VLAN 10 and 50 to pass through.

Comprehensive Selective QinQ Configuration Example

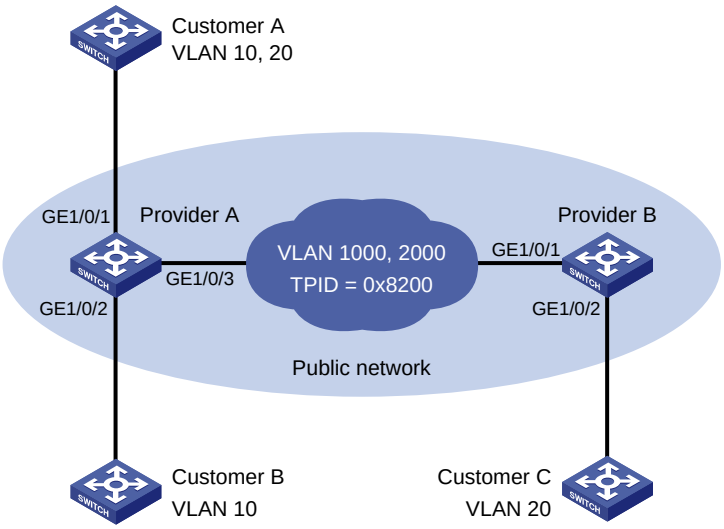
Network requirements

- Provider A and Provider B are edge devices on the service provider network and are interconnected through trunk ports. They belong to SVLAN 1000 and SVLAN 2000 separately.
- Customer A, Customer B and Customer C are edge devices on the customer network.
- Third-party devices with a TPID value of 0x8200 are deployed between Provider A and Provider B.

Make configuration to achieve the following:

- VLAN 10 frames of Customer A and Customer B can be forwarded to each other across SVLAN 1000;
- VLAN 20 frames of Customer A and Customer C can be forwarded to each other across SVLAN 2000.

Figure 1-5 Network diagram for comprehensive selective QinQ configuration



Configuration procedure

Note

Make sure that the devices in the service provider network have been configured to allow QinQ packets to pass through.

1) Configuration on Provider A

- Configure GigabitEthernet 1/0/1

Configure GigabitEthernet 1/0/1 as a hybrid port to permit frames of VLAN 1000 and VLAN 2000 to pass through, and configure GigabitEthernet 1/0/1 to send packets of these VLANs with tags removed.

```
<ProviderA> system-view
```

```
[ProviderA] interface gigabitethernet 1/0/1
[ProviderA-GigabitEthernet1/0/1] port link-type hybrid
[ProviderA-GigabitEthernet1/0/1] port hybrid vlan 1000 2000 untagged
```

Tag CVLAN 10 frames with SVLAN 1000.

```
[ProviderA-GigabitEthernet1/0/1] qinq vid 1000
[ProviderA-GigabitEthernet1/0/1-vid-1000] raw-vlan-id inbound 10
[ProviderA-GigabitEthernet1/0/1-vid-1000] quit
```

Tag CVLAN 20 frames with SVLAN 2000.

```
[ProviderA-GigabitEthernet1/0/1] qinq vid 2000
[ProviderA-GigabitEthernet1/0/1-vid-2000] raw-vlan-id inbound 20
[ProviderA-GigabitEthernet1/0/1-vid-2000] quit
[ProviderA-GigabitEthernet1/0/1] quit
```

- Configure GigabitEthernet 1/0/2

Configure GigabitEthernet 1/0/2 as a hybrid port to permit frames of VLAN 1000 to pass through, and configure GigabitEthernet 1/0/2 to send packets of VLAN 1000 with tag removed.

```
[ProviderA] interface gigabitethernet 1/0/2
[ProviderA-GigabitEthernet1/0/2] port link-type hybrid
[ProviderA-GigabitEthernet1/0/2] port hybrid vlan 1000 untagged
```

Tag CVLAN 10 frames with SVLAN 1000.

```
[ProviderA-GigabitEthernet1/0/2] qinq vid 1000
[ProviderA-GigabitEthernet1/0/2-vid-1000] raw-vlan-id inbound 10
[ProviderA-GigabitEthernet1/0/2-vid-1000] quit
[ProviderA-GigabitEthernet1/0/2] quit
```

- Configure GigabitEthernet 1/0/3

Configure GigabitEthernet 1/0/3 as a trunk port to permit frames of VLAN 1000 and VLAN 2000 to pass through.

```
[ProviderA] interface gigabitethernet 1/0/3
[ProviderA-GigabitEthernet1/0/3] port link-type trunk
[Sysname-GigabitEthernet1/0/3] port trunk permit vlan 1000 2000
```

Set the TPID value in the outer tag to 0x8200.

```
[ProviderA-GigabitEthernet1/0/3] quit
[ProviderA] qinq ethernet-type service-tag 8200
```

2) Configuration on Provider B

- Configure GigabitEthernet 1/0/1

Configure GigabitEthernet 1/0/1 as a trunk port to permit frames of VLAN 1000 and VLAN 2000 to pass through.

```
<ProviderB> system-view
[ProviderB] interface gigabitethernet 1/0/1
[ProviderB-GigabitEthernet1/0/1] port link-type trunk
[ProviderB-GigabitEthernet1/0/1] port trunk permit vlan 1000 2000
```

- Configure GigabitEthernet 1/0/2

Configure GigabitEthernet 1/0/2 as a hybrid port to permit frames of VLAN 2000 to pass through, and configure GigabitEthernet 1/0/2 to send packets of VLAN 2000 with tag removed.

```
[ProviderB] interface gigabitethernet 1/0/2
```

```
[ProviderB-GigabitEthernet1/0/2] port link-type hybrid
[ProviderB-GigabitEthernet1/0/2] port hybrid vlan 2000 untagged
```

Tag CVLAN 20 frames with SVLAN 2000.

```
[ProviderB-GigabitEthernet1/0/2] qinq vid 2000
[ProviderB-GigabitEthernet1/0/2-vid-2000] raw-vlan-id inbound 20
```

Set the TPID value in the outer tag to 0x8200.

```
[ProviderA-GigabitEthernet1/0/3] quit
[ProviderA] qinq ethernet-type service-tag 8200
```

3) Configuration on third-party devices

Configure the third-party devices between Provider A and Provider B as follows: configure the port connecting GigabitEthernet 1/0/3 of Provider A and that connecting GigabitEthernet 1/0/1 of Provider B to allow tagged frames of VLAN 1000 and VLAN 2000 to pass through.

Table of Contents

1 BPDU Tunneling Configuration

When configuring BPDU tunneling, go to these sections for information you are interested in:

Introduction to BPDU Tunneling

To avoid loops in your network, you can enable the Spanning Tree Protocol (STP) on your device. Here, the term STP is in a broad sense. It includes STP, RSTP, and MSTP. STP calculates the topology of a network by multicasting bridge protocol data units (BPDUs) at Layer 2. As these BPDUs can be received and processed by all STP-enabled devices, this prevents each network from correctly calculating its independent spanning tree.

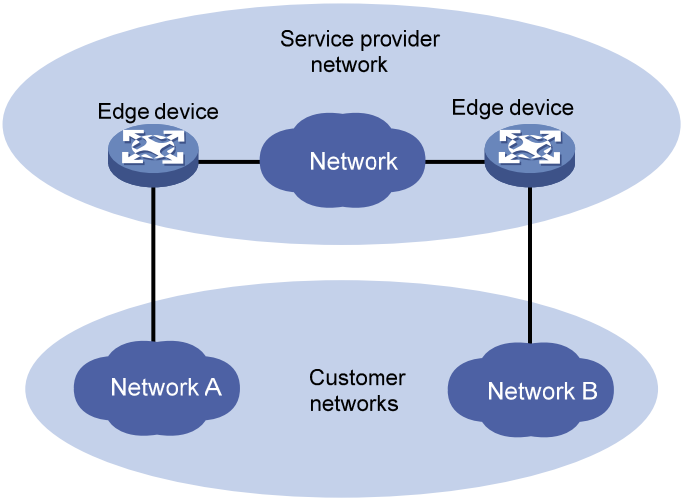
To allow each network to calculate an independent spanning tree with STP, BPDU tunneling was introduced.

BPDU tunneling delivers the following benefits:

- BPDUs can be transmitted transparently. BPDUs of the same customer network can be broadcast in a specific VLAN across the service provider network, so that the geographically dispersed networks of the same customer can implement consistent spanning tree calculation across the service provider network.
- BPDUs of different customer networks can be confined within different VLANs for transmission on the service provider network. Thus, each customer network can perform independent spanning tree calculation.

As shown in Figure 1-1, the upper part is the service provider network, and the lower part represents the customer networks. The customer networks include network A and network B. Enabling the BPDU tunneling function on the edge devices across the service provider network allows BPDUs of the customer networks to be transparently transmitted in the service provider network, and allows each customer network to implement independent spanning tree calculation, without affecting each other.

Figure 1-1 Network hierarchy of BPDU tunneling



- At the input side of the service provider network, the edge device changes the destination MAC address of a BPDU from a customer network from 0x0180-C200-0000 to a special multicast MAC address, 0x010F-E200-0003 by default. In the service provider's network, the modified BPDUs are forwarded as data packets in the user VLAN.
- At the output side of the service provider network, the edge device recognizes the BPDU with the destination MAC address of 0x010F-E200-0003 and restores its original destination MAC address 0x0180-C200-0000. Then, the device removes the outer VLAN tag, and sends the BPDU to the destination customer network.

Note

Make sure, through configuration, that the VLAN tag of the BPDU is neither changed nor removed during its transparent transmission in the service provider network; otherwise, the system will fail to transparently transmit the customer network BPDU correctly.

Configuring BPDU Transparent Transmission

Perform the following tasks to configure BPDU transparent transmission:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. • Settings made in interface view take effect only on the current port. • Settings made in Layer-2 aggregate interface view take effect only on the Layer-2 aggregate interface. • Settings made in port group view take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Disable STP on the port(s)		**undo stp enable**	Required
Enable BPDU tunneling for STP on the port(s)		**bpdu-tunnel dot1q stp**	Required By default, BPDU tunneling for STP is disabled.

Configuring Destination Multicast MAC Address for BPDU Tunnel Frames

By default, the destination multicast MAC address for BPDU tunnel frames is 0x010F-E200-0003. You can modify it to 0x0100-0CCD-CDD0, 0x0100-0CCD-CDD1 or 0x0100-0CCD-CDD2 through the following configuration.

Follow these steps to configure destination multicast MAC address for BPDU tunnel frames:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the destination multicast MAC address for BPDU tunnel frames	**bpdu-tunnel tunnel-dmac** *mac-address*	Optional 0x010F-E200-0003 by default.

Note

For BPDU tunnel frames to be recognized, the destination multicast MAC addresses configured for BPDU tunneling must be the same on the edge devices on the service provider network.

BPDU Tunneling Configuration Example

Network requirements

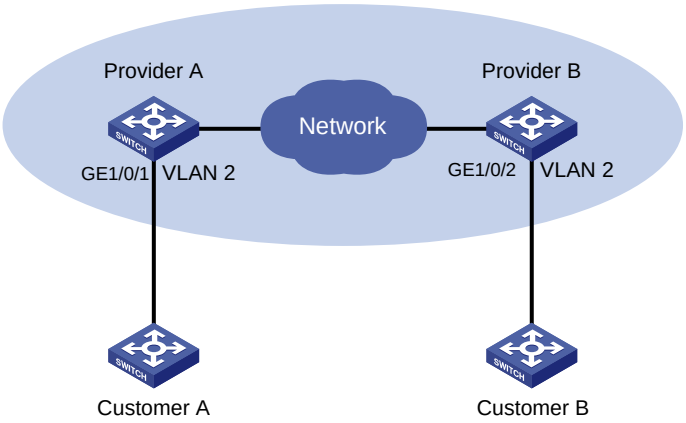
- Customer A and Customer B are customer network edge devices.

- Provider A and Provider B are service provider network edge devices, which are interconnected through configured trunk ports.

The configuration is required to satisfy the following requirements:

- Geographically dispersed customer network access devices Customer A and Customer B can implement consistent spanning tree calculation across the service provider network.
- The destination multicast MAC address configured for BPDU tunnel frames is 0x0100-0CCD-CDD0.

Figure 1-2 Network diagram for BPDU tunneling configuration



Configuration procedure

1) Configuration on Provider A

Configure the destination multicast MAC address for BPDU tunnel frames as 0x0100-0CCD-CDD0.

```
<ProviderA> system-view
[ProviderA] bpdu-tunnel tunnel-dmac 0100-0ccd-cdd0
```

Configure GigabitEthernet 1/0/1 to transmit packets through VLAN 2.

```
[ProviderA] vlan 2
[ProviderA-vlan2] quit
[ProviderA] interface GigabitEthernet 1/0/1
[ProviderA-GigabitEthernet1/0/1] port access vlan 2
```

Configure GigabitEthernet 1/0/1 to transmit BPDUs transparently.

```
[ProviderA-GigabitEthernet1/0/1] undo stp enable
[ProviderA-GigabitEthernet1/0/1] bpdu-tunnel dot1q stp
```

2) Configuration on Provider B

Configure the destination multicast MAC address for BPDU tunnel frames as 0x0100-0CCD-CDD0.

```
<ProviderB> system-view
[ProviderB] bpdu-tunnel tunnel-dmac 0100-0ccd-cdd0
```

Configure GigabitEthernet 1/0/2 to transmit packets through VLAN 2.

```
[ProviderB] vlan 2
[ProviderB-vlan2] quit
[ProviderB] interface GigabitEthernet 1/0/2
[ProviderB-GigabitEthernet1/0/2] port access vlan 2
```

Configure GigabitEthernet 1/0/2 to transmit BPDUs transparently.

```
[ProviderB-GigabitEthernet1/0/2] undo stp enable
[ProviderB-GigabitEthernet1/0/2] bpdu-tunnel dot1q stp
```

Table of Contents

1 VLAN Mapping Configuration

When configuring VLAN mapping, go to these sections for information you are interested in:

VLAN Mapping Overview

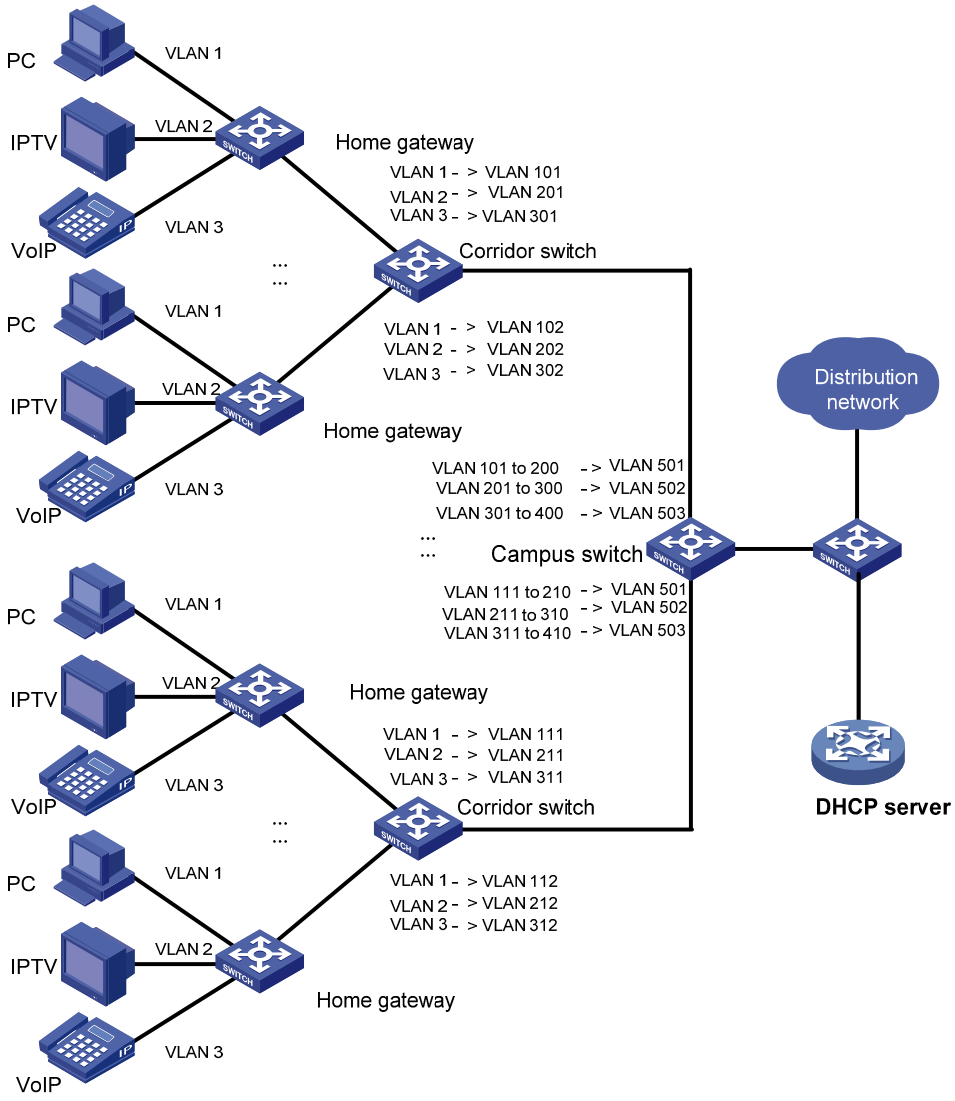
VLAN mapping maps the customer VLANs (CVLANs) to service-provider VLANs (SVLANs). Types of VLAN mapping include:

- One-to-one VLAN mapping that maps the CVLAN ID in the VLAN tag to the SVLAN ID.
- Many-to-one VLAN mapping that maps the CVLAN IDs in the VLAN tags of traffic of more than two VLANs to the same SVLAN ID.
- Two-to-two VLAN mapping that maps traffic with outer and inner VLAN IDs to the service-provider outer and the inner VLAN IDs.

The following sections present the scenario to which these VLAN mapping types apply.

One-to-One VLAN Mapping and Many-to-One VLAN Mapping

Figure 1-1 Scenario for one-to-one/many-to-one VLAN mapping

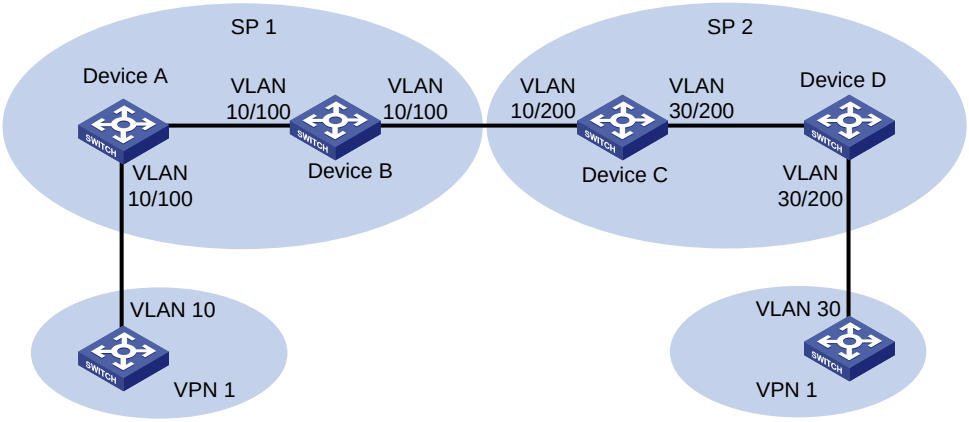


One-to-one VLAN mapping and many-to-one VLAN mapping are mainly applied in networking environments as shown in Figure 1-1. In such a network, different VLANs are used for transmitting different services (PC, IPTV, and VoIP for example) of a home user. Furthermore, to differentiate home users that are using the same service, you need to perform one-to-one VLAN mapping to map the service traffic to different VLANs by user on the corridor switches. However, an access device on the distribution layer is likely unable to support the number of VLANs required for this type of VLAN mapping.

To reduce the number of VLANs required on the edge device at the distribution layer, you can adopt many-to-one VLAN mapping. This type of VLAN mapping maps the VLANs carrying the same service of different users to the same VLAN while isolating the service traffic of different users.

Two-to-Two VLAN Mapping

Figure 1-2 Scenario for two-to-two VLAN mapping



Two-to-two VLAN mapping are mainly applied in networking environments as shown in Figure 1-2. In Figure 1-2, two VPN 1 users located in different regions communicate with each other across two service provider (SP) networks: SP 1 and SP 2.

Assuming that the VPN 1 user using VLAN 10 sends a packet to the VPN 1 user using VLAN 30, the following describes how this packet is processed on its way to the destination:

1) When the packet tagged with VLAN 10 arrives at the edge of the SP 1 network, Device A tags the packet with VLAN 100, the VLAN ID assigned to the VPN 1 user in SP 1 by using QinQ or Selective QinQ. The packet thus becomes double-tagged.
2) When the double-tagged packet enters the SP 2 network, Device C replaces the outer VLAN tag (VLAN 100) with VLAN 200, the VLAN ID assigned by SP 2 to the VPN 1 user. For the packet to reach the VPN 1 user in VLAN 30, Device C replaces the inner tag (VLAN 10) of the packet with VLAN 30. This double-tag to double-tag replacement is called two-to-two VLAN mapping.

Note

For more information about QinQ and Selective QinQ, refer to *QinQ Configuration* in the *Access Volume*.

Basic Concepts of VLAN Mapping

Before you configure VLAN mappings, be aware of the following concepts, which will be used throughout this document.

- Uplink traffic: Traffic transmitted from a home user network to a distribution network or from a user network to an SP network.
- Downlink traffic: Traffic transmitted from a distribution network to a home user network or from an SP network to a user network.
- Uplink port: A port transmitting uplink traffic and receiving downlink traffic.
- Downlink port: A port transmitting downlink traffic and receiving uplink traffic.

- Uplink policy: A QoS policy containing VLAN mappings for uplink traffic.
- Downlink policy: A QoS policy containing VLAN mappings for downlink traffic.

How VLAN Mapping Is Implemented

This section describes how VLAN mapping is implemented on your device.

One-to-one VLAN mapping

On the downlink port			
For uplink traffic		For downlink traffic	
Do…	Based on…	Do…	Based on…
Replace the customer VLAN (CVLAN) with the service provider VLAN (SVLAN)	Uplink policy in the inbound direction	Replace the SVLAN with the original CVLAN	Downlink policy in the outbound direction

Note

For information about QoS policies, refer to *QoS Configuration* in the *QoS Volume*.

Many-to-one VLAN mapping

On the downlink port		On the uplink port	
For uplink traffic		For downlink traffic	
Do...	Based on…	Do...	Based on…
Map all specified customer VLANs (CVLANs) to one service provider VLAN (SVLAN)	Uplink policy in the inbound direction	Replace the SVLAN with the original CVLAN	DHCP snooping address table, which contains mappings of the SVLAN, IP address, MAC address, and CVLAN for DHCP clients

Note

- For information about DHCP snooping, refer to *DHCP Configuration* in the *IP Services Volume*.
- For information about QoS policies, refer to *QoS Configuration* in the *QoS Volume*.

Two-to-two VLAN mapping

Note

In two-to-two VLAN mapping, the outer VLAN and the inner VLAN carried in a double-tagged uplink frame received at the downlink port on the edge device of an SP network are called the original SVLAN and CVLAN, and the VLANs that the edge device substitutes for the original SVLAN and CVLAN are called the new SVLAN and CVLAN.

On the downlink port				On the uplink port	
For uplink traffic		For downlink traffic		For uplink traffic	
Do...	Based on…	Do...	Based on…	Do...	Based on…
Replace the original SVLAN with the new SVLAN	Uplink policy in the inbound direction	Replace the new SVLAN and CVLAN with the original SVLAN and CVLAN	Downlink policy in the outbound direction	Replace the original CVLAN with the new CVLAN	Uplink policy in the outbound direction

VLAN Mapping Configuration Task List

You need to configure VLAN mapping on your device depending on its position in the network.

Complete the following tasks to configure VLAN mapping:

Task	Remarks
Configuring One-to-One VLAN Mapping	Optional Perform this configuration on the corridor switches shown in Figure 1-1.
Configuring Many-to-One VLAN Mapping	Optional Perform this configuration on the campus switches shown in Figure 1-1.
Configuring Two-to-Two VLAN Mapping	Optional Perform this configuration on an edge device connecting two SP networks. An example is Device C in the SP 2 network in Figure 1-2.

For VLAN mapping to work, you are required to do the following in addition to configuring QoS policies:

- Enable ARP detection to send ARP packets to the CPU to allow modification of the VLAN attributes carried in the packets, which is impossible with the normal ARP packet processing procedure. For information about ARP detection, refer to ARP Configuration in the IP Services Volume.
- Enable the dynamic address binding support of IP Source Guard to filter packets received on a port based on the source IP address and MAC address bindings created dynamically to prevent illegal packets from passing through the port. For information about this feature, refer to IP Source Guard Configuration in the Security Volume.

- For many-to-one VLAN mapping, enable customer-side QinQ on the downlink port and service provider-side QinQ on the uplink port.
- To save system resources, disable user bindings recording on the DHCP snooping trusted ports that forward DHCP packets. For information about this feature, refer to DHCP Configuration in the IP Services Volume.

Configuring One-to-One VLAN Mapping

Perform one-to-one VLAN mapping on the corridor switches shown in Figure 1-1 to use VLANs to isolate different services of different users.

Configuring One-to-One VLAN Mapping

Configuration prerequisites

The CVLAN-to-SVLAN mappings have been planned.

Configuration procedure

Follow these steps to configure a one-to-one VLAN mapping:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Create a CVLAN and a SVLAN	Create a VLAN	**vlan** *vlan-id*	Required By default, only the default VLAN (VLAN 1) exists. Repeat these steps for all CVLANs and SVLANs involved in VLAN mapping.
	Exit to system view	**quit**	
Configure an uplink policy to map the CVLAN to the SVLAN		Refer to Table 1-1	Required
Configure a downlink policy to map the SVLAN to the original CVLAN		Refer to Table 1-2	Required
Enter interface view of the downlink port		**interface** *interface-type interface-number*	—
Set the link type of the downlink port to trunk		**port link-type trunk**	Required
Configure the downlink port to permit the specified CVLANs and SVLANs or all VLANs to pass through		**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only VLAN 1 to pass through.
Enable basic QinQ on the port		**qinq enable**	Required
Apply the uplink policy to the inbound direction of the downlink port		**qos apply policy** *policy-name* **inbound**	Required
Apply the downlink policy to the outbound direction of the downlink port		**qos apply policy** *policy-name* **outbound**	Required
Exit to system view		**quit**	—
Enter the interface view of the uplink port		**interface** *interface-type interface-number*	—

To do...	Use the command...	Remarks
Set the link type of the uplink port to trunk	**port link-type trunk**	Required
Configure the uplink port to permit the specified SVLANs to pass through	**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only VLAN 1 to pass through.

Table 1-1 Configure an uplink policy

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Specify the CVLAN for the VLAN mapping	**if-match customer-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the SVLAN for the VLAN mapping	**remark service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required
Map the CVLAN to the SVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name*	Required
Exit to system view	**quit**	—

Table 1-2 Configure a downlink policy

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Specify the SVLAN for the VLAN mapping	**if-match service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the CVLAN for the VLAN mapping	**remark customer-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required

To do...	Use the command...	Remarks
Map the SVLAN to the CVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name*	Required
Exit to system view	**quit**	—

Configuring Many-to-One VLAN Mapping

Perform many-to-one VLAN mapping on the campus switches shown in Figure 1-1 to carry the same service of different users using the same VLAN on the service provider's network.

Configuring Many-to-One VLAN Mapping

Configuration prerequisites

- All service terminals of home users are using DHCP for obtaining an IP address. For how to get an IP address through DHCP, refer to DHCP Configuration in the IP Services Volume.
- The CVLAN-to-SVLAN mappings have been planned.

Configuration procedure

Follow these steps to configure a many-to-one VLAN mapping:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enable DHCP snooping		**dhcp-snooping**	Required Disabled by default.
Enable ARP detection on the CVLANs and the SVLAN for the VLAN mapping	Create a VLAN and enter VLAN view	**vlan** *vlan-id*	Required Disabled by default. Repeat these steps for all CVLANs and the SVLAN that the VLAN mapping involves.
	Enable ARP detection	**arp detection enable**	
	Exit to system view	**quit**	
Configure an uplink policy to map the CVLANs to the same SVLAN		Refer to Table 1-3	Required
Enter the interface view of the downlink port		**interface** *interface-type interface-number*	—
Set the link type of the downlink port to trunk		**port link-type trunk**	Required
Configure the downlink port to permit the specified CVLANs and SVLANs to pass through		**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only VLAN 1 to pass through.
Enable customer side QinQ		**qinq enable downlink**	Required Disabled by default.
Apply the uplink policy to the downlink port in the inbound direction		**qos apply policy** *policy-name* **inbound**	Required

To do...	Use the command...	Remarks
Exit to system view	**quit**	—
Enter the interface view of the uplink port	**interface** *interface-type interface-number*	—
Configure the uplink port as a DHCP snooping trusted port	**dhcp-snooping trust**	Required By default, all ports with DHCP snooping enabled are DHCP snooping untrusted ports.
Configure the uplink port as an ARP trusted port	**arp detection trust**	Required By default, all ports are ARP untrusted ports.
Set the link type of the uplink port to trunk	**port link-type trunk**	Required
Configure the uplink port to permit the specified SVLANs to pass through	**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only VLAN 1 to pass through.
Enable service provider side QinQ	**qinq enable uplink**	Required Disabled by default.

Caution

- To defend against attacks, you are recommended to enable ARP detection for each CVLAN.
- Before applying a QoS policy to the downlink port, enable customer-side QinQ on the port; before disabling customer-side QinQ on the downlink port, remove the QoS policy.
- To change a VLAN mapping, you must first use the **reset dhcp-snooping** command to clear the corresponding DHCP snooping address binding entry (refer to *DHCP Commands* in the *IP Services Volume*) or disable the dynamic address binding function of IP Source Guard on the downlink port and then enable dynamic address binding again (refer to *IP Source Guard Commands* in the *Security Volume*).
- You can configure the **qinq enable uplink** command or the **qinq enable downlink** command in port group view to make it take effect on all ports in the port group.

Table 1-3 Configure an uplink policy

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* **operator or**	Required
Specify the CVLANs for the VLAN mapping	**if-match customer-vlan-id** { *vlan-id-list* | *vlan-id1* **to** *vlan-id2* }	Required
Exit to system view	**quit**	—

To do...	Use the command...	Remarks
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the SVLAN for the VLAN mapping	**remark service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required
Map the CVLANs to the SVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name* **mode dot1q-tag-manipulation**	Required
Exit to system view	**quit**	—

Configuring Two-to-Two VLAN Mapping

Note

In two-to-two VLAN mapping, the outer VLAN and the inner VLAN carried in a double-tagged uplink frame received at the downlink port on the edge device of an SP network are called the original SVLAN and CVLAN, and the VLANs that the edge device substitutes for the original SVLAN and CVLAN are called the new SVLAN and CVLAN.

Perform two-to-two VLAN mapping on the edge device that connects two SP networks, on Device C in Figure 1-2 for example.

Follow these steps to configure a two-to-two VLAN mapping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure an uplink policy for the uplink port to replace the original CVLAN with the new CVLAN	Refer to Table 1-4.	Required
Configure an uplink policy for the downlink port to replace the original SVLAN with the new SVLAN	Refer to Table 1-5.	Required
Configure a downlink policy for the downlink port to replace the new SVLAN and CVLAN with the original SVLAN and CVLAN	Refer to Table 1-6.	Required
Enter the interface view of the downlink port	**interface** *interface-type interface-number*	—
Configure the downlink port as a trunk port	**port link-type trunk**	Required

To do...	Use the command...	Remarks
Configure the downlink port to permit the packets of the SVLANs to pass through	**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only the packets of VLAN 1 to pass through.
Apply the uplink policy for the downlink port to the inbound direction of the downlink port	**qos apply policy** *policy-name* **inbound**	Required
Apply the downlink policy for the downlink port to the outbound direction of the downlink port	**qos apply policy** *policy-name* **outbound**	Required
Exit to system view	**quit**	—
Enter the interface view of the uplink port	**interface** *interface-type interface-number*	—
Configure the uplink port as a trunk port	**port link-type trunk**	Required
Configure the uplink port to permit the packets of the SVLANs to pass through	**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, a trunk port permits only the packets of VLAN 1 to pass through.
Apply the uplink policy for the uplink port to the outbound direction of the uplink port	**qos apply policy** *policy-name* **outbound**	Required

Table 1-4 Configure an uplink policy for the uplink port

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Specify the original CVLAN for the VLAN mapping	**if-match customer-vlan-id** *vlan-id-value*	Required
Specify the new SVLAN for the VLAN mapping	**if-match service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the new CVLAN used for replacing the original CVLAN	**remark customer-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required

To do...	Use the command...	Remarks
Map the original CVLAN and the new SVLAN to the new CVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name*	Required
Exit to system view	**quit**	—

Table 1-5 Configure an uplink policy for the downlink port

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Specify the original CVLAN for the VLAN mapping	**if-match customer-vlan-id** *vlan-id-value*	Required
Specify the original SVLAN for the VLAN mapping	**if-match service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the new SVLAN used for replacing the original SVLAN	**remark service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required
Map the original SVLAN and CVLAN to the new SVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name*	Required
Exit to system view	**quit**	—

Table 1-6 Configure a downlink policy for the downlink port

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create a class and enter class view	**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Specify the new CVLAN for the VLAN mapping	**if-match customer-vlan-id** *vlan-id-value*	Required
Specify the new SVLAN for the VLAN mapping	**if-match service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Specify the original CVLAN used for replacing the new CVLAN	**remark customer-vlan-id** *vlan-id-value*	Required

To do...	Use the command...	Remarks
Specify the original SVLAN used for replacing the new SVLAN	**remark service-vlan-id** *vlan-id-value*	Required
Exit to system view	**quit**	—
Create a QoS policy and enter QoS policy view	**qos policy** *policy-name*	Required
Map the new CVLAN and SVLAN to the original CVLAN and SVLAN by associating the traffic class with the traffic behavior	**classifier** *tcl-name* **behavior** *behavior-name*	Required
Exit to system view	**quit**	—

VLAN Mapping Configuration Examples

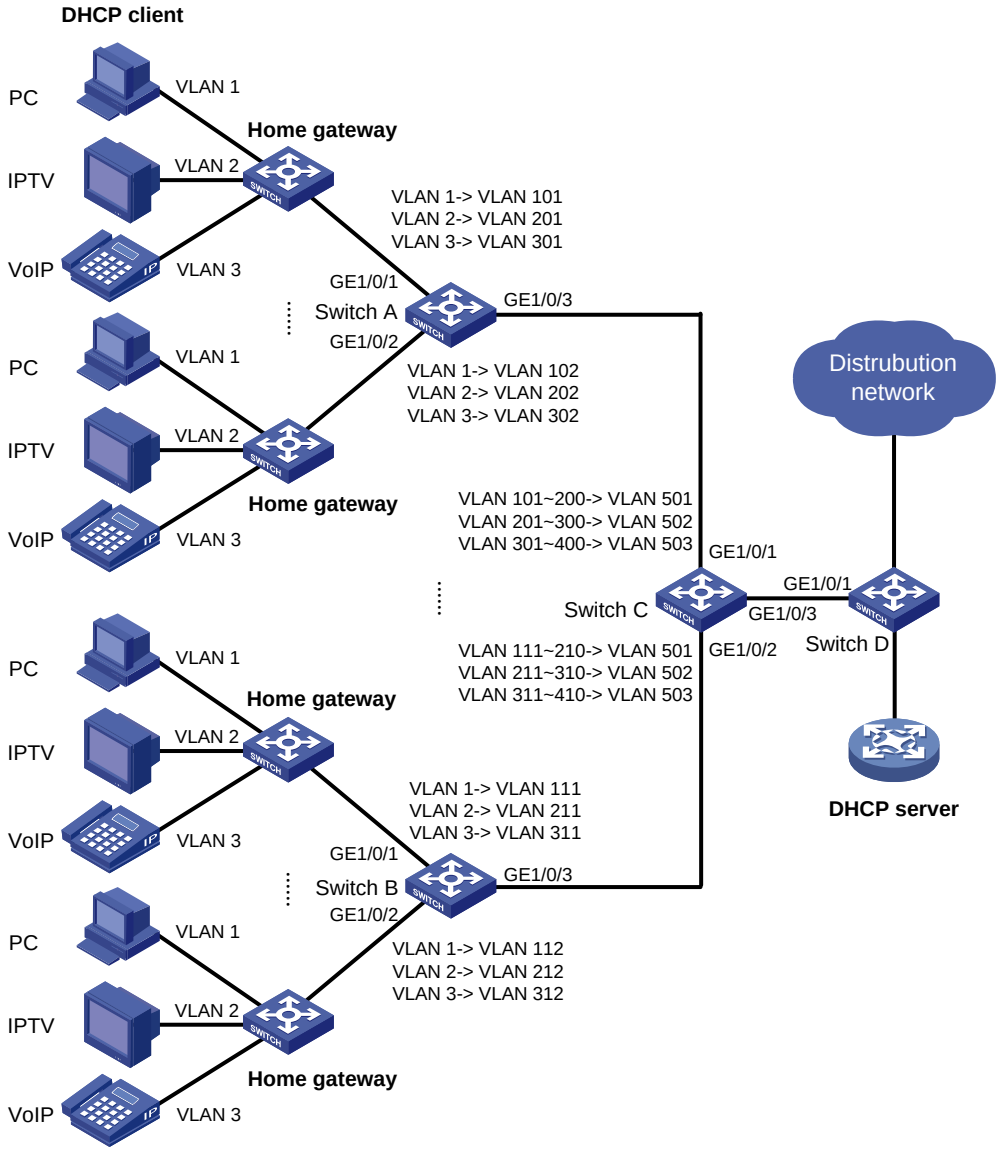
One-to-One/Many-to-One VLAN Mapping Configuration Example

Network requirements

To save VLAN resources, use one VLAN to carry a type of service traffic from Switch C at the campus network edge while isolating the traffic of a home user from the traffic of all other home users in the VLAN.

Use VLAN 501 for PC traffic, VLAN 502 for IPTV traffic, and VLAN 503 for VoIP traffic.

Figure 1-3 Network diagram for one-to-one/many-to-one VLAN mapping configuration



Configuration procedure

1) Configuration on Switch A

Create the CVLANs and the SVLANs.

```
<SwitchA> system-view
[SwitchA] vlan 2 to 3
[SwitchA] vlan 101 to 102
[SwitchA] vlan 201 to 202
[SwitchA] vlan 301 to 302
```

Configure uplink policies to map the CVLANs to the SVLANs.

```
[SwitchA] traffic classifier c1
[SwitchA-classifier-c1] if-match customer-vlan-id 1
[SwitchA-classifier-c1] traffic classifier c2
[SwitchA-classifier-c2] if-match customer-vlan-id 2
[SwitchA-classifier-c2] traffic classifier c3
[SwitchA-classifier-c3] if-match customer-vlan-id 3
```

```
[SwitchA-classifier-c3] quit
[SwitchA] traffic behavior b1
[SwitchA-behavior-b1] remark service-vlan-id 101
[SwitchA-behavior-b1] traffic behavior b2
[SwitchA-behavior-b2] remark service-vlan-id 201
[SwitchA-behavior-b2] traffic behavior b3
[SwitchA-behavior-b3] remark service-vlan-id 301
[SwitchA-behavior-b3] traffic behavior b4
[SwitchA-behavior-b4] remark service-vlan-id 102
[SwitchA-behavior-b4] traffic behavior b5
[SwitchA-behavior-b5] remark service-vlan-id 202
[SwitchA-behavior-b5] traffic behavior b6
[SwitchA-behavior-b6] remark service-vlan-id 302
[SwitchA-behavior-b6] quit
[SwitchA] qos policy p1
[SwitchA-policy-p1] classifier c1 behavior b1
[SwitchA-policy-p1] classifier c2 behavior b2
[SwitchA-policy-p1] classifier c3 behavior b3
[SwitchA-policy-p1] quit
[SwitchA] qos policy p2
[SwitchA-policy-p2] classifier c1 behavior b4
[SwitchA-policy-p2] classifier c2 behavior b5
[SwitchA-policy-p2] classifier c3 behavior b6
[SwitchA-policy-p2] quit
```

Configure downlink policies to map the SVLANs to the original CVLANs.

```
[SwitchA] traffic classifier c11
[SwitchA-classifier-c11] if-match service-vlan-id 101
[SwitchA-classifier-c11] traffic classifier c22
[SwitchA-classifier-c22] if-match service-vlan-id 201
[SwitchA-classifier-c22] traffic classifier c33
[SwitchA-classifier-c33] if-match service-vlan-id 301
[SwitchA-classifier-c33] traffic classifier c44
[SwitchA-classifier-c44] if-match service-vlan-id 102
[SwitchA-classifier-c44] traffic classifier c55
[SwitchA-classifier-c55] if-match service-vlan-id 202
[SwitchA-classifier-c55] traffic classifier c66
[SwitchA-classifier-c66] if-match service-vlan-id 302
[SwitchA-classifier-c66] quit
[SwitchA] traffic behavior b11
[SwitchA-behavior-b11] remark customer-vlan-id 1
[SwitchA-behavior-b11] traffic behavior b22
[SwitchA-behavior-b22] remark customer-vlan-id 2
[SwitchA-behavior-b22] traffic behavior b33
[SwitchA-behavior-b33] remark customer-vlan-id 3
 [SwitchA-behavior-b33] quit
[SwitchA] qos policy p11
[SwitchA-policy-p11] classifier c11 behavior b11
```

```
[SwitchA-policy-p11] classifier c22 behavior b22
[SwitchA-policy-p11] classifier c33 behavior b33
[SwitchA-policy-p11] quit
[SwitchA] qos policy p22
[SwitchA-policy-p22] classifier c44 behavior b11
[SwitchA-policy-p22] classifier c55 behavior b22
[SwitchA-policy-p22] classifier c66 behavior b33
[SwitchA-policy-p22] quit
```

Configure GigabitEthernet 1/0/1 to permit frames of the specified CVLANs and SLVANs to pass through.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-type trunk
[SwitchA-GigabitEthernet1/0/1] port trunk permit vlan 1 2 3 101 201 301
```

Enable basic QinQ on GigabitEthernet 1/0/1.

```
[SwitchA-GigabitEthernet1/0/1] qinq enable
```

Apply the uplink policy **p1** to the inbound direction of GigabitEthernet 1/0/1.

```
[SwitchA-GigabitEthernet1/0/1] qos apply policy p1 inbound
```

Apply the downlink policy **p11** to the outbound direction of GigabitEthernet 1/0/1.

```
[SwitchA-GigabitEthernet1/0/1] qos apply policy p11 outbound
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to permit frames of the specified CVLANs and SVLANs to pass through.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-type trunk
[SwitchA-GigabitEthernet1/0/2] port trunk permit vlan 1 2 3 102 202 302
```

Enable basic QinQ on GigabitEthernet 1/0/2.

```
[SwitchA-GigabitEthernet1/0/2] qinq enable
```

Apply the uplink policy **p2** to the inbound direction of GigabitEthernet 1/0/2.

```
[SwitchA-GigabitEthernet1/0/2] qos apply policy p2 inbound
```

Apply the downlink policy **p22** to the outbound direction of GigabitEthernet 1/0/2.

```
[SwitchA-GigabitEthernet1/0/2] qos apply policy p22 outbound
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 to permit frames of the specified SVLANs to pass through.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-type trunk
[SwitchA-GigabitEthernet1/0/3] port trunk permit vlan 101 201 301 102 202 302
```

2) Configuration on Switch B

Create the CVLANs and the SVLANs.

```
<SwitchB> system-view
[SwitchB] vlan 2 to 3
[SwitchB] vlan 111 to 112
[SwitchB] vlan 211 to 212
[SwitchB] vlan 311 to 312
```

Configure uplink policies to map the CVLANs to the SVLANs.

```
[SwitchB] traffic classifier c1
[SwitchB-classifier-c1] if-match customer-vlan-id 1
[SwitchB-classifier-c1] traffic classifier c2
[SwitchB-classifier-c2] if-match customer-vlan-id 2
[SwitchB-classifier-c2] traffic classifier c3
[SwitchB-classifier-c3] if-match customer-vlan-id 3
 [SwitchB-classifier-c3] quit
[SwitchB] traffic behavior b1
[SwitchB-behavior-b1] remark service-vlan-id 111
[SwitchB-behavior-b1] traffic behavior b2
[SwitchB-behavior-b2] remark service-vlan-id 211
[SwitchB-behavior-b2] traffic behavior b3
[SwitchB-behavior-b3] remark service-vlan-id 311
[SwitchB-behavior-b3] traffic behavior b4
[SwitchB-behavior-b4] remark service-vlan-id 112
[SwitchB-behavior-b4] traffic behavior b5
[SwitchB-behavior-b5] remark service-vlan-id 212
[SwitchB-behavior-b5] traffic behavior b6
[SwitchB-behavior-b6] remark service-vlan-id 312
[SwitchB-behavior-b6] quit
[SwitchB] qos policy p1
[SwitchB-policy-p1] classifier c1 behavior b1
[SwitchB-policy-p1] classifier c2 behavior b2
[SwitchB-policy-p1] classifier c3 behavior b3
[SwitchB-policy-p1] quit
[SwitchB] qos policy p2
[SwitchB-policy-p2] classifier c1 behavior b4
[SwitchB-policy-p2] classifier c2 behavior b5
[SwitchB-policy-p2] classifier c3 behavior b6
[SwitchB-policy-p2] quit
```

Configure downlink policies to map the SVLANs to the original CVLANs.

```
[SwitchB] traffic classifier c11
[SwitchB-classifier-c11] if-match service-vlan-id 111
[SwitchB-classifier-c11] traffic classifier c22
[SwitchB-classifier-c22] if-match service-vlan-id 211
[SwitchB-classifier-c22] traffic classifier c33
[SwitchB-classifier-c33] if-match service-vlan-id 311
[SwitchB-classifier-c33] traffic classifier c44
[SwitchB-classifier-c44] if-match service-vlan-id 112
[SwitchB-classifier-c44] traffic classifier c55
[SwitchB-classifier-c55] if-match service-vlan-id 212
[SwitchB-classifier-c55] traffic classifier c66
[SwitchB-classifier-c66] if-match service-vlan-id 312
[SwitchB-classifier-c66] quit
[SwitchB] traffic behavior b11
[SwitchB-behavior-b11] remark customer-vlan-id 1
```

```
[SwitchB-behavior-b11] traffic behavior b22
[SwitchB-behavior-b22] remark customer-vlan-id 2
[SwitchB-behavior-b22] traffic behavior b33
[SwitchB-behavior-b33] remark customer-vlan-id 3
 [SwitchB-behavior-b33] quit
[SwitchB] qos policy p11
[SwitchB-policy-p11] classifier c11 behavior b11
[SwitchB-policy-p11] classifier c22 behavior b22
[SwitchB-policy-p11] classifier c33 behavior b33
[SwitchB-policy-p11] quit
[SwitchB] qos policy p22
[SwitchB-policy-p22] classifier c44 behavior b11
[SwitchB-policy-p22] classifier c55 behavior b22
[SwitchB-policy-p22] classifier c66 behavior b33
[SwitchB-policy-p22] quit
```

Configure GigabitEthernet 1/0/1 to permit frames of the specified CVLANs and SVLANs to pass through.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-type trunk
[SwitchB-GigabitEthernet1/0/1] port trunk permit vlan 1 2 3 111 211 311
```

Enable basic QinQ on GigabitEthernet 1/0/1.

```
[SwitchB-GigabitEthernet1/0/1] qinq enable
```

Apply the uplink policy **p1** to the inbound direction of GigabitEthernet 1/0/1.

```
[SwitchB-GigabitEthernet1/0/1] qos apply policy p1 inbound
```

Apply the downlink policy **p11** to the outbound direction of GigabitEthernet 1/0/1.

```
[SwitchB-GigabitEthernet1/0/1] qos apply policy p11 outbound
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to permit frames of the specified CVLANs and SLVANs to pass through.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type trunk
[SwitchB-GigabitEthernet1/0/2] port trunk permit vlan 1 2 3 112 212 312
```

Enable basic QinQ on GigabitEthernet 1/0/2.

```
[SwitchB-GigabitEthernet1/0/2] qinq enable
```

Apply the uplink policy **p2** to the inbound direction of GigabitEthernet 1/0/2.

```
[SwitchB-GigabitEthernet1/0/2] qos apply policy p2 inbound
```

Apply the downlink policy **p22** to the outbound direction of GigabitEthernet 1/0/2.

```
[SwitchB-GigabitEthernet1/0/2] qos apply policy p22 outbound
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 to permit frames of the specified SVLANs to pass through.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-type trunk
[SwitchB-GigabitEthernet1/0/3] port trunk permit vlan 111 211 311 112 212 312
```

3) Configuration on Switch C

Enable DHCP snooping.

```
<SwitchC> system-view
[SwitchC] dhcp-snooping
```

Enable ARP detection on each VLAN involved in VLAN mapping.

```
[SwitchC] vlan 101
[SwitchC-vlan101] arp detection enable
[SwitchC-vlan101] vlan 201
[SwitchC-vlan201] arp detection enable
[SwitchC-vlan201] vlan 301
[SwitchC-vlan301] arp detection enable
[SwitchC-vlan301] vlan 102
[SwitchC-vlan102] arp detection enable
[SwitchC-vlan102] vlan 202
[SwitchC-vlan202] arp detection enable
[SwitchC-vlan202] vlan 302
[SwitchC-vlan302] arp detection enable
[SwitchC-vlan302] vlan 111
[SwitchC-vlan111] arp detection enable
[SwitchC-vlan111] vlan 211
[SwitchC-vlan211] arp detection enable
[SwitchC-vlan211] vlan 311
[SwitchC-vlan311] arp detection enable
[SwitchC-vlan311] vlan 112
[SwitchC-vlan112] arp detection enable
[SwitchC-vlan112] vlan 212
[SwitchC-vlan212] arp detection enable
[SwitchC-vlan212] vlan 312
[SwitchC-vlan312] arp detection enable
[SwitchC-vlan312] vlan 501
[SwitchC-vlan501] arp detection enable
[SwitchC-vlan501] vlan 502
[SwitchC-vlan502] arp detection enable
[SwitchC-vlan502] vlan 503
[SwitchC-vlan503] arp detection enable
[SwitchC-vlan503] quit
```

Configure uplink policies to map the CVLANs for the same service of different users to the same SVLAN.

```
[SwitchC] traffic classifier c1
[SwitchC-classifier-c1] if-match customer-vlan-id 101 to 200
[SwitchC-classifier-c1] traffic classifier c2
[SwitchC-classifier-c2] if-match customer-vlan-id 201 to 300
[SwitchC-classifier-c2] traffic classifier c3
[SwitchC-classifier-c3] if-match customer-vlan-id 301 to 400
[SwitchC-classifier-c3] traffic classifier c4
[SwitchC-classifier-c4] if-match customer-vlan-id 111 to 210
```

```
[SwitchC-classifier-c4] traffic classifier c5
[SwitchC-classifier-c5] if-match customer-vlan-id 211 to 310
[SwitchC-classifier-c5] traffic classifier c6
[SwitchC-classifier-c6] if-match customer-vlan-id 311 to 410
[SwitchC-classifier-c6] quit
[SwitchC] traffic behavior b1
[SwitchC-behavior-b1] remark service-vlan-id 501
[SwitchC-behavior-b1] traffic behavior b2
[SwitchC-behavior-b2] remark service-vlan-id 502
[SwitchC-behavior-b2] traffic behavior b3
[SwitchC-behavior-b3] remark service-vlan-id 503
[SwitchC-behavior-b3] quit
[SwitchC] qos policy p1
[SwitchC-policy-p1] classifier c1 behavior b1 mode dot1q-tag-manipulation
[SwitchC-policy-p1] classifier c2 behavior b2 mode dot1q-tag-manipulation
[SwitchC-policy-p1] classifier c3 behavior b3 mode dot1q-tag-manipulation
[SwitchC-policy-p1] quit
[SwitchC] qos policy p2
[SwitchC-policy-p2] classifier c4 behavior b1 mode dot1q-tag-manipulation
[SwitchC-policy-p2] classifier c5 behavior b2 mode dot1q-tag-manipulation
[SwitchC-policy-p2] classifier c6 behavior b3 mode dot1q-tag-manipulation
[SwitchC-policy-p2] quit
```

Configure GigabitEthernet 1/0/1 to permit frames of the specified CVLANs and SVLANs to pass through.

```
[SwitchC] interface gigabitethernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] port link-type trunk
[SwitchC-GigabitEthernet1/0/1] port trunk permit vlan 101 201 301 102 202 302 501 502 503
```

Enable customer-side QinQ on GigabitEthernet 1/0/1.

```
[SwitchC-GigabitEthernet1/0/1] qinq enable downlink
```

Apply the uplink policy **p1** to the inbound direction of GigabitEthernet 1/0/1.

```
[SwitchC-GigabitEthernet1/0/1] qos apply policy p1 inbound
[SwitchC-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to permit frames of the specified CVLANs and SVLANs to pass through.

```
[SwitchC] interface gigabitethernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] port link-type trunk
[SwitchC-GigabitEthernet1/0/2] port trunk permit vlan 111 211 311 112 212 312 501 502 503
```

Enable customer-side QinQ on GigabitEthernet 1/0/2.

```
[SwitchC-GigabitEthernet1/0/2] qinq enable downlink
```

Apply the uplink policy **p2** to the inbound direction of GigabitEthernet 1/0/2.

```
[SwitchC-GigabitEthernet1/0/2] qos apply policy p2 inbound
[SwitchC-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 to permit frames of the specified SVLANs to pass through.

```
[SwitchC] interface gigabitethernet 1/0/3
```

```
[SwitchC-GigabitEthernet1/0/3] port link-type trunk
[SwitchC-GigabitEthernet1/0/3] port trunk permit vlan 501 502 503
```

Configure GigabitEthernet 1/0/3 as a DHCP snooping trusted port.

```
[SwitchC-GigabitEthernet1/0/3] dhcp-snooping trust
```

Configure GigabitEthernet 1/0/3 as an ARP trusted port.

```
[SwitchC-GigabitEthernet1/0/3] arp detection trust
```

Enable SP-side QinQ on GigabitEthernet 1/0/3.

```
[SwitchC-GigabitEthernet1/0/3] qinq enable uplink
```

4) Configuration on Switch D

Enable DHCP snooping.

```
<SwitchD> system-view
[SwitchD] dhcp-snooping
```

Configure GigabitEthernet 1/0/1 to permit frames of the specified SVLANs to pass through.

```
[SwitchD] interface gigabitethernet 1/0/1
[SwitchD-GigabitEthernet1/0/1] port link-type trunk
[SwitchD-GigabitEthernet1/0/1] port trunk permit vlan 501 502 503
```

Configure GigabitEthernet 1/0/1 as a DHCP snooping trusted port and disable DHCP snooping to record the IP-to-MAC bindings for DHCP clients on it.

```
[SwitchD-GigabitEthernet1/0/1] dhcp-snooping trust no-user-binding
```

Two-to-Two VLAN Mapping Configuration Example

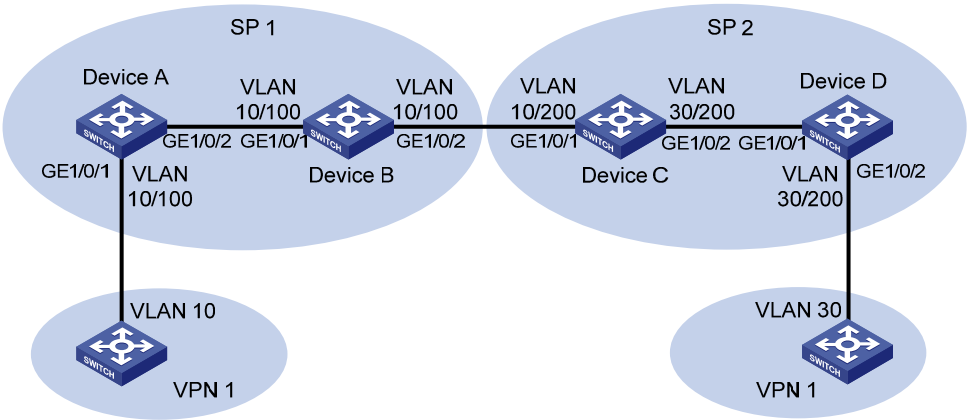
Network requirements

As shown in Figure 1-4:

- Two users of the same VPN are in VLAN 10 and VLAN 30 respectively.
- SP 1 assigns VLAN 100 to VPN 1 users, and SP 2 assigns VLAN 200 to VPN 1 users.

Configure two-to-two VLAN mappings to allow the users that belong to the same VPN but are located in different regions to communicate with each other across the networks of SP 1 and SP 2.

Figure 1-4 Network diagram for two-to-two VLAN mapping configuration



Configuration procedure

1) Configuration on Device A

Configure QinQ function on GigabitEthernet 1/0/1 to add outer VLAN tag 100 to the traffic tagged with VLAN 10.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port access vlan 100
[DeviceA-GigabitEthernet1/0/1] qinq enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Configure the uplink port GigabitEthernet 1/0/2 to permit frames of VLAN 100 to pass through.

```
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 100
```

2) Configuration on Device B

Configure GigabitEthernet 1/0/1 to permit frames of VLAN 100 to pass through.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 100
[DeviceB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to permit frames of VLAN 100 to pass through.

```
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 100
```

3) Configuration on Device C

Specify the original CVLAN and SVLAN in the VLAN mapping for VPN 1 traffic received on GigabitEthernet 1/0/1.

```
<DeviceC> system-view
[DeviceC] traffic classifier downlink_in
[DeviceC-classifier-downlink_in] if-match customer-vlan-id 10
[DeviceC-classifier-downlink_in] if-match service-vlan-id 100
[DeviceC-classifier-downlink_in] quit
```

Specify the new SVLAN used for replacing the original SVLAN for incoming VPN 1 traffic on GigabitEthernet 1/0/1.

```
[DeviceC] traffic behavior downlink_in
[DeviceC-behavior-downlink_in] remark service-vlan-id 200
[DeviceC-behavior-downlink_in] quit
```

Configure an uplink policy to map the original SVLAN and CVLAN to the new SVLAN.

```
[DeviceC] qos policy downlink_in
[DeviceC-qospolicy-downlink_in] classifier downlink_in behavior downlink_in
[DeviceC-qospolicy-downlink_in] quit
```

Specify the new CVLAN and SVLAN in the VLAN mapping for outgoing VPN 1 traffic on GigabitEthernet 1/0/1.

```
[DeviceC] traffic classifier downlink_out
[DeviceC-classifier-downlink_out] if-match customer-vlan-id 30
[DeviceC-classifier-downlink_out] if-match service-vlan-id 200
[DeviceC-classifier-downlink_out] quit
```

Specify the original CVLAN and SVLAN for outgoing VPN 1 traffic on GigabitEthernet 1/0/1.

```
[DeviceC] traffic behavior downlink_out
[DeviceC-behavior-downlink_out] remark customer-vlan-id 10
[DeviceC-behavior-downlink_out] remark service-vlan-id 100
[DeviceC-behavior-downlink_out] quit
```

Configure a downlink policy to map the new CVLAN and SVLAN to the original CVLAN and SVLAN for the outgoing VPN 1 traffic on GigabitEthernet 1/0/1.

```
[DeviceC] qos policy downlink_out
[DeviceC-qospolicy-downlink_out] classifier downlink_out behavior downlink_out
[DeviceC-qospolicy-downlink_out] quit
```

Specify the original CVLAN and the new SVLAN in the VLAN mapping for outgoing VPN 1 traffic on GigabitEthernet 1/0/2.

```
[DeviceC] traffic classifier uplink_out
[DeviceC-classifier-uplink_out] if-match customer-vlan-id 10
[DeviceC-classifier-uplink_out] if-match service-vlan-id 200
[DeviceC-classifier-uplink_out] quit
```

Specify the new CVLAN used for replacing the original CVLAN for outgoing VPN 1 traffic on GigabitEthernet 1/0/2.

```
[DeviceC] traffic behavior uplink_out
[DeviceC-behavior-uplink_out] remark customer-vlan-id 30
[DeviceC-behavior-uplink_out] quit
```

Configure an uplink policy to map the original CVLAN and the new SVLAN to the new CVLAN for outgoing VPN 1 traffic on GigabitEthernet 1/0/2.

```
[DeviceC] qos policy uplink_out
[DeviceC-qospolicy-uplink_out] classifier uplink_out behavior uplink_out
[DeviceC-qospolicy-uplink_out] quit
```

Apply uplink and downlink policies to GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 200
[DeviceC-GigabitEthernet1/0/1] qos apply policy downlink_in inbound
[DeviceC-GigabitEthernet1/0/1] qos apply policy downlink_out outbound
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 200
[DeviceC-GigabitEthernet1/0/2] qos apply policy uplink_out outbound
[DeviceC-GigabitEthernet1/0/2] quit
```

4) Configuration on Device D

Configure QinQ function on GigabitEthernet 1/0/2 to add outer VLAN tag 200 to the traffic tagged with VLAN 30.

```
<DeviceD> system-view
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] port access vlan 200
[DeviceD-GigabitEthernet1/0/2] qinq enable
```

Configure GigabitEthernet 1/0/1 to permit frames of VLAN 200 to pass through.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 200
```

Table of Contents

1 Ethernet OAM Configuration

When configuring the Ethernet OAM function, go to these sections for information you are interested in:

Ethernet OAM Overview

Ethernet OAM (operation, administration, and maintenance) is a tool monitoring Layer-2 link status by sending OAM protocol data units (OAMPDUs) between devices. It helps network administrators manage their networks effectively.

Currently, Ethernet OAM is mainly used to address common link-related issues on the “last mile.” By enabling Ethernet OAM on two devices connected by a point-to-point connection, you can monitor the status of the link. Ethernet OAM provides the following functions:

- Link performance monitoring, for detecting link errors
- Fault detection and alarm, for reporting link errors to the administrators
- Loopback testing, for detecting link errors through non-OAMPDUs

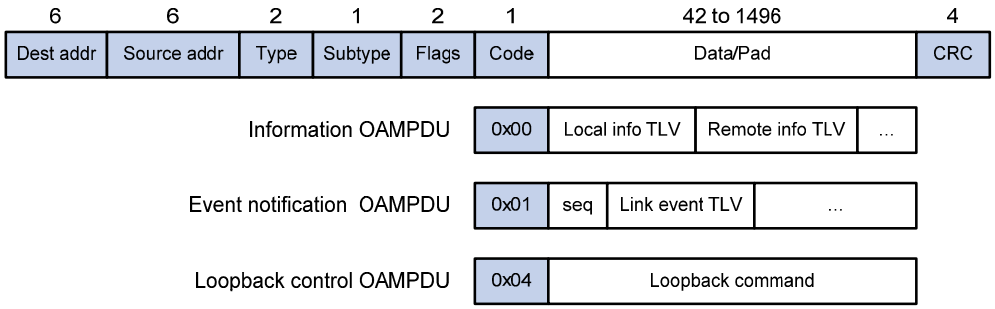
Note

Throughout this document, a port with Ethernet OAM enabled is called an Ethernet OAM entity or an OAM entity.

Types of Ethernet OAMPDUs

Figure 1-1 shows the formats of different types of OAMPDUs.

Figure 1-1 Formats of different types of Ethernet OAMPDUs



The fields in an OAMPDU are described as follows:

Table 1-1 Description of the fields in an OAMPDU

Field	Description
Dest addr	Destination MAC address of the Ethernet OAMPDU. It is a slow protocol multicast address 0180c2000002.
Source addr	Source MAC address of the Ethernet OAMPDU. It is the bridge MAC address of the sending side and is a unicast MAC address.
Type	Type of the encapsulated protocol in the Ethernet OAMPDU. The value is 0x8809.
Subtype	The specific protocol being encapsulated in the Ethernet OAMPDU. The value is 0x03.
Flags	Status information of an Ethernet OAM entity.
Code	Type of the Ethernet OAMPDU

Table 1-2 shows the function of the three types of OAMPDUs.

Table 1-2 Functions of different types of OAMPDUs

OAMPDU type	Function
Information OAMPDU	Used for transmitting state information of an Ethernet OAM entity (including the information about the local device and remote devices, and customized information) to the remote Ethernet OAM entity and maintaining OAM connections
Event Notification OAMPDU	Used by link monitoring to notify the remote OAM entity when it detects problems on the link in between.
Loopback Control OAMPDU	Used for remote loopback control. By inserting the information used to enable/disable loopback to a loopback control OAMPDU, you can enable/disable loopback on a remote OAM entity.

Ethernet OAM Implementation

This section describes the working procedures of Ethernet OAM.

Ethernet OAM connection establishment

Ethernet OAM connection is the base of all the other Ethernet OAM functions. OAM connection establishment is also known as the Discovery phase, where an Ethernet OAM entity discovers remote OAM entities and establishes sessions with them.

In this phase, interconnected OAM entities notify the peer of their OAM configuration information and the OAM capabilities of the local nodes by exchanging Information OAMPDUs and determine whether Ethernet OAM connections can be established. An Ethernet OAM connection can be established only when the settings concerning Loopback, link detecting, and link event of the both sides match. After an Ethernet OAM connection is established, Ethernet OAM takes effect on it.

As for Ethernet OAM connection establishment, a device can operate in active Ethernet OAM mode or passive Ethernet OAM mode. Table 1-3 compares active Ethernet OAM mode with passive Ethernet OAM mode.

Table 1-3 Active Ethernet OAM mode and passive Ethernet OAM mode

Item	Active Ethernet OAM mode	Passive Ethernet OAM mode
Initiating OAM Discovery	Available	Unavailable
Responding to OAM Discovery	Available	Available
Transmitting Information OAMPDUs	Available	Available
Transmitting Event Notification OAMPDUs	Available	Available
Transmitting Information OAMPDUs with the Data/Pad field being empty	Available	Available
Transmitting Loopback Control OAMPDUs	Available	Unavailable
Responding to Loopback Control OAMPDUs	Available (if both sides operate in active OAM mode)	Available

Note

- OAM connections can be initiated only by OAM entities operating in active OAM mode, while those operating in passive mode wait and respond to the connection requests sent by their peers.
- No OAM connection can be established between OAM entities operating in passive OAM mode.

After an Ethernet OAM connection is established, the Ethernet OAM entities on both sides exchange Information OAMPDUs periodically to keep the Ethernet OAM connection valid. If an Ethernet OAM entity receives no Information OAMPDU for five seconds, the Ethernet OAM connection is disconnected.

Note

The interval to send Information OAMPDUs is determined by a timer. Up to ten Information OAMPDUs can be sent in a second.

Link monitoring

Error detection in an Ethernet is difficult, especially when the physical connection in the network is not disconnected but network performance is degrading gradually. Link monitoring is used to detect and indicate link faults in various environments. Ethernet OAM implements link monitoring through the exchange of Event Notification OAMPDUs. Upon detecting a link error event listed in Table 1-4, the local OAM entity sends an Event Notification OAMPDU to notify the remote OAM entity. With the log information, network administrators can keep track of network status in time. Table 1-4 describes the link events.

Table 1-4 Ethernet OAM link error events

Ethernet OAM link events	Description
Errored symbol event	An errored symbol event occurs when the number of detected symbol errors over a specific detection interval exceeds the predefined threshold.
Errored frame event	An errored frame event occurs when the number of detected error frames over a specific interval exceeds the predefined threshold.
Errored frame period event	An errored frame period event occurs if the number of frame errors in specific number of received frames exceeds the predefined threshold.
Errored frame seconds event	When the number of error frame seconds detected on a port over a detection interval reaches the error threshold, an errored frame seconds event occurs.

Note

- The system transforms the period of detecting errored frame period events into the maximum number of 64-byte frames that a port can send in the specific period, that is, the system takes the maximum number of frames sent as the period. The maximum number of frames sent is calculated using this formula: the maximum number of frames = interface bandwidth (bps) × errored frame period event detection period (in ms)/(64 × 8 × 1000)
- If errored frames appear in a certain second, this second is called an errored frame second.

Remote fault detection

In a network where traffic is interrupted due to device failures or unavailability, the flag field defined in Ethernet OAMPDUs allows an Ethernet OAM entity to send error information to its peer. It can identify the critical link error events listed in Table 1-5.

Table 1-5 Critical link error events

Ethernet OAM link events	Description
Link Fault	Peer link signal is lost.
Dying Gasp	An unexpected fault, such as power failure, occurred.
Critical event	An undetermined critical event happened.

As Information OAMPDUs are exchanged periodically across established OAM connections, an Ethernet OAM entity can inform one of its OAM peers of link faults through Information OAMPDUs. Therefore, the network administrator can keep track of link status in time through the log information and troubleshoot in time.

Remote loopback testing

Remote loopback testing is available only after the Ethernet OAM connection is established. With remote loopback enabled, the Ethernet OAM entity operating in active Ethernet OAM mode issues remote loopback requests and the peer responds to them. If the peer operates in the loopback mode, it returns all the PDUs except Ethernet OAMPDUs to the senders along the original paths.

Performing remote loopback testing periodically helps to detect network faults in time. Furthermore, performing remote loopback testing by network segments helps to locate network faults.

Standards and Protocols

Ethernet OAM is defined in IEEE 802.3h.

Ethernet OAM Configuration Task List

Complete the following tasks to configure Ethernet OAM:

Task		Remarks
Configuring Basic Ethernet OAM Functions		Required
Configuring Link Monitoring	Configuring Errored Symbol Event Detection	Optional
	Configuring Errored Frame Event Detection	Optional
	Configuring Errored Frame Period Event Detection	Optional
	Configuring Errored Frame Seconds Event Detection	Optional
Enabling OAM Loopback Testing		Optional

Configuring Basic Ethernet OAM Functions

As for Ethernet OAM connection establishment, a device can operate in active mode or passive mode. After Ethernet OAM is enabled on an Ethernet port, according to its Ethernet OAM mode, the Ethernet port establishes an Ethernet OAM connection with its peer port.

Follow these steps to configure basic Ethernet OAM functions:

To do…	Use the command…	Remarks
Enter system view	**System-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Set Ethernet OAM operating mode	**oam mode** { **active** | **passive** }	Optional The default is active Ethernet OAM mode.
Enable Ethernet OAM on the current port	**oam enable**	Required Ethernet OAM is disabled by default.

Configuring Link Monitoring

Note

After Ethernet OAM connections are established, the link monitoring periods and thresholds configured in this section take effect on all Ethernet ports automatically.

Configuring Errored Symbol Event Detection

An errored symbol event occurs when the number of detected symbol errors over a specific detection interval exceeds the predefined threshold.

Follow these steps to configure errored symbol event detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the errored symbol event detection interval	**oam errored-symbol period** *period-value*	Optional 1 second by default
Configure the errored symbol event triggering threshold	**oam errored-symbol threshold** *threshold-value*	Optional 1 by default

Configuring Errored Frame Event Detection

An errored frame event occurs when the number of detected error frames over a specific interval exceeds the predefined threshold.

Follow these steps to configure errored frame event detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the errored frame event detection interval	**oam errored-frame period** *period-value*	Optional 1 second by default
Configure the errored frame event triggering threshold	**oam errored-frame threshold** *threshold-value*	Optional 1 by default

Configuring Errored Frame Period Event Detection

An errored frame period event occurs if the number of frame errors in specific number of received frames exceeds the predefined threshold.

Follow these steps to configure errored frame period event detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the errored frame period event detection period	**oam errored-frame-period period** *period-value*	Optional 1000 milliseconds by default
Configure the errored frame period event triggering threshold	**oam errored-frame-period threshold** *threshold-value*	Optional 1 by default

Configuring Errored Frame Seconds Event Detection

An errored frame seconds event occurs when the number of error frame seconds detected on a port over a detection interval exceeds the error threshold.

Follow these steps to configure errored frame seconds event detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the errored frame seconds event detection interval	**oam errored-frame-seconds period** *period-value*	Optional 60 second by default
Configure the errored frame seconds event triggering threshold	**oam errored-frame-seconds threshold** *threshold-value*	Optional 1 by default

Caution

Make sure the errored frame seconds triggering threshold is less than the errored frame seconds detection interval. Otherwise, no errored frame seconds event can be generated.

Enabling OAM Loopback Testing

Follow these steps to enable Ethernet OAM loopback testing:

To do…	Use the command…	Remarks
Enter system view	**System-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Enable Ethernet OAM loopback testing	**oam loopback**	Required Disabled by default.

Note

- Ethernet OAM loopback testing is available only after the Ethernet OAM connection is established and can be performed only by the Ethernet OAM entities operating in active Ethernet OAM mode.
- Loopback testing is available only on full-duplex links that support remote loopback at both ends.
- Ethernet OAM loopback testing needs the support of the peer hardware.
- Enabling Ethernet OAM loopback testing interrupts data communications. After Ethernet OAM loopback testing is disabled, all the ports involved will shut down and then come up. Ethernet OAM loopback testing is disabled when you execute the **undo oam enable** command to disable Ethernet OAM, when you execute the **undo oam loopback** command to disable Ethernet OAM loopback testing, or when the Ethernet OAM connection times out.
- Ethernet OAM loopback testing is only applicable to individual links. It is not applicable to link aggregation member ports or service loopback group member ports. In addition, you cannot assign ports where Ethernet OAM loopback testing is being performed to link aggregation groups or service loopback groups. For more information about link aggregation groups and service loopback groups, refer to *Link Aggregation Configuration* and *Service Loopback Group Configuration* in the *Access Volume*.
- Enabling internal loopback test on a port in remote loopback test can terminate the remote loopback test. For more information about loopback test, refer to *Ethernet Interface Configuration* in the *Access Volume*.

Displaying and Maintaining Ethernet OAM Configuration

To do…	Use the command…	Remarks
Display global Ethernet OAM configuration	**display oam configuration**	Available in any view
Display the statistics on critical events after an Ethernet OAM connection is established	**display oam critical-event** [**interface** *interface-type interface-number*]	
Display the statistics on Ethernet OAM link error events after an Ethernet OAM connection is established or after you clear the statistics	**display oam link-event** { **local** | **remote** } [**interface** *interface-type interface-number*]	
Display the information about an Ethernet OAM connection	**display oam** { **local** | **remote** } [**interface** *interface-type interface-number*]	

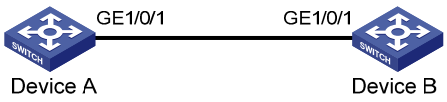
To do…	Use the command…	Remarks
Clear statistics on Ethernet OAM packets and Ethernet OAM link error events	**reset oam** [**interface** *interface-type interface-number*]	Available in user view only

Ethernet OAM Configuration Example

Network requirements

- Enable Ethernet OAM on Device A and Device B to manage links on data link layer.
- Monitor link performance and collect statistics about the error frames received by Device A.

Figure 1-2 Network diagram for Ethernet OAM configuration



Configuration procedure

1) Configure Device A

Configure GigabitEthernet 1/0/1 to operate in passive Ethernet OAM mode and enable Ethernet OAM for it.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] oam mode passivez
[DeviceA-GigabitEthernet1/0/1] oam enable
[DeviceA-GigabitEthernet1/0/1] quit
```

Set the errored frame detection interval to 20 seconds and set the errored frame event triggering threshold to 10.

```
[DeviceA] oam errored-frame period 20
[DeviceA] oam errored-frame threshold 10
```

2) Configure Device B

Configure GigabitEthernet 1/0/1 to operate in active Ethernet OAM mode (the default) and enable Ethernet OAM for it.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] oam mode active
[DeviceB-GigabitEthernet1/0/1] oam enable
[DeviceB-GigabitEthernet1/0/1] quit
```

3) Verify the configuration

Use the **display oam configuration** command to display the Ethernet OAM configuration. For example:

Display the Ethernet OAM configuration on Device A.

```
[DeviceA] display oam configuration
Configuration of the link event window/threshold :
```

```
------------------------------------------------------------------------------
Errored-symbol Event period(in seconds)          :     1
Errored-symbol Event threshold                   :     1
Errored-frame Event period(in seconds)           :     20
Errored-frame Event threshold                    :     10
Errored-frame-period Event period(in ms)         :     1000
Errored-frame-period Event threshold             :     1
Errored-frame-seconds Event period(in seconds)   :     60
Errored-frame-seconds Event threshold            :     1
```

Use the **display oam link-event** command to display the statistics about Ethernet OAM link events. For example:

Display Ethernet OAM link event statistics of the remote end of Device B.

```
[DeviceB] display oam link-event remote
Port :GigabitEthernet1/0/1
Link Status :Up
OAMRemoteErrFrameEvent : (ms = milliseconds)
------------------------------------------------------------------------
Event Time Stamp          : 5789        Errored FrameWindow   : 10(100ms)
Errored Frame Threshold   : 1           Errored Frame         : 3
Error Running Total       : 35          Event Running Total   : 17
```

The above information indicates that 35 errors occurred since Ethernet OAM is enabled on Device A, 17 of which are caused by error frames. The link is instable.

Table of Contents

1 Connectivity Fault Detection Configuration

When configuring CFD, go to these sections for information you are interested in:

Overview

Connectivity Fault Detection (CFD) is an end-to-end per-VLAN link layer Operations, Administration and Maintenance (OAM) mechanism used for link connectivity detection, fault verification, and fault location.

Basic Concepts in CFD

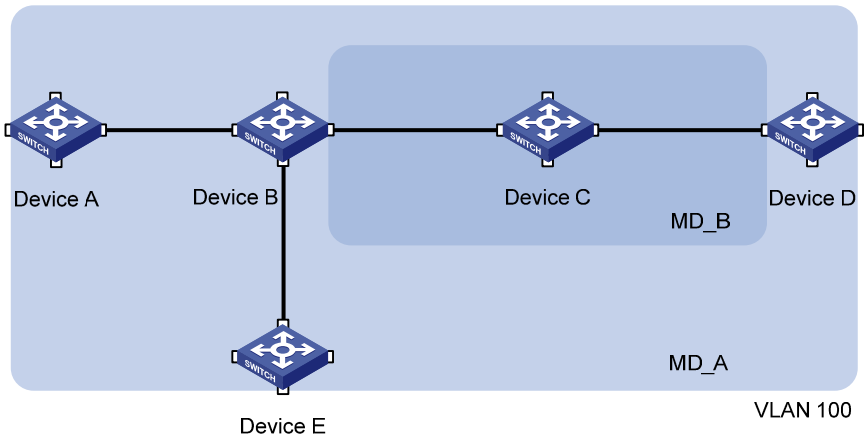
Maintenance domain

A maintenance domain (MD) defines the network where CFD plays its role. The MD boundary is defined by some maintenance association end points MEPs configured on the ports. A MD is identified by an MD name.

To locate faults exactly, CFD introduces eight levels (from 0 to 7) to MDs. The bigger the number, the higher the level and the larger the area covered. Domains can touch or nest (if the outer domain has a higher level than the nested one) but cannot intersect or overlap.

MD levels facilitate fault location and make fault location more accurate. As shown in Figure 1-1, MD_A in light blue nests MD_B in dark blue. If a connectivity fault is detected at the boundary of MD_A, any of the devices in MD_A, including Device A through Device E, may fail. In this case, if a connectivity fault is also detected at the boundary of MD_B, the failure points may be any of Device B through Device D. If the devices in MD_B operate normally, you can be sure that at least Device C is operational.

Figure 1-1 Two nested MDs



CFD exchanges messages and performs operations on a per-domain basis. By planning MDs properly in a network, you can use CFD to locate failure points rapidly.

Maintenance association

A maintenance association (MA) is a set of maintenance points (MPs) in a MD. An MA is identified by the “MD name + MA name”.

An MA serves a VLAN. Packets sent by the MPs in an MA carry the corresponding VLAN tag. An MP can receive packets sent by other MPs in the same MA.

Maintenance point

An MP is configured on a port and belongs to an MA. MPs fall into two types: maintenance association end points (MEPs) and maintenance association intermediate points (MIPs).

- MEP

Each MEP is identified by an integer called a MEP ID. The MEPs of an MD define the range and boundary of the MD. The MA and MD that a MEP belongs to define the VLAN attribute and level of the packets sent by the MEP. MEPs fall into inward-facing MEPs and outward-facing MEPs.

The level of a MEP determines the levels of packets that the MEP can process. The packets transmitted from a MEP carry the level of the MEP. An MEP forwards packets at a higher level and processes packet of its level or lower. The processing procedure is specific to packets in the same VLAN. Packets of different VLANs are independent.

The direction of a MEP determines the position of the MD relative to the port. In Figure 1-2, outward-facing MEPs are configured on the two ports. In Figure 1-3, inward-facing MEPs are configured on the two ports.

An outward-facing MEP communicates through the wire side connected to the port; an inward-facing MEP communicates through the relay function side.

Figure 1-2 Outward-facing MEP

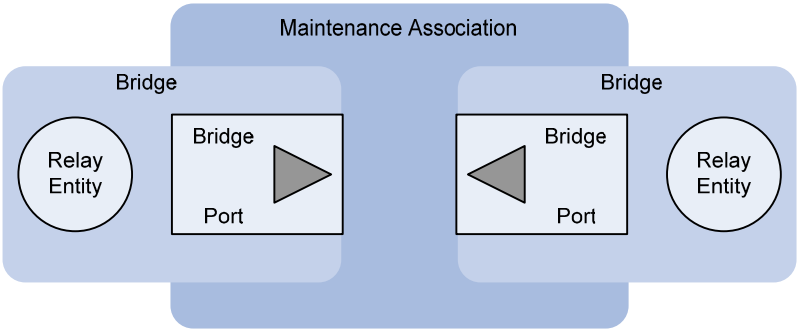
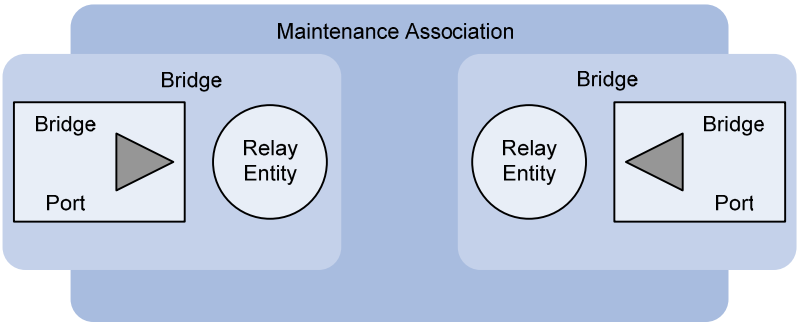


Figure 1-3 Inward-facing MEP



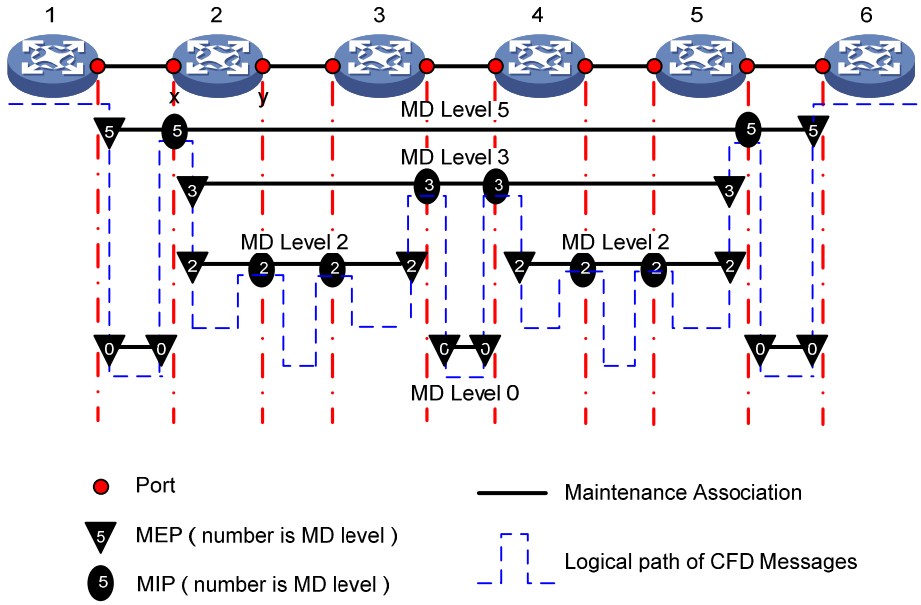
- MIP

A MIP is internal to an MD. It cannot send CFD packets actively; however, it can handle and respond to CFD packets. The MA and MD that a MIP belongs to define the VLAN attribute and level of the packets received.

By cooperating with MEPs, a MIP can perform a function similar to ping and traceroute. Like a MEP, a MIP forwards packets at a higher level without any processing.

Figure 1-4 demonstrates a grading example of the CFD module. In the figure, there are six devices, labeled 1 through 6 respectively. Suppose each device has two ports, and MEPs and MIPs are configured on some of these ports. Four levels of MDs are designed in this example, the bigger the number, the higher the level and the larger the area covered. In this example, the X port of device 2 is configured with the following MPs: a level 5 MIP, a level 3 inward-facing MEP, a level 2 inward-facing MEP, and a level 0 outward-facing MEP.

Figure 1-4 Levels of MPs



Basic Functions of CFD

CFD works effectively only in properly-configured networks. Its functions, which are implemented through the MPs, include:

- Continuity check (CC);
- Loopback (LB)
- Linktrace (LT)

Continuity check

Continuity check is responsible for checking the connectivity between MEPs. Connectivity faults are usually caused by device faults or configuration errors. This function is implemented through periodic sending of continuity check messages (CCMs) by the MEPs. As a multicast message, a CCM sent by one MEP is intended to be received by all the other MEPs in the same MA. If a MEP fails to receive the CCMs within 3.5 sending periods, the link is regarded as faulty and a corresponding log is generated. When multiple MEPs send CCMs at the same time, the multipoint-to-multipoint link check is achieved.

Loopback

Similar to ping at the IP layer, loopback is responsible for verifying the connectivity between a local device and a remote device. To implement this function, the local MEP sends loopback messages (LBMs) to the remote MEP. Depending on whether the local MEP can receive a loopback reply message (LBR) from the remote MEP, the link state between the two can be verified. LBMs and LBRs are unicast messages. They are used to verify the connectivity between two points.

Linktrace

Linktrace is responsible for identifying the path between the source MEP and the destination MEP. This function is implemented in the following way: the source MEP multicasts linktrace messages (LTMs) to the destination MEP. After receiving the messages, the destination MEP and the MIPs that the LTMs pass send back linktrace reply messages (LTRs) to the source MEP. Based on the reply messages, the

source MEP can identify the path to the destination MEP. Note that LTMs are multicast frames while LTRs are unicast frames.

Protocols and Standards

The CFD function is implemented in accordance with IEEE P802.1ag.

CFD Configuration Task List

For CFD to work effectively, you should first design the network by performing the following tasks:

- Grade the MDs in the entire network, and define the boundary of each MD.
- Assign a name for each MD. Make sure that the same MD has the same name on different devices.
- Define the MA in each MD according to the VLAN you want to monitor.
- Assign a name for each MA. Make sure that the same MA in the same MD has the same name on different devices.
- At the edges of MD and MA, MPs should be designed at the device port. MEPs can be designed on devices or ports that are not at the edges.

Complete the following tasks to configure CFD:

Tasks	Remarks
Basic Configuration Tasks	Required These configurations are the foundation for other configuration tasks.
Configuring CC on MEPs	Required Configuring the MEPs to send CCMs to manage link connectivity
Configuring LB on MEPs	Optional Checking link state by testing link connectivity
Configuring LT on MEPs	Optional Tracing link fault and finding the path between the source MEP and target MEP

Note

- A port blocked by STP cannot receive, send, or respond to CFD messages, however, if the port is configured as an outward-facing MEP, it can still receive and send CCM messages even if it is blocked by STP.
- Only Ethernet ports support CFD.

Basic Configuration Tasks

Basic configuration tasks include:

- Configuring Service Instance
- Configuring MEP
- Configuring MIP Generation Rules

Note

Based on the network design, you should configure MEPs or the rules for generating MIPs on each device. However, before doing this you must first configure the service instance.

Configuring Service Instance

A service instance is indicated by an integer to represent an MA in an MD. The MD and MA define the level and VLAN attribute of the messages handled by the MPs in a service instance.

Follow these steps to configure a service instance:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable CFD	**cfd enable**	Required CFD is disabled by default.
Create an MD	**cfd md** *md-name* **level** *level-value*	Required Not created by default
Create an MA	**cfd ma** *ma-name* **md** *md-name* **vlan** *vlan-id*	Required Not created by default
Create a service instance	**cfd service-instance** *instance-id* **md** *md-name* **ma** *ma-name*	Required Not created by default

Caution

- These configuration tasks are the foundation for other CFD configuration tasks.
- The last three steps in the table above must be performed strictly in order.

Configuring MEP

MEPs are functional entities in a service instance. CFD is implemented through operations on MEPs, which provides such functions as CC, LB, LT and gives prompts on error CCMs and cross connections. As a MEP is configured on a service instance, the MD level and VLAN attribute of the service instance become the attribute of the MEP.

Follow these steps to configure a MEP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Configure a MEP	**cfd mep** *mep-id* **service-instance** *instance-id* { **inbound** | **outbound** }	Required Not configured by default

To do...	Use the command...	Remarks
Configure a remote MEP for a MEP in the same service instance	**cfd remote-mep** *remote-mep-id* **service-instance** *instance-id* **mep** *mep-id*	Required No remote MEP is configured for a MEP by default.
Enable the MEP	**cfd mep service-instance** *instance-id* **mep** *mep-id* **enable**	Required Disabled by default

Configuring MIP Generation Rules

As functional entities in a service instance, MIPs deal with LBM and LTM messages.

MIPs are generated on each port according to some rules. You can choose appropriate MIP generation rules based on your network design.

Follow these steps to configure the rules for generating MIPs:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the rules for generating MIPs	**cfd mip-rule** { **explicit** | **default** } **service-instance** *instance-id*	Required By default, neither the MIPs nor the rules for generating MIPs are configured.

MIPs are generated on each port automatically according to the rules specified in the **cfd mip-rule** command. If a port has no MIP, the system will check the MAs in each MD (from low to high levels), and follow the rules in Table 1-1 to create or not create MIPs (within a single VLAN):

Table 1-1 Rules for generating MIP

MIP exists on low level MA	The cfd mip-rule command is configured as	MEP exists on low level MA	Create MIP or not
Yes	—	—	No
No	Explicit	No	No
		Yes	Yes
	Default	—	Yes

Each of the following actions or cases can cause MIPs to be created or deleted after you have configured the **cfd mip-rule** command:

- Enabling CFD (use the **cfd enable** command)
- Creating or deleting the MEPs on a port
- Changes occur to the VLAN attribute of a port
- The rule specified in the **cfd mip-rule** command changes

Configuring CC on MEPs

After the CC function is configured, MEPs can send CCMs mutually to check the connectivity between them.

Configuration Prerequisites

Before configuring this function, you should first complete the MEP configuration.

Configuring Procedure

Follow these steps to configure CC on a MEP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the interval field value in the CCM messages sent by MEPs	**cfd cc interval** *interval-field-value* **service-instance** *instance-id*	Optional By default, the interval filed value is 5.
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Enable CCM sending on a MEP	**cfd cc service-instance** *instance-id* **mep** *mep-id* **enable**	Required Disabled by default

The relationship between the interval field value in the CCM messages, the interval between CCM messages and the timeout time of the remote MEP is illustrated in Table 1-2.

Table 1-2 Relationship of the interval field value, the interval between CCM messages and the timeout time of the remote MEP

The interval field value	The interval between CCM messages	The timeout time of the remote MEP
5	1 second	3.5 seconds
6	10 seconds	35 seconds
7	60 seconds	210 seconds

Caution

On different devices, the MEPs belonging to the same MD and MA should be configured with the same time interval for CCMs sending.

Configuring LB on MEPs

The LB function can verify the link state between two ends after CC detects a link fault.

Configuration Prerequisites

Before configuring this function, you should first complete the MEP and MIP configuration tasks.

Configuration Procedure

Follow these steps to configure LB on MEP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable LB	**cfd loopback service-instance** *instance-id* **mep** *mep-id* { **target-mep** *target-mep-id* | **target-mac** *mac-address* } [**number** *loopback-number*]	Required Disabled by default

Configuring LT on MEPs

LT can trace the path between the specified MEP and the target MEP, and can also locate link faults by sending LT messages automatically. The two functions are implemented in the following way:

- To implement the first function, the specified MEP first sends LTM messages to the target MEP. Based on the LTR messages in response to the LTM messages, the path between the two MEPs can be identified.
- In the latter case, after LT messages automatic sending is enabled, if a MEP fails to receive the CCMs from the remote MEP within 3.5 sending intervals, the link between the two is regarded as faulty and LTMs will be sent out. Based on the LTRs that echo back, the fault source can be located.

Configuration Prerequisites

Before configuring this function, you should first complete MEP and MIP configuration tasks.

Finding the Path Between a Source MEP and a Target MEP

Follow these steps to find the path between a source MEP and a target MEP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Find the path between a source MEP and a target MEP	**cfd linktrace service-instance** *instance-id* **mep** *mep-id* { **target-mep** *target-mep-id* | **target-mac** *mac-address* } [**ttl** *ttl-value*] [**hw-only**]	Required

Enabling Automatic LT Messages Sending

Follow these steps to enable automatic LT messages sending:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable automatic LT messages sending	**cfd linktrace auto-detection** [**size** *size-value*]	Required Disabled by default

Displaying and Maintaining CFD

To do...	Use the command...	Remarks
Display CFD status	**display cfd status**	Available in any view
Display MD configuration information	**display cfd md**	Available in any view
Display MA configuration information	**display cfd ma** [[*ma-name*] **md** *md-name*]	Available in any view
Display service instance configuration information	**display cfd service-instance** [*instance-id*]	Available in any view
Display MP information	**display cfd mp** [**interface** *interface-type interface-number*]	Available in any view
Display the attribute and running information of the MEPs	**display cfd mep** *mep-id* **service-instance** *instance-id*	Available in any view
Display LTR information received by a MEP	**display cfd linktrace-reply** [**service-instance** *instance-id* [**mep** *mep-id*]]	Available in any view
Display the information of a remote MEP	**display cfd remote-mep service-instance** *instance-id* **mep** *mep-id*	Available in any view
Display the content of the LTR that responds to LTM messages	**display cfd linktrace-reply auto-detection** [**size** *size-value*]	Available in any view

CFD Configuration Examples

Configuring Service Instance

Network requirements

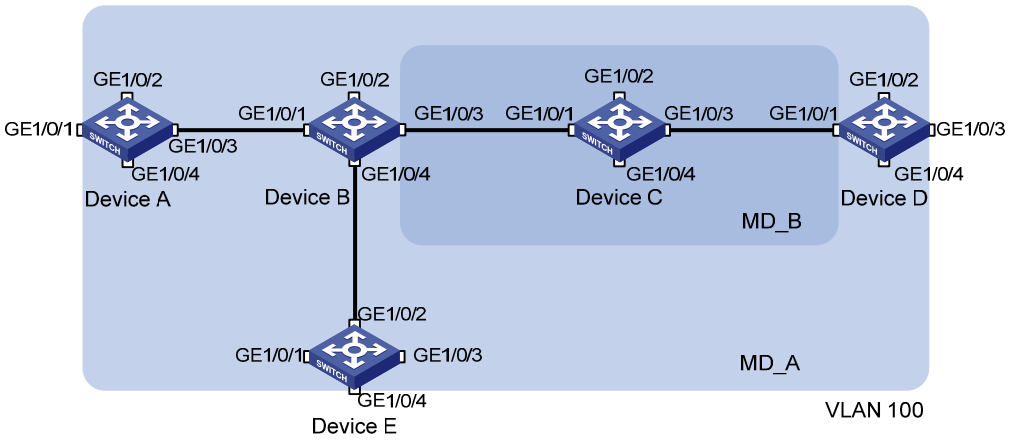
As shown in Figure 1-5, there are five devices in the MDs. Each device has four ports belonging to VLAN 100. The light blue square frame and the blue one specify two different MDs.

- Two MDs, MD_A (indicated by the light blue square frame, with level 5) and MD_B (indicated by the blue square frame, with level 3)) are designed in this network.
- Define the edge ports of each MD, and define the MD of each port.
- The VLAN IDs of each MA in the two MDs are all 100.

According to the network diagram as shown in Figure 1-5, You should perform the following configurations:

- Configure MD_A on Device A and Device E
- Configure MD_B on Device C
- Configure MD_A and MD_B on Device B and Device D
- Configure an MA in each MD
- Configure a service instance for each MA

Figure 1-5 Network diagram for MD configuration



Configuration procedure

1) Configuration on Device A (configuration on Device E is the same as that on Device A)

```
<DeviceA> system-view
[DeviceA] cfd enable
[DeviceA] cfd md MD_A level 5
[DeviceA] cfd ma MA_MD_A md MD_A vlan 100
[DeviceA] cfd service-instance 1 md MD_A ma MA_MD_A
```

2) Configuration on Device C

```
<DeviceC> system-view
[DeviceC] cfd enable
[DeviceC] cfd md MD_B level 3
[DeviceC] cfd ma MA_MD_B md MD_B vlan 100
[DeviceC] cfd service-instance 2 md MD_B ma MA_MD_B
```

3) Configuration on Device B (configuration on Device D is the same as that on Device B)

```
<DeviceB> system-view
[DeviceB] cfd enable
[DeviceB] cfd md MD_A level 5
[DeviceB] cfd ma MA_MD_A md MD_A vlan 100
[DeviceB] cfd service-instance 1 md MD_A ma MA_MD_A
[DeviceB] cfd md MD_B level 3
[DeviceB] cfd ma MA_MD_B md MD_B vlan 100
[DeviceB] cfd service-instance 2 md MD_B ma MA_MD_B
```

After the above configuration, you can use the commands **display cfd md**, **display cfd ma** and **display cfd service-instance** to verify your configuration.

Configuring MEP and Enabling CC on it

Network requirements

After finishing service instance configuration, you can start to design the MEPs.

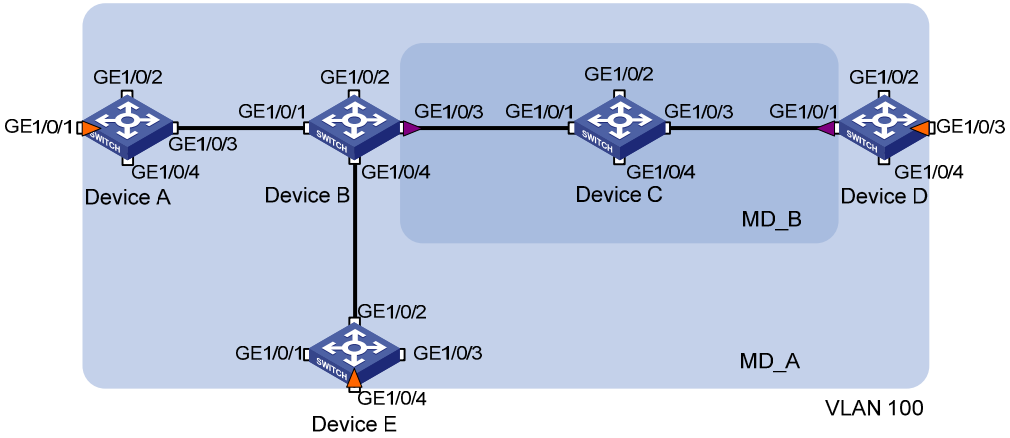
- MEPs are configured at the edge or border of MDs. Find the edge port of each MD.
- Decide the MEP direction (inward-facing or outward-facing) on each edge port based on the MD position.
- Assign a unique ID to each MEP in an MA.

- Decide the remote MEP for each MEP, and enable these MEPs.

According to the network diagram as shown in Figure 1-6, perform the following configurations:

- In MD_A, there are three edge ports: GigabitEthernet 1/0/1 on Device A, GigabitEthernet 1/0/3 on Device D and GigabitEthernet 1/0/4 on Device E. Configure inward-facing MEPs on these ports respectively.
- In MD_B, there are two edge ports: GigabitEthernet 1/0/3 on Device B and GigabitEthernet 1/0/1 on Device D. Configure outward-facing MEPs on the two ports respectively.
- In MD_A and MD_B, each MEP checks the messages from other MEPs.

Figure 1-6 Network diagram of MD and MEP configuration



Configuration procedure

1) On Device A

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] cfd mep 1001 service-instance 1 inbound
[DeviceA-GigabitEthernet1/0/1] cfd remote-mep 5001 service-instance 1 mep 1001
[DeviceA-GigabitEthernet1/0/1] cfd remote-mep 4002 service-instance 1 mep 1001
[DeviceA-GigabitEthernet1/0/1] cfd mep service-instance 1 mep 1001 enable
[DeviceA-GigabitEthernet1/0/1] cfd cc service-instance 1 mep 1001 enable
```

2) On Device B

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] cfd mep 2001 service-instance 2 outbound
[DeviceB-GigabitEthernet1/0/3] cfd remote-mep 4001 service-instance 2 mep 2001
[DeviceB-GigabitEthernet1/0/3] cfd mep service-instance 2 mep 2001 enable
[DeviceB-GigabitEthernet1/0/3] cfd cc service-instance 2 mep 2001 enable
```

3) On Device D

```
<DeviceD> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] cfd mep 4001 service-instance 2 outbound
[DeviceD-GigabitEthernet1/0/1] cfd remote-mep 2001 service-instance 2 mep 4001
[DeviceD-GigabitEthernet1/0/1] cfd mep service-instance 2 mep 4001 enable
[DeviceD-GigabitEthernet1/0/1] cfd cc service-instance 2 mep 4001 enable
[DeviceD-GigabitEthernet1/0/1] interface gigabitethernet 1/0/3
[DeviceD-GigabitEthernet1/0/3] cfd mep 4002 service-instance 1 inbound
```

```
[DeviceD-GigabitEthernet1/0/3] cfd remote-mep 1001 service-instance 1 mep 4002
[DeviceD-GigabitEthernet1/0/3] cfd remote-mep 5001 service-instance 1 mep 4002
[DeviceD-GigabitEthernet1/0/3] cfd mep service-instance 1 mep 4002 enable
[DeviceD-GigabitEthernet1/0/3] cfd cc service-instance 1 mep 4002 enable
```

4) On Device E

```
<DeviceE> system-view
[DeviceE] interface gigabitethernet 1/0/4
[DeviceE-GigabitEthernet1/0/4] cfd mep 5001 service-instance 1 inbound
[DeviceE-GigabitEthernet1/0/4] cfd remote-mep 1001 service-instance 1 mep 5001
[DeviceE-GigabitEthernet1/0/4] cfd remote-mep 4002 service-instance 1 mep 5001
[DeviceE-GigabitEthernet1/0/4] cfd mep service-instance 1 mep 5001 enable
[DeviceE-GigabitEthernet1/0/4] cfd cc service-instance 1 mep 5001 enable
```

After the above configuration, you can use the commands **display cfd mp** and **display cfd mep** to verify your configuration.

Configuring the Rules for Generating MIPs

Network requirements

After finishing MEP configuration, you can continue to configure the MIPs.

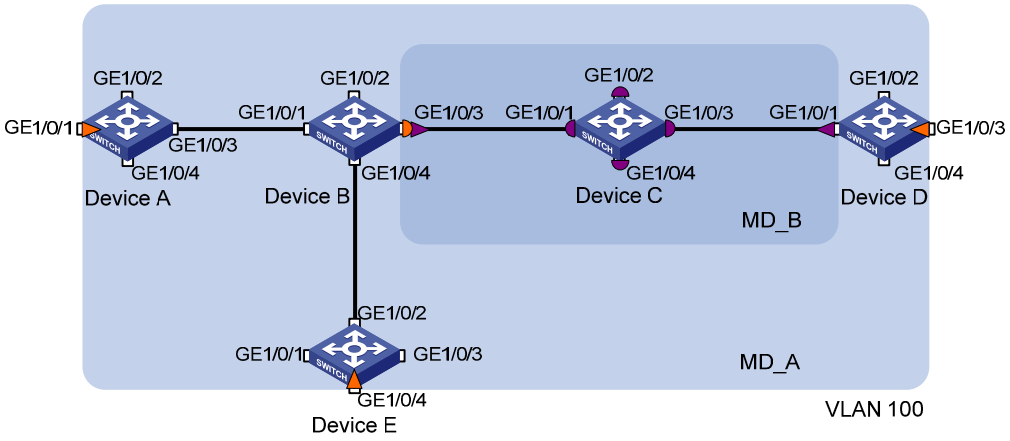
MIPs, which are generated by some rules, are configured in the following way:

- Decide the device on which MIPs are to be configured.
- Choose suitable rules for MIP generation. By default, MIP is not configured on a device. If MIPs are to be configured on each port in the MD, you should choose the **default** rule. If MIPs are to be configured only when the low level MDs having MEP, you should choose the **explicit** rule.

According to the diagram as shown in Figure 1-7, perform the following configurations:

- In MD_A, Device B is designed to have MIPs when its port is configured with low level MEPs. In this case, port GigabitEthernet 1/0/3 is configured with MEPs of MD_B, and the MIPs of MD_A can be configured on this port. Based on the design, you should configure the MIP generation rule of MD_A to explicit on Device B.
- The MIPs of MD_B are designed on Device C, and are configured on all ports. Based on this design, the MIP generation rule should be configured as default.

Figure 1-7 Network diagram of MD and MP configuration



Configuration procedure

1) Configure Device B

```
<DeviceB> system-view
[DeviceB] cfd mip-rule explicit service-instance 1
```

2) Configure Device C

```
<DeviceC> system-view
[DeviceC] cfd mip-rule default service-instance 2
```

After the above operation, you can use the **display cfd mp** command to verify your configuration.

Configuring LB on MEPs

Network requirements

Use the LB function to trace the fault source after CC detects a link fault.

As shown in Figure 1-6, enable LB on Device A so that Device A can send LBM messages to MEPs on Device D.

Configuration procedure

Configure Device A

```
<DeviceA> system-view
[DeviceA] cfd loopback service-instance 1 mep 1001 target-mep 4002
```

Configuring LT on MEPs

Network requirements

Use the LT function to find the path and locate the fault after you obtain the state of the entire network through the CC.

As shown in Figure 1-6, enable LT on Device A so that Device A can send LTM messages to the MEP on Device D.

Configuration procedure

Configure Device A

```
<DeviceA> system-view
[DeviceA] cfd linktrace service-instance 1 mep 1001 target-mep 4002
```

Table of Contents

1 MSTP Configuration

When configuring MSTP, go to these sections for information you are interested in:

MSTP Overview

Introduction to STP

Why STP?

The Spanning Tree Protocol (STP) was developed based on the 802.1d standard of IEEE to eliminate loops at the data link layer in a local area network (LAN). Devices running this protocol detect loops in the network by exchanging information with one another and eliminate loops by selectively blocking certain ports to prune the loop structure into a loop-free tree structure. This avoids proliferation and infinite cycling of packets that would occur in a loop network and prevents decreased performance of network devices caused by duplicate packets received.

In the narrow sense, STP refers to IEEE 802.1d STP; in the broad sense, STP refers to the IEEE 802.1d STP and various enhanced spanning tree protocols derived from that protocol.

Protocol Packets of STP

STP uses bridge protocol data units (BPDUs), also known as configuration messages, as its protocol packets.

STP-enabled network devices exchange BPDUs to establish a spanning tree. BPDUs contain sufficient information for the network devices to complete spanning tree calculation.

In STP, BPDUs come in two types:

- Configuration BPDUs, used for calculating a spanning tree and maintaining the spanning tree topology.
- Topology change notification (TCN) BPDUs, used for notifying the concerned devices of network topology changes, if any.

Basic concepts in STP

1) Root bridge

A tree network must have a root; hence the concept of root bridge was introduced in STP.

There is one and only one root bridge in the entire network, and the root bridge can change along with changes of the network topology. Therefore, the root bridge is not fixed.

After network convergence, the root bridge generates and sends out configuration BPDUs at a certain interval, and other devices just forward the BPDUs. This mechanism ensures stable topologies.

2) Root port

On a non-root bridge, the port nearest to the root bridge is called the root port. The root port is responsible for communication with the root bridge. Each non-root bridge has one and only one root port. The root bridge has no root port.

3) Designated bridge and designated port

The following table describes designated bridges and designated ports.

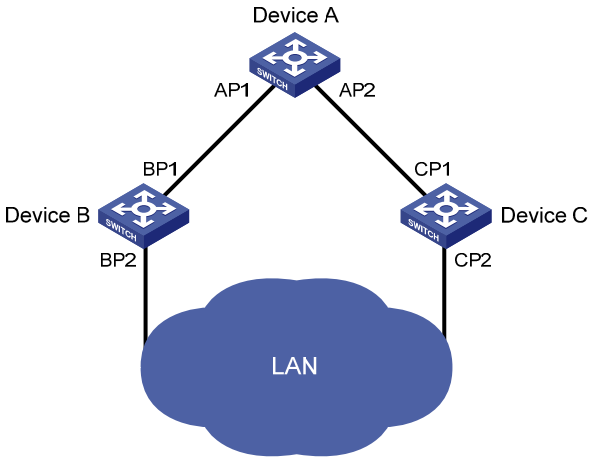
Table 1-1 Description of designated bridges and designated ports:

Classification	Designated bridge	Designated port
For a device	A device directly connected with the local device and responsible for forwarding BPDUs to the local device	The port through which the designated bridge forwards BPDUs to this device
For a LAN	The device responsible for forwarding BPDUs to this LAN segment	The port through which the designated bridge forwards BPDUs to this LAN segment

As shown in Figure 1-1, AP1 and AP2, BP1 and BP2, and CP1 and CP2 are ports on Device A, Device B, and Device C respectively.

- If Device A forwards BPDUs to Device B through AP1, the designated bridge for Device B is Device A, and the designated port of Device B is port AP1 on Device A.
- Two devices are connected to the LAN: Device B and Device C. If Device B forwards BPDUs to the LAN, the designated bridge for the LAN is Device B, and the designated port for the LAN is the port BP2 on Device B.

Figure 1-1 A schematic diagram of designated bridges and designated ports



Note

All the ports on the root bridge are designated ports.

Path cost

Path cost is a reference value used for link selection in STP. By calculating path costs, STP selects relatively robust links and blocks redundant links, and finally prunes the network into a loop-free tree.

How STP works

The devices on a network exchange BPDUs to identify the network topology. Configuration BPDUs contain sufficient information for the network devices to complete spanning tree calculation. Important fields in a configuration BPDU include:

- Root bridge ID: consisting of the priority and MAC address of the root bridge.
- Root path cost: the cost of the path to the root bridge.
- Designated bridge ID: consisting of the priority and MAC address of the designated bridge.
- Designated port ID: designated port priority plus port name.
- Message age: age of the configuration BPDU while it propagates in the network.
- Max age: maximum age of the configuration BPDU.
- Hello time: configuration BPDU transmission interval.
- Forward delay: the delay used by STP bridges to transit the state of the root and designated ports to forwarding.

Note

For simplicity, the descriptions and examples below involve only four fields of configuration BPDUs:

- Root bridge ID (represented by device priority)
- Root path cost (related to the rate of the link connected to the port)
- Designated bridge ID (represented by device priority)
- Designated port ID (represented by port name)

Calculation process of the STP algorithm

1) Initial state

Upon initialization of a device, each port generates a BPDU with itself as the root bridge, in which the root path cost is 0, designated bridge ID is the device ID, and the designated port is the local port.

2) Selection of the optimum configuration BPDU

Each device sends out its configuration BPDU and receives configuration BPDUs from other devices.

The process of selecting the optimum configuration BPDU is as follows:

Table 1-2 Selection of the optimum configuration BPDU

Step	Actions
1	Upon receiving a configuration BPDU on a port, the device performs the following: • If the received configuration BPDU has a lower priority than that of the configuration BPDU generated by the port, the device discards the received configuration BPDU and does not process the configuration BPDU of this port. • If the received configuration BPDU has a higher priority than that of the configuration BPDU generated by the port, the device replaces the content of the configuration BPDU generated by the port with the content of the received configuration BPDU.
2	The device compares the configuration BPDUs of all the ports and chooses the optimum configuration BPDU.

Note

The following are the principles of configuration BPDU comparison:

- The configuration BPDU that has the lowest root bridge ID has the highest priority.
- If all the configuration BPDUs have the same root bridge ID, their root path costs are compared. Assume that the root path cost in a configuration BPDU plus the path cost of a receiving port is S. The configuration BPDU with the smallest S value has the highest priority.
- If all configuration BPDUs have the same ports value, their designated bridge IDs, designated port IDs, and the IDs of the receiving ports are compared in sequence. The configuration BPDU containing a smaller ID wins out.

3) Selection of the root bridge

Initially, each STP-enabled device on the network assumes itself to be the root bridge, with the root bridge ID being its own device ID. By exchanging configuration BPDUs, the devices compare their root bridge IDs to elect the device with the smallest root bridge ID as the root bridge.

4) Selection of the root port and designated ports on a non-root device

The process of selecting the root port and designated ports is as follows:

Table 1-3 Selection of the root port and designated ports

Step	Description
1	A non-root-bridge device regards the port on which it received the optimum configuration BPDU as the root port.
2	Based on the configuration BPDU and the path cost of the root port, the device calculates a designated port configuration BPDU for each of the rest ports. • The root bridge ID is replaced with that of the configuration BPDU of the root port. • The root path cost is replaced with that of the configuration BPDU of the root port plus the path cost of the root port. • The designated bridge ID is replaced with the ID of this device. • The designated port ID is replaced with the ID of this port.

Step	Description
3	The device compares the calculated configuration BPDU with the configuration BPDU on the port of which the port role is to be defined, and acts depending on the comparison result: • If the calculated configuration BPDU is superior, the device considers this port as the designated port, and replaces the configuration BPDU on the port with the calculated configuration BPDU, which will be sent out periodically. • If the configuration BPDU on the port is superior, the device blocks this port without updating its configuration BPDU. The blocked port can receive BPDUs but not send BPDUs or forward data.

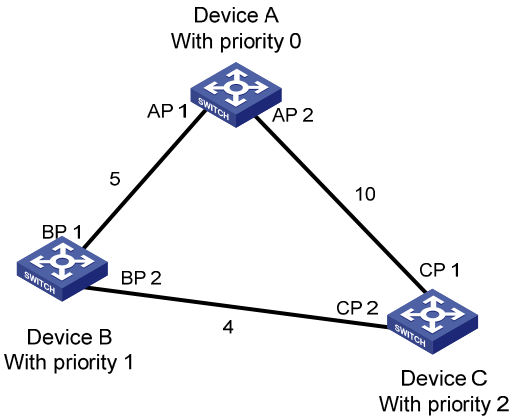
Note

When the network topology is stable, only the root port and designated ports forward traffic, while other ports are all in the blocked state – they receive BPDUs but do not forward BPDUs or user traffic.

A tree-shape topology forms upon successful election of the root bridge, the root port on each non-root bridge and the designated ports.

The following is an example of how the STP algorithm works. As shown in Figure 1-2, assume that the priority of Device A is 0, the priority of Device B is 1, the priority of Device C is 2, and the path costs of these links are 5, 10 and 4 respectively.

Figure 1-2 Network diagram for the STP algorithm



- Initial state of each device

The following table shows the initial state of each device.

Table 1-4 Initial state of each device

Device	Port name	BPDU of port
Device A	AP1	{0, 0, 0, AP1}
	AP2	{0, 0, 0, AP2}
Device B	BP1	{1, 0, 1, BP1}
	BP2	{1, 0, 1, BP2}

Device	Port name	BPDU of port
Device C	CP1	{2, 0, 2, CP1}
	CP2	{2, 0, 2, CP2}

- Comparison process and result on each device

The following table shows the comparison process and result on each device.

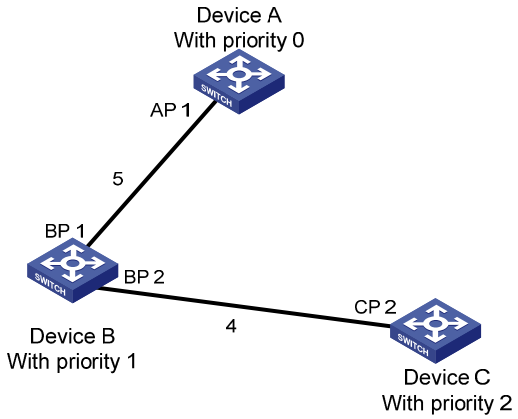
Table 1-5 Comparison process and result on each device

Device	Comparison process	BPDU of port after comparison
Device A	• Port AP1 receives the configuration BPDU of Device B {1, 0, 1, BP1}. Device A finds that the configuration BPDU of the local port {0, 0, 0, AP1} is superior to the received configuration BPDU, and therefore discards the received configuration BPDU. • Port AP2 receives the configuration BPDU of Device C {2, 0, 2, CP1}. Device A finds that the BPDU of the local port {0, 0, 0, AP2} is superior to the received configuration BPDU, and therefore discards the received configuration BPDU. • Device A finds that both the root bridge and designated bridge in the configuration BPDUs of all its ports are itself, so it assumes itself to be the root bridge. In this case, it does not make any change to the configuration BPDU of each port, and starts sending out configuration BPDUs periodically.	AP1: {0, 0, 0, AP1} AP2: {0, 0, 0, AP2}
Device B	• Port BP1 receives the configuration BPDU of Device A {0, 0, 0, AP1}. Device B finds that the received configuration BPDU is superior to the configuration BPDU of the local port {1, 0, 1, BP1}, and updates the configuration BPDU of BP1. • Port BP2 receives the configuration BPDU of Device C {2, 0, 2, CP2}. Device B finds that the configuration BPDU of the local port {1, 0, 1, BP2} is superior to the received configuration BPDU, and therefore discards the received configuration BPDU.	BP1: {0, 0, 0, AP1} BP2: {1, 0, 1, BP2}
	• Device B compares the configuration BPDUs of all its ports, and determines that the configuration BPDU of BP1 is the optimum configuration BPDU. Then, it uses BP1 as the root port, the configuration BPDUs of which will not be changed. • Based on the configuration BPDU of BP1 and the path cost of the root port (5), Device B calculates a designated port configuration BPDU for BP2 {0, 5, 1, BP2}. • Device B compares the calculated configuration BPDU {0, 5, 1, BP2} with the configuration BPDU of BP2. As the calculated BPDU is superior, BP2 will act as the designated port, and the configuration BPDU on this port will be replaced with the calculated configuration BPDU, which will be sent out periodically.	Root port BP1: {0, 0, 0, AP1} Designated port BP2: {0, 5, 1, BP2}

Device	Comparison process	BPDU of port after comparison
Device C	• Port CP1 receives the configuration BPDU of Device A {0, 0, 0, AP2}. Device C finds that the received configuration BPDU is superior to the configuration BPDU of the local port {2, 0, 2, CP1}, and updates the configuration BPDU of CP1. • Port CP2 receives the configuration BPDU of port BP2 of Device B {1, 0, 1, BP2} before the configuration BPDU is updated. Device C finds that the received configuration BPDU is superior to the configuration BPDU of the local port {2, 0, 2, CP2}, and therefore updates the configuration BPDU of CP2.	CP1: {0, 0, 0, AP2} CP2: {1, 0, 1, BP2}
	After comparison: • The configuration BPDU of CP1 is elected as the optimum configuration BPDU, so CP1 is identified as the root port, the configuration BPDUs of which will not be changed. • Device C compares the calculated designated port configuration BPDU {0, 10, 2, CP2} with the configuration BPDU of CP2, and CP2 becomes the designated port, and the configuration BPDU of this port will be replaced with the calculated configuration BPDU.	Root port CP1: {0, 0, 0, AP2} Designated port CP2: {0, 10, 2, CP2}
	• Then, port CP2 receives the updated configuration BPDU of Device B {0, 5, 1, BP2}. Because the received configuration BPDU is superior to its own configuration BPDU, Device C launches a BPDU update process. • At the same time, port CP1 receives periodic configuration BPDUs from Device A. Device C does not launch an update process after comparison.	CP1: {0, 0, 0, AP2} CP2: {0, 5, 1, BP2}
	After comparison: • Because the root path cost of CP2 (9) (root path cost of the BPDU (5) plus path cost corresponding to CP2 (4)) is smaller than the root path cost of CP1 (10) (root path cost of the BPDU (0) + path cost corresponding to CP2 (10)), the BPDU of CP2 is elected as the optimum BPDU, and CP2 is elected as the root port, the messages of which will not be changed. • After comparison between the configuration BPDU of CP1 and the calculated designated port configuration BPDU, port CP1 is blocked, with the configuration BPDU of the port unchanged, and the port will not receive data from Device A until a spanning tree calculation process is triggered by a new event, for example, the link from Device B to Device C going down.	Blocked port CP2: {0, 0, 0, AP2} Root port CP2: {0, 5, 1, BP2}

After the comparison processes described in the table above, a spanning tree with Device A as the root bridge is established as shown in Figure 1-3.

Figure 1-3 The final calculated spanning tree



Note

The spanning tree calculation process in this example is only simplified process.

The BPDU forwarding mechanism in STP

- Upon network initiation, every switch regards itself as the root bridge, generates configuration BPDUs with itself as the root, and sends the configuration BPDUs at a regular hello interval.
- If it is the root port that received a configuration BPDU and the received configuration BPDU is superior to the configuration BPDU of the port, the device increases the message age carried in the configuration BPDU following a certain rule and starts a timer to time the configuration BPDU while sending out this configuration BPDU through the designated port.
- If the configuration BPDU received on a designated port has a lower priority than the configuration BPDU of the local port, the port immediately sends out its own configuration BPDU in response.
- If a path becomes faulty, the root port on this path will no longer receive new configuration BPDUs and the old configuration BPDUs will be discarded due to timeout. In this case, the device will generate a configuration BPDU with itself as the root and send out the BPDUs and TCN BPDUs. This triggers a new spanning tree calculation process to establish a new path to restore the network connectivity.

However, the newly calculated configuration BPDU will not be propagated throughout the network immediately, so the old root ports and designated ports that have not detected the topology change continue forwarding data along the old path. If the new root ports and designated ports begin to forward data as soon as they are elected, a temporary loop may occur.

STP timers

STP calculation involves three important timing parameters: forward delay, hello time, and max age.

- Forward delay is the delay time for device state transition.

A path failure can cause spanning tree re-calculation to adapt the spanning tree structure to the change. However, the resulting new configuration BPDU cannot propagate throughout the network immediately. If the newly elected root ports and designated ports start to forward data right away, a temporary loop is likely to occur.

For this reason, as a mechanism for state transition in STP, the newly elected root ports or designated ports require twice the forward delay time before transiting to the forwarding state to ensure that the new configuration BPDU has propagated throughout the network.

- Hello time is the time interval at which a device sends hello packets to the surrounding devices to ensure that the paths are fault-free.
- Max age is a parameter used to determine whether a configuration BPDU held by the device has expired. A configuration BPDU beyond the max age will be discarded.

Introduction to MSTP

Why MSTP

1) Weakness of STP and RSTP

STP does not support rapid state transition of ports. A newly elected root port or designated port must wait twice the forward delay time before transiting to the forwarding state, even if it is a port on a point-to-point link or an edge port, which directly connects to a user terminal rather than to another device or a shared LAN segment.

The Rapid Spanning Tree Protocol (RSTP) is an optimized version of STP. RSTP allows a newly elected root port or designated port to enter the forwarding state much quicker under certain conditions than in STP. As a result, it takes a shorter time for the network to converge.

Note

- In RSTP, a newly elected root port can enter the forwarding state rapidly if this condition is met: The old root port on the device has stopped forwarding data and the upstream designated port has started forwarding data.
- In RSTP, a newly elected designated port can enter the forwarding state rapidly if this condition is met: The designated port is an edge port or a port connected with a point-to-point link. If the designated port is an edge port, it can enter the forwarding state directly; if the designated port is connected with a point-to-point link, it can enter the forwarding state immediately after the device undergoes handshake with the downstream device and gets a response.

Although RSTP supports rapid network convergence, it has the same drawback as STP does: All bridges within a LAN share the same spanning tree, so redundant links cannot be blocked based on VLAN, and the packets of all VLANs are forwarded along the same spanning tree.

2) Features of MSTP

The Multiple Spanning Tree Protocol (MSTP) overcomes the shortcomings of STP and RSTP. In addition to the support for rapid network convergence, it also allows data flows of different VLANs to be forwarded along separate paths, thus providing a better load sharing mechanism for redundant links. For description about VLANs, refer to *VLAN Configuration* in the *Access Volume*.

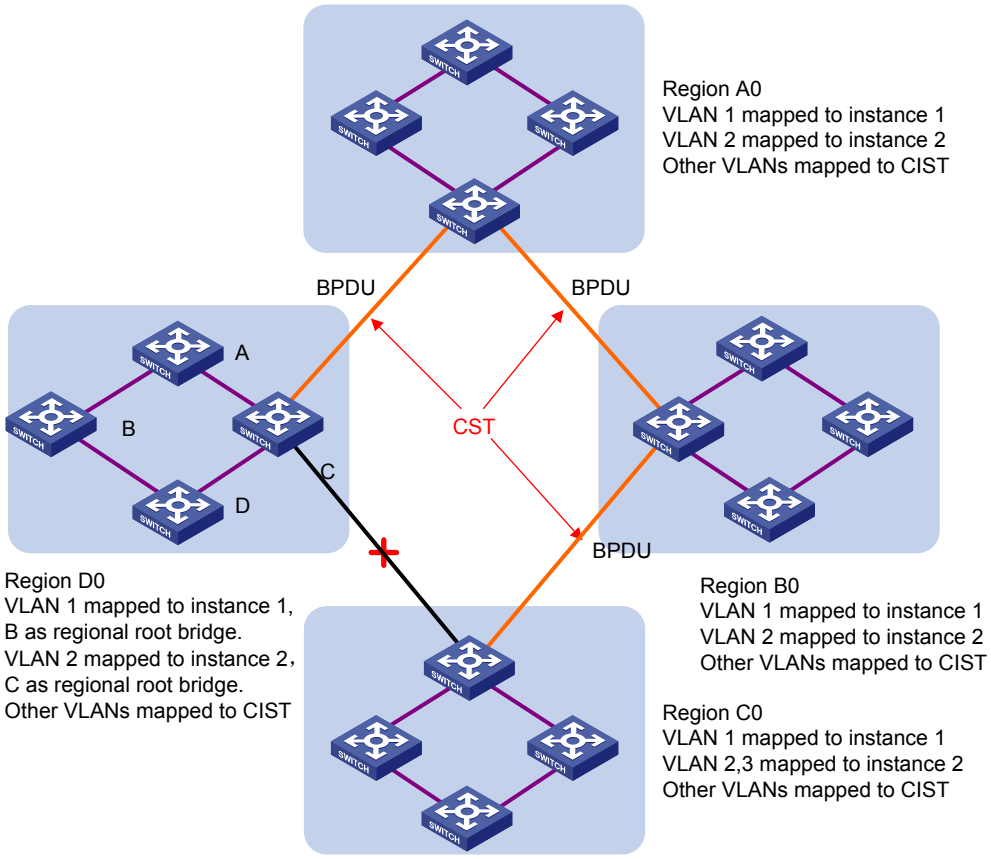
MSTP features the following:

- MSTP supports mapping VLANs to MST instances (MSTIs) by means of a VLAN-to-MSTI mapping table. MSTP can reduce communication overheads and resource usage by mapping multiple VLANs to one MSTI.

- MSTP divides a switched network into multiple regions, each containing multiple spanning trees that are independent of one another.
- MSTP prunes a loop network into a loop-free tree, thus avoiding proliferation and endless cycling of packets in a loop network. In addition, it provides multiple redundant paths for data forwarding, thus supporting load balancing of VLAN data.
- MSTP is compatible with STP and RSTP.

Basic concepts in MSTP

Figure 1-4 Basic concepts in MSTP



Assume that all devices in Figure 1-4 are running MSTP. This section explains some basic concepts of MSTP.

1) MST region

A multiple spanning tree region (MST region) consists of multiple devices in a switched network and the network segments among them. These devices have the following characteristics:

- All are MSTP-enabled,
- They have the same region name,
- They have the same VLAN-to-MSTI mapping configuration,
- They have the same MSTP revision level configuration, and
- They are physically linked with one another.

For example, all the devices in region A0 in Figure 1-4 have the same MST region configuration:

- The same region name,
- The same VLAN-to-MSTI mapping configuration (VLAN 1 is mapped to MSTI 1, VLAN 2 to MSTI 2, and the rest to the common and internal spanning tree (CIST, that is, MSTI 0), and
- The same MSTP revision level (not shown in the figure).

Multiple MST regions can exist in a switched network. You can use an MSTP command to assign multiple devices to the same MST region.

2) VLAN-to-MSTI mapping table

As an attribute of an MST region, the VLAN-to-MSTI mapping table describes the mapping relationships between VLANs and MSTIs. In Figure 1-4, for example, the VLAN-to-MSTI mapping table of region A0 is as follows: VLAN 1 is mapped to MSTI 1, VLAN 2 to MSTI 2, and the rest to CIST. MSTP achieves load balancing by means of the VLAN-to-MSTI mapping table.

3) IST

An internal spanning tree (IST) is a spanning tree that runs in an MST region.

ISTs in all MST regions and the common spanning tree (CST) jointly constitute the common and internal spanning tree (CIST) of the entire network. An IST is a section of the CIST.

In Figure 1-4, for example, the CIST has a section in each MST region, and this section is the IST in the respective MST region.

4) CST

The CST is a single spanning tree that connects all MST regions in a switched network. If you regard each MST region as a “device”, the CST is a spanning tree calculated by these “devices” through STP or RSTP. For example, the red lines in Figure 1-4 represent the CST.

5) CIST

Jointly constituted by ISTs and the CST, the CIST is a single spanning tree that connects all devices in a switched network.

In Figure 1-4, for example, the ISTs in all MST regions plus the inter-region CST constitute the CIST of the entire network.

6) MSTI

Multiple spanning trees can be generated in an MST region through MSTP, one spanning tree being independent of another. Each spanning tree is referred to as a multiple spanning tree instance (MSTI). In Figure 1-4, for example, multiple spanning trees can exist in each MST region, each spanning tree corresponding to the specific VLAN(s). These spanning trees are called MSTIs.

7) Regional root bridge

The root bridge of the IST or an MSTI within an MST region is the regional root bridge of the IST or the MSTI. Based on the topology, different spanning trees in an MST region may have different regional roots.

For example, in region D0 in Figure 1-4, the regional root of MSTI 1 is device B, while that of MSTI 2 is device C.

8) Common root bridge

The common root bridge is the root bridge of the CIST.

In Figure 1-4, for example, the common root bridge is a device in region A0.

9) Boundary port

A boundary port is a port that connects an MST region to another MST region, or to a single spanning-tree region running STP, or to a single spanning-tree region running RSTP. In Figure 1-4, for example, if a device in region A0 is interconnected with the first port of a device in region D0 and the common root bridge of the entire switched network is located in region A0, the first port of that device in region D0 is the boundary port of region D0.

During MSTP calculation, a boundary port's role on an MSTI is consistent with its role on the CIST. But that is not true with master ports. A master port on MSTIs is a root port on the CIST.

10) Roles of ports

MSTP calculation involves these port roles: root port, designated port, master port, alternate port, backup port, and so on.

- Root port: a port responsible for forwarding data to the root bridge.
- Designated port: a port responsible for forwarding data to the downstream network segment or device.
- Master port: A port on the shortest path from the current region to the common root bridge, connecting the MST region to the common root bridge. If the region is seen as a node, the master port is the root port of the region on the CST. The master port is a root port on IST/CIST and still a master port on the other MSTIs.
- Alternate port: The standby port for the root port and the master port. When the root port or master port is blocked, the alternate port becomes the new root port or master port.
- Backup port: The backup port of a designated port. When the designated port is blocked, the backup port becomes a new designated port and starts forwarding data without delay. A loop occurs when two ports of the same MSTP device are interconnected. Therefore, the device will block either of the two ports, and the backup port is that port to be blocked.

A port can play different roles in different MSTIs.

Figure 1-5 Port roles

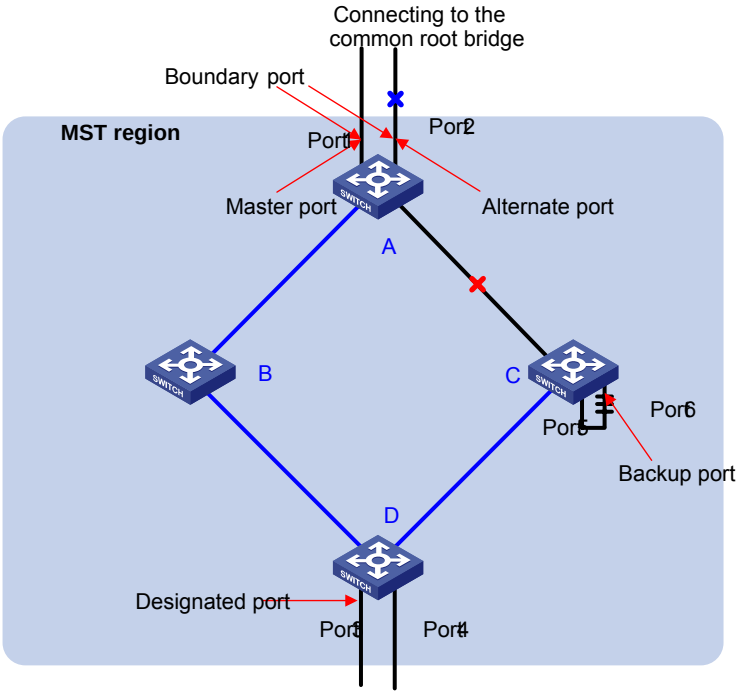


Figure 1-5 helps understand these concepts. In this figure:

- Devices A, B, C, and D constitute an MST region.
- Port 1 and port 2 of device A connect to the common root bridge.
- Port 5 and port 6 of device C form a loop.
- Port 3 and port 4 of device D connect downstream to other MST regions.

11) Port states

In MSTP, port states fall into the following three:

- Forwarding: the port learns MAC addresses and forwards user traffic;
- Learning: the port learns MAC addresses but does not forward user traffic;
- Discarding: the port neither learns MAC addresses nor forwards user traffic.

Note

When in different MSTIs, a port can be in different states.

A port state is not exclusively associated with a port role. Table 1-6 lists the port state(s) supported by each port role ("√" indicates that the port supports this state, while "—" indicates that the port does not support this state).

Table 1-6 Ports states supported by different port roles

Port role (right) / Port state (below)	Root port/master port	Designated port	Alternate port	Backup port
Forwarding	√	√	—	—
Learning	√	√	—	—
Discarding	√	√	√	√

How MSTP works

MSTP divides an entire Layer 2 network into multiple MST regions, which are interconnected by a calculated CST. Inside an MST region, multiple spanning trees are calculated, each being called an MSTI. Among these MSTIs, MSTI 0 is the IST, while all the others are MSTIs. Similar to STP, MSTP uses configuration BPDUs to calculate spanning trees. The only difference between the two protocols is that an MSTP BPDU carries the MSTP configuration on the device from which this BPDU is sent.

1) CIST calculation

The calculation of a CIST tree is also the process of configuration BPDU comparison. During this process, the device with the highest priority is elected as the root bridge of the CIST. MSTP generates an IST within each MST region through calculation, and, at the same time, MSTP regards each MST region as a single device and generates a CST among these MST regions through calculation. The CST and ISTs constitute the CIST of the entire network.

2) MSTI calculation

Within an MST region, MSTP generates different MSTIs for different VLANs based on the VLAN-to-MSTI mappings. MSTP performs a separate calculation process, which is similar to spanning tree calculation in STP, for each spanning tree. For details, refer to How STP works.

In MSTP, a VLAN packet is forwarded along the following paths:

- Within an MST region, the packet is forwarded along the corresponding MSTI.
- Between two MST regions, the packet is forwarded along the CST.

Implementation of MSTP on devices

MSTP is compatible with STP and RSTP. STP and RSTP protocol packets can be recognized by devices running MSTP and used for spanning tree calculation.

In addition to basic MSTP functions, many special functions are provided for ease of management, as follows:

- Root bridge hold
- Root bridge backup
- Root guard
- BPDU guard
- Loop guard
- TC-BPDU guard

Protocols and Standards

MSTP is documented in:

- IEEE 802.1d: Spanning Tree Protocol
- IEEE 802.1w: Rapid Spanning Tree Protocol
- IEEE 802.1s: Multiple Spanning Tree Protocol

Configuration Task List

Before configuring MSTP, you need to know the position of each device in each MSTI: root bridge or leave node. In each MSTI, one, and only one device acts as the root bridge, while all others as leaf nodes.

Complete these tasks to configure MSTP:

Task		Remarks
Configuring the Root Bridge	Configuring an MST Region	Required
	Specifying the Root Bridge or a Secondary Root Bridge	Optional
	Configuring the Work Mode of an MSTP Device	Optional
	Configuring the Priority of the Current Device	Optional
	Configuring the Maximum Hops of an MST Region	Optional
	Configuring the Network Diameter of a Switched Network	Optional
	Configuring Timers of MSTP	Optional
	Configuring the Timeout Factor	Optional
	Configuring the Maximum Port Rate	Optional
	Configuring Ports as Edge Ports	Optional
	Setting the Link Type of a Port to P2P	Optional
	Configuring the Mode a Port Uses to Recognize/Send MSTP Packets	Optional
	Enabling the Output of Port State Transition Information	Optional
	Enabling the MSTP Feature	Required

Task		Remarks
Configuring Leaf NodesConfiguring Leaf Nodes	Configuring an MST Region	Required
	Configuring the Work Mode of an MSTP Device	Optional
	Configuring the Timeout Factor	Optional
	Configuring the Maximum Port Rate	Optional
	Configuring Ports as Edge Ports	Optional
	Configuring Path Costs of Ports	Optional
	Configuring Port Priority	Optional
	Setting the Link Type of a Port to P2P	Optional
	Configuring the Mode a Port Uses to Recognize/Send MSTP Packets	Optional
	Enabling the Output of Port State Transition Information	Optional
	Enabling the MSTP Feature	Required
Performing mCheck		Optional
Configuring Digest Snooping		Optional
Configuring No Agreement Check		Optional
Configuring Protection Functions		Optional

Note

- If both GVRP and MSTP are enabled on a device at the same time, GVRP packets will be forwarded along the CIST. Therefore, if you wish to advertise a certain VLAN within the network through GVRP in this case, make sure that this VLAN is mapped to the CIST (MSTI 0) when configuring the VLAN-to-MSTI mapping table. For the detailed information of GVRP, refer to *GVRP Configuration* of the *Access Volume*.
- MSTP is mutually exclusive with any of the following functions on a port: service loopback, RRPP, Smart Link, and BPDU tunnel.
- Configurations made in Layer-2 aggregate interface view can take effect only on the aggregate interface; configurations made on an aggregation member port can take effect only after the port is removed from the aggregation group. For detailed information about link aggregation, refer to *Link Aggregation Configuration* in the *Access Volume*.
- After you enable MSTP on a Layer-2 aggregate interface, the system performs MSTP calculation on the Layer-2 aggregate interface but not on the aggregation member ports. The MSTP enable state and forwarding state of each selected port in an aggregation group is consistent with those of the corresponding Layer-2 aggregate interface.
- Though the member port of an aggregation group does not participate in MSTP calculation, the port still reserves its MSTP configurations for participating MSTP calculation after leaving the aggregation group.

Configuring the Root Bridge

Configuring an MST Region

Configuration procedure

Follow these steps to configure an MST region:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MST region view	**stp region-configuration**	—
Configure the MST region name	**region-name** *name*	Optional The MST region name is the MAC address by default.
Configure the VLAN-to-MSTI mapping table	**instance** *instance-id* **vlan** *vlan-list*	Optional Use either command. All VLANs in an MST region are mapped to MSTI 0 by default.
	vlan-mapping modulo *modulo*	
Configure the MSTP revision level of the MST region	**revision-level** *level*	Optional 0 by default
Activate MST region configuration manually	**active region-configuration**	Required
Display all the configuration information of the MST region	**check region-configuration**	Optional
Display the currently effective MST region configuration information	**display stp region-configuration**	The **display** command can be executed in any view.

Note

Two or more MSTP-enabled devices belong to the same MST region only if they are configured to have the same MST region name, the same VLAN-to-MSTI mapping entries in the MST region and the same MST region revision level, and they are interconnected via a physical link.

The configuration of MST region–related parameters, especially the VLAN-to-MSTI mapping table, will cause MSTP to launch a new spanning tree calculation process, which may result in network topology instability. To reduce the possibility of topology instability caused by configuration, MSTP will not immediately launch a new spanning tree calculation process when processing MST region–related configurations; instead, such configurations will take effect only after you:

- activate the MST region–related parameters using the **active region-configuration** command, or
- enable MSTP using the **stp enable** command.

Configuration example

Configure the MST region name to be “info”, the MSTP revision level to be 1, and VLAN 2 through VLAN 10 to be mapped to MSTI 1 and VLAN 20 through VLAN 30 to MSTI 2.

```
<Sysname> system-view
[Sysname] stp region-configuration
[Sysname-mst-region] region-name info
[Sysname-mst-region] instance 1 vlan 2 to 10
[Sysname-mst-region] instance 2 vlan 20 to 30
[Sysname-mst-region] revision-level 1
[Sysname-mst-region] active region-configuration
```

Specifying the Root Bridge or a Secondary Root Bridge

MSTP can determine the root bridge of a spanning tree through MSTP calculation. Alternatively, you can specify the current device as the root bridge using the commands provided by the system.

Specifying the current device as the root bridge of a specific spanning tree

Follow these steps to specify the current device as the root bridge of a specific spanning tree:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Specify the current device as the root bridge of a specific spanning tree	**stp** [**instance** *instance-id*] **root primary**	Required By default, a device does not function as the root bridge.

Specifying the current device as a secondary root bridge of a specific spanning tree

Follow these steps to specify the current device as a secondary root bridge of a specific spanning tree:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Specify the current device as a secondary root bridge of a specific spanning tree	**stp** [**instance** *instance-id*] **root secondary**	Required By default, a device does not function as a secondary root bridge.

Note that:

- After specifying the current device as the root bridge or a secondary root bridge, you cannot change the priority of the device.
- You can configure the current device as the root bridge or a secondary root bridge of an MSTI, which is specified by **instance** *instance-id* in the command. If you set *instance-id* to 0, the current device will be the root bridge or a secondary root bridge of the CIST.
- The current device has independent roles in different MSTIs. It can act as the root bridge or a secondary root bridge of one instance while it can also act as the root bridge or a secondary root bridge of another MSTI. However, the same device cannot be the root bridge and a secondary root bridge in the same MSTI at the same time.

- There is one and only one root bridge in effect in a spanning tree instance. If two or more devices have been designated to be root bridges of the same spanning tree instance, MSTP will select the device with the lowest MAC address as the root bridge.
- You can specify multiple secondary root bridges for the same instance. Namely, you can specify secondary root bridges for the same instance on two or more than two devices.
- When the root bridge of an instance fails or is shut down, the secondary root bridge (if you have specified one) can take over the role of the primary root bridge. However, if you specify a new primary root bridge for the instance at this time, the secondary root bridge will not become the root bridge. If you have specified multiple secondary root bridges for an instance, when the root bridge fails, MSTP will select the secondary root bridge with the lowest MAC address as the new root bridge.
- Alternatively, you can also specify the current device as the root bridge by setting the priority of the device to 0. For the device priority configuration, refer to Configuring the Priority of the Current Device.

Configuration example

Specify the current device as the root bridge of MSTI 1 and a secondary root bridge of MSTI 2.

```
<Sysname> system-view
[Sysname] stp instance 1 root primary
[Sysname] stp instance 2 root secondary
```

Configuring the Work Mode of an MSTP Device

MSTP and RSTP can recognize each other's protocol packets, so they are mutually compatible. However, STP is unable to recognize MSTP packets. For hybrid networking with legacy STP devices and for full interoperability with RSTP-enabled devices, MSTP supports three work modes: STP-compatible mode, RSTP mode, and MSTP mode.

- In STP-compatible mode, all ports of the device send out STP BPDUs,
- In RSTP mode, all ports of the device send out RSTP BPDUs. If the device detects that it is connected with a legacy STP device, the port connecting with the legacy STP device will automatically migrate to STP-compatible mode.
- In MSTP mode, all ports of the device send out MSTP BPDUs. If the device detects that it is connected with a legacy STP device, the port connecting with the legacy STP device will automatically migrate to STP-compatible mode.

Configuration procedure

Follow these steps to configure the MSTP work mode:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the work mode of MSTP	**stp mode** { **stp** | **rstp** | **mstp** }	Optional MSTP mode by default

Configuration example

Configure MSTP to work in STP-compatible mode.

```
<Sysname> system-view
```

```
[Sysname] stp mode stp
```

Configuring the Priority of the Current Device

The priority of a device determines whether it can be elected as the root bridge of a spanning tree. A lower value indicates a higher priority. By setting the priority of a device to a low value, you can specify the device as the root bridge of the spanning tree. An MSTP-enabled device can have different priorities in different MSTIs.

Configuration procedure

Follow these steps to configure the priority of the current device in a specified MSTI:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the priority of the current device in a specified MSTI	**stp** [**instance** *instance-id*] **priority** *priority*	Optional 32768 by default

Caution

- After specifying the current device as the root bridge or a secondary root bridge, you cannot change the priority of the device.
- During root bridge selection, if all devices in a spanning tree have the same priority, the one with the lowest MAC address will be selected as the root bridge of the spanning tree.

Configuration example

Set the device priority in MSTI 1 to 4096.

```
<Sysname> system-view
[Sysname] stp instance 1 priority 4096
```

Configuring the Maximum Hops of an MST Region

By setting the maximum hops of an MST region, you can restrict the region size. The maximum hops configured on the regional root bridge will be used as the maximum hops of the MST region.

The regional root bridge always sends a configuration BPDU with a hop count set to the maximum value. When a switch receives this configuration BPDU, it decrements the hop count by 1 and uses the new hop count in the BPDUs it propagates. When the hop count of a BPDU reaches 0, it is discarded by the device that received it. Thus, devices beyond the reach of the maximum hop can no longer take part in spanning tree calculation, and thereby the size of the MST region is confined.

All the devices other than the root bridge in the MST region use the maximum hop value set for the root bridge.

Configuration procedure

Follow these steps to configure the maximum number of hops of the MST region:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the maximum hops of the MST region	**stp max-hops** *hops*	Optional 20 by default

Note

A larger maximum hops setting means a larger size of the MST region. Only the maximum hops configured on the regional root bridge can restrict the size of the MST region.

Configuration example

Set the maximum hops of the MST region to 30.

```
<Sysname> system-view
[Sysname] stp max-hops 30
```

Configuring the Network Diameter of a Switched Network

Any two stations in a switched network are interconnected through a specific path composed of a series of devices. The network diameter is the number of devices on the path composed of the most devices.

Configuration procedure

Follow these steps to configure the network diameter of the switched network:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the network diameter of the switched network	**stp bridge-diameter** *bridge-number*	Optional 7 by default

Note

- The network diameter is a parameter that indicates the network size. A bigger network diameter represents a larger network size.
- Based on the network diameter you configured, MSTP automatically sets an optimal hello time, forward delay, and max age for the device.
- The configured network diameter is effective for the CIST only, and not for MSTIs. Each MST region is considered as a device.

Configuration example

Set the network diameter of the switched network to 6.

```
<Sysname> system-view
[Sysname] stp bridge-diameter 6
```

Configuring Timers of MSTP

MSTP involves three timers: forward delay, hello time and max age. You can configure these three parameters for MSTP to calculate spanning trees.

Configuration procedure

Follow these steps to configure the timers of MSTP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the forward delay timer	**stp timer forward-delay** *centi-seconds*	Optional 1,500 centiseconds (15 seconds) by default
Configure the hello timer	**stp timer hello** *centi-seconds*	Optional 200 centiseconds (2 seconds) by default
Configure the max age timer	**stp timer max-age** *centi-seconds*	Optional 2,000 centiseconds (20 seconds) by default

These three timers set on the root bridge of the CIST apply on all the devices on the entire switched network.

Caution

- The length of the forward delay time is related to the network diameter of the switched network. Typically, the larger the network diameter is, the longer the forward delay time should be. Note that if the forward delay setting is too small, temporary redundant paths may be introduced; if the forward delay setting is too big, it may take a long time for the network to converge. We recommend that you use the default setting.
- An appropriate hello time setting enables the device to timely detect link failures on the network without using excessive network resources. If the hello time is set too long, the device will take packet loss as a link failure and trigger a new spanning tree calculation process; if the hello time is set too short, the device will send repeated configuration BPDUs frequently, which adds to the device burden and causes waste of network resources. We recommend that you use the default setting.
- If the max age time setting is too small, the network devices will frequently launch spanning tree calculations and may take network congestion as a link failure; if the max age setting is too large, the network may fail to timely detect link failures and fail to timely launch spanning tree calculations, thus reducing the auto-sensing capability of the network. We recommend that you use the default setting.

The settings of hello time, forward delay and max age must meet the following formulae; otherwise network instability will frequently occur.

- 2 × (forward delay – 1 second) ⩾ max age
- Max age ⩾ 2 × (hello time + 1 second)

We recommend that you specify the network diameter with the **stp root primary** command and let MSTP automatically calculate optimal settings of these three timers.

Configuration example

Set the forward delay to 1,600 centiseconds, hello time to 300 centiseconds, and max age to 2,100 centiseconds.

```
<Sysname> system-view
[Sysname] stp timer forward-delay 1600
[Sysname] stp timer hello 300
[Sysname] stp timer max-age 2100
```

Configuring the Timeout Factor

After the network topology is stabilized, each non-root-bridge device forwards configuration BPDUs to the downstream devices at the interval of hello time to check whether any link is faulty. Typically, if a device does not receive a BPDU from the upstream device within nine times the hello time, it will assume that the upstream device has failed and start a new spanning tree calculation process.

In a very stable network, this kind of spanning tree calculation may occur because the upstream device is busy. In this case, you can avoid such unwanted spanning tree calculation by lengthening the timeout time.

Configuration procedure

Follow these steps to configure the timeout factor:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the timeout factor of the device	**stp timer-factor** *number*	Optional 3 by default

Note

- Timeout time = timeout factor × 3 × hello time.
- Typically, we recommend that you set the timeout factor to 5, or 6, or 7 for a stable network.

Configuration example

Set the timeout factor to 6.

```
<Sysname> system-view
[Sysname] stp timer-factor 6
```

Configuring the Maximum Port Rate

The maximum rate of a port refers to the maximum number of MSTP packets that the port can send within each hello time. The maximum rate of a port is related to the physical status of the port and the network structure.

Configuration procedure

Follow these steps to configure the maximum rate of a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the maximum rate of the port(s)		**stp transmit-limit** *packet-number*	Optional 10 by default

Note

If the maximum rate setting of a port is too big, the port will send a large number of MSTP packets within each hello time, thus using excessive network resources. We recommend that you use the default setting.

Configuration example

Set the maximum transmission rate of port GigabitEthernet 1/0/1 to 5.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp transmit-limit 5
```

Configuring Ports as Edge Ports

If a port directly connects to a user terminal rather than another device or a shared LAN segment, this port is regarded as an edge port. When a network topology change occurs, an edge port will not cause a temporary loop. Because a device does not know whether a port is directly connected to a terminal, you need to manually configure the port to be an edge port. After that, this port can transition rapidly from the blocked state to the forwarding state without delay.

Configuration procedure

Follow these steps to specify a port or a group of ports as edge port(s):

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the port(s) as edge port(s)		**stp edged-port enable**	Required All Ethernet ports are non-edge ports by default.

Note

- With BPDU guard disabled, when a port set as an edge port receives a BPDU from another port, it will become a non-edge port again. To restore the edge port, re-enable it.
- If a port directly connects to a user terminal, configure it as an edge port and enable BPDU guard for it. This enables the port to transition to the forwarding state fast while ensuring network security.

Configuration example

Configure GigabitEthernet 1/0/1 to be an edge port.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp edged-port enable
```

Setting the Link Type of a Port to P2P

A point-to-point link is a link directly connecting two devices. If the two ports across a point-to-point link are root ports or designated ports, the ports can rapidly transition to the forwarding state after a proposal-agreement handshake process.

Configuration procedure

Follow these steps to set the type of a connected link to P2P:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Set the link type to P2P		**stp point-to-point** { **auto** | **force-false** | **force-true** }	Optional The default setting is **auto**; namely the port automatically detects whether its link is point-to-point.

Note

- A Layer-2 aggregate interface can be configured to connect to a point-to-point link. If a port works in auto-negotiation mode and the negotiation result is full duplex, this port can be configured as connecting to a point-to-point link.
- If a port is configured as connecting to a point-to-point link, the setting takes effect for the port in all MSTIs. If the physical link to which the port connects is not a point-to-point link and you force it to be a point-to-point link by configuration, the configuration may incur a temporary loop.

Configuration example

Configure port GigabitEthernet 1/0/1 as connecting to a point-to-point link.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp point-to-point force-true
```

Configuring the Mode a Port Uses to Recognize/Send MSTP Packets

A port can send/recognize MSTP packets of two formats:

- 802.1s-compliant standard format, and
- Compatible format

By default, the packet format recognition mode of a port is **auto**, namely the port automatically distinguishes the two MSTP packet formats, and determines the format of packets it will send based on the recognized format. You can configure the MSTP packet format to be used by a port. After the configuration, when working in MSTP mode, the port sends and receives only MSTP packets of the format you have configured to communicate with devices that send packets of the same format.

Configuration procedure

Follow these steps to configure the MSTP packet format to be supported by a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the mode the port uses to recognize/send MSTP packets		**stp compliance** { **auto** | **dot1s** | **legacy** }	Optional **auto** by default

Note

- MSTP provides the MSTP packet format incompatibility guard function. In MSTP mode, if a port is configured to recognize/send MSTP packets in a mode other than **auto**, and if it receives a packet in a format different from the specified type, the port will become a designated port and remain in the discarding state to prevent the occurrence of a loop.
- MSTP provides the MSTP packet format frequent change guard function. If a port receives MSTP packets of different formats frequently, this means that the MSTP packet format configuration contains errors. In this case, if the port is working in MSTP mode, it will be disabled for protection. Those ports closed thereby can be restored only by the network administers.

Configuration example

Configure GigabitEthernet 1/0/1 to receive and send standard-format MSTP packets.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp compliance dot1s
```

Enabling the Output of Port State Transition Information

In a large-scale, MSTP-enabled network, there are a large number of MSTIs, so ports may frequently transition from one state to another. In this situation, you can enable devices to output the port state transition information of all MSTIs or the specified MSTI so as to monitor the port states in real time.

Follow these steps to enable output of port state transition information:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable output of port state transition information of all MSTIs or a particular MSTI	**stp port-log** { **all** | **instance** *instance-id* }	Optional This function is enabled by default.

Enabling the MSTP Feature

Configuration procedure

Follow these steps to enable the MSTP feature:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enable the MSTP feature for the device		**stp enable**	Required Enabled by default.
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable the MSTP feature for the port(s)		**stp enable**	Optional By default, MSTP is enabled on all ports.

Note

- MSTP takes effect when it is enabled both globally and on the port.
- You must enable MSTP for the device before any other MSTP-related configuration can take effect.
- To control MSTP flexibly, you can use the **undo stp enable** command to disable the MSTP feature for certain ports so that they will not take part in spanning tree calculation and thus to save the device's CPU resources.

Configuration example

Enable MSTP globally and disable MSTP for port GigabitEthernet 1/0/1.

```
<Sysname> system-view
[Sysname] stp enable
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] undo stp enable
```

Configuring Leaf Nodes

Configuring an MST Region

Refer to Configuring an MST Region in the section about root bridge configuration.

Configuring the Work Mode of MSTP

Refer to Configuring the Work Mode of an MSTP Device in the section about root bridge configuration.

Configuring the Timeout Factor

Refer to Configuring the Timeout Factor in the section about root bridge configuration.

Configuring the Maximum Transmission Rate of Ports

Refer to Configuring the Maximum Port Rate in the section about root bridge configuration.

Configuring Ports as Edge Ports

Refer to Configuring Ports as Edge Ports in the section about root bridge configuration.

Configuring Path Costs of Ports

Path cost is a parameter related to the rate of a port. On an MSTP-enabled device, a port can have different path costs in different MSTIs. Setting appropriate path costs allows VLAN traffic flows to be forwarded along different physical links, thus to achieve VLAN-based load balancing.

The device can automatically calculate the default path cost; alternatively, you can also configure the path cost for ports.

Specifying a standard that the device uses when calculating the default path cost

You can specify a standard for the device to use in automatic calculation for the default path cost. The device supports the following standards:

- **dot1d-1998**: The device calculates the default path cost for ports based on IEEE 802.1d-1998.
- **dot1t**: The device calculates the default path cost for ports based on IEEE 802.1t.
- **legacy**: The device calculates the default path cost for ports based on a private standard.

Follow these steps to specify a standard for the device to use when calculating the default path cost:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Specify a standard for the device to use when calculating the default path costs for ports of the device	**stp pathcost-standard** { **dot1d-1998 | dot1t** | **legacy** }	Optional The default standard used by the device is **legacy**.

Table 1-7 Link speed vs. path cost

Link speed	Duplex state	802.1d-1998	802.1t	Private standard
0	—	65535	200,000,000	200,000
10 Mbps	Single Port	100	2,000,000	2,000
	Aggregate Link 2 Ports	100	1,000,000	1,800
	Aggregate Link 3 Ports	100	666,666	1,600
	Aggregate Link 4 Ports	100	500,000	1,400
100 Mbps	Single Port	19	200,000	200
	Aggregate Link 2 Ports	19	100,000	180
	Aggregate Link 3 Ports	19	66,666	160
	Aggregate Link 4 Ports	19	50,000	140
1000 Mbps	Single Port	4	20,000	20
	Aggregate Link 2 Ports	4	10,000	18
	Aggregate Link 3 Ports	4	6,666	16
	Aggregate Link 4 Ports	4	5,000	14
10 Gbps	Single Port	2	2,000	2
	Aggregate Link 2 Ports	2	1,000	1
	Aggregate Link 3 Ports	2	666	1
	Aggregate Link 4 Ports	2	500	1

Note

When calculating path cost for an aggregate interface, 802.1d-1998 does not take into account the number of member ports in its aggregation group as 802.1t does. The calculation formula of 802.1t is: Path Cost = 200,000,000/link speed (in 100 kbps), where link speed is the sum of the link speed values of the non-blocked ports in the aggregation group.

Configuring Path Costs of Ports

Follow these steps to configure the path cost of ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the path cost of the port(s)		**stp** [**instance** *instance-id*] **cost** *cost*	Required By default, MSTP automatically calculates the path cost of each port.

Caution

- If you change the standard that the device uses in calculating the default path cost, the port path cost value set through the **stp cost** command will be invalid.
- When the path cost of a port is changed, MSTP will re-calculate the role of the port and initiate a state transition. If you use 0 as *instance-id*, you are setting the path cost of the CIST.

Configuring Port Priority

The priority of a port is an important factor in determining whether the port can be elected as the root port of a device. If all other conditions are the same, the port with the highest priority will be elected as the root port.

On an MSTP-enabled device, a port can have different priorities in different MSTIs, and the same port can play different roles in different MSTIs, so that data of different VLANs can be propagated along different physical paths, thus implementing per-VLAN load balancing. You can set port priority values based on the actual networking requirements.

Configuration procedure

Follow these steps to configure the priority of a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure a priority for the port(s)		**stp** [**instance** *instance-id*] **port priority** *priority*	Optional 128 for all Ethernet ports by default.

Note

- When the priority of a port is changed, MSTP will re-calculate the role of the port and initiate a state transition.
- Generally, a lower configured value indicates a higher priority. If you configure the same priority value for all the ports on a device, the specific priority of a port depends on the index number of the port. Changing the priority of a port triggers a new spanning tree calculation process.

Configuration example

Set the priority of port GigabitEthernet 1/0/1 to 16 in MSTI 1.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp instance 1 port priority 16
```

Setting the Link Type of a Port to P2P

Refer to Setting the Link Type of a Port to P2P in the section about root bridge configuration.

Configuring the Mode a Port Uses to Recognize/Send MSTP Packets

Refer to Configuring the Mode a Port Uses to Recognize/Send MSTP Packets in the section about root bridge configuration.

Enabling Output of Port State Transition Information

Refer to Enabling the Output of Port State Transition Information in the section about root bridge configuration.

Enabling the MSTP Feature

Refer to Enabling the MSTP Feature in the section about root bridge configuration.

Performing mCheck

Ports on an MSTP-enabled device have three working modes: STP compatible mode, RSTP mode, and MSTP mode.

In a switched network, if a port on the device running MSTP (or RSTP) connects to a device running STP, this port will automatically migrate to the STP-compatible mode. However, if the device running STP is removed, the port on the MSTP (or RSTP) device will not be able to migrate automatically to the MSTP (or RSTP) mode, but will remain working in the STP-compatible mode. In this case, you can perform an mCheck operation to force the port to migrate to the MSTP (or RSTP) mode.

You can perform mCheck on a port through two approaches, which lead to the same result.

Configuration Prerequisites

- MSTP has been correctly configured on the device.
- MSTP is configured to operate in MSTP mode or RSTP-compatible mode.

Configuration Procedure

Performing mCheck globally

Follow these steps to perform global mCheck:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Perform mCheck	**stp mcheck**	Required

Performing mCheck in interface view

Follow these steps to perform mCheck in interface view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	—
Perform mCheck	**stp mcheck**	Required

Configuration Example

Perform mCheck on port GigabitEthernet1/0/1.

1) Method 1: Perform mCheck globally.

```
<Sysname> system-view
[Sysname] stp mcheck
```

2) Method 2: Perform mCheck in interface view.

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp mcheck
```

Configuring Digest Snooping

As defined in IEEE 802.1s, interconnected devices are in the same region only when the region-related configuration (domain name, revision level, VLAN-to-MSTI mappings) on them is identical. An MSTP enabled device identifies devices in the same MST region via checking the configuration ID in BPDU packets. The configuration ID includes the region name, revision level, configuration digest that is in 16-byte length and is the result calculated via the HMAC-MD5 algorithm based on VLAN-to-MSTI mappings.

Since MSTP implementations differ with vendors, the configuration digests calculated using private keys is different; hence different vendors' devices in the same MST region can not communicate with each other.

Enabling the Digest Snooping feature on the port connecting the local device to another vendor's device in the same MST region can make the two devices communicate with each other.

Configuration Prerequisites

Associated devices of different vendors are interconnected and run MSTP.

Configuration Procedure

Follow these steps to configure Digest Snooping:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable digest snooping on the interface or port group		**stp config-digest-snooping**	Required Not enabled by default
Return to system view		**quit**	—
Enable global digest snooping		**stp config-digest-snooping**	Required Not enabled by default

Caution

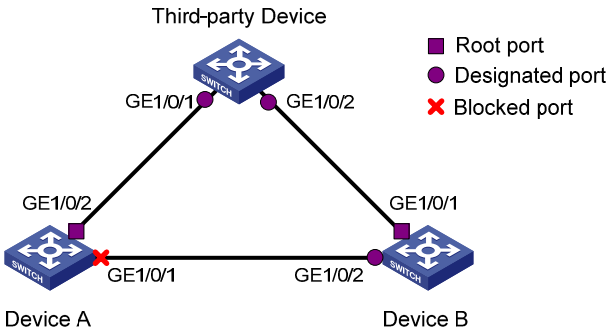
- You can enable Digest Snooping on only a device that is connected to another vendor's device that uses its private key to calculate the configuration digest.
- With the Digest Snooping feature enabled, comparison of configuration digest is not needed for in-the-same-region check, so the VLAN-to-MSTI mappings must be the same on associated ports.
- With global Digest Snooping enabled, modification of VLAN-to-MSTI mappings and removing of the current region configuration using the **undo stp region-configuration** command are not allowed. You can only modify the region name and revision level.
- You need to enable this feature both globally and on associated ports to make it take effect. It is recommended to enable the feature on all associated ports first and then globally, making all configured ports take effect, and disable the feature globally to disable it on all associated ports.
- It is not recommended to enable Digest Snooping on MST region edge ports to avoid loops.
- It is recommended to enable Digest Snooping first and then MSTP. Do not configure Digest Snooping when the network works well to avoid traffic interruption.

Configuration Example

Network requirements

- Device A and Device B connect to a third-party's device and all the devices are in the same region.
- Enable Digest Snooping on Device A and Device B so that the three routers can communicate with one another.

Figure 1-6 Digest Snooping configuration



Configuration procedure

1) Enable Digest Snooping on Device A.

Enable Digest Snooping on GigabitEthernet1/0/1.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] stp config-digest-snooping
[DeviceA-GigabitEthernet1/0/1] quit
```

Enable global Digest Snooping.

```
[DeviceA] stp config-digest-snooping
```

2) Enable Digest Snooping on Device B (the same as above, omitted)

Configuring No Agreement Check

In RSTP and MSTP, two types of messages are used for rapid state transition on designated ports:

- Proposal: sent by designated ports to request rapid transition
- Agreement: used to acknowledge rapid transition requests

Both RSTP and MSTP devices can perform rapid transition on a designated port only when the port receives an agreement packet from the downstream device. The differences between RSTP and MSTP devices are:

- For MSTP, the downstream device's root port sends an agreement packet only after it receives an agreement packet from the upstream device.
- For RSTP, the down stream device sends an agreement packet regardless of whether an agreement packet from the upstream device is received.

Figure 1-7 shows the rapid state transition mechanism on MSTP designated ports.

Figure 1-7 Rapid state transition of an MSTP designated port

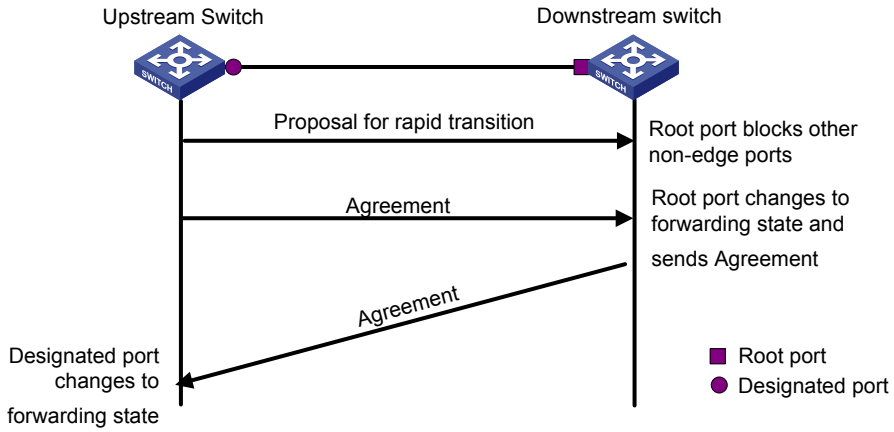
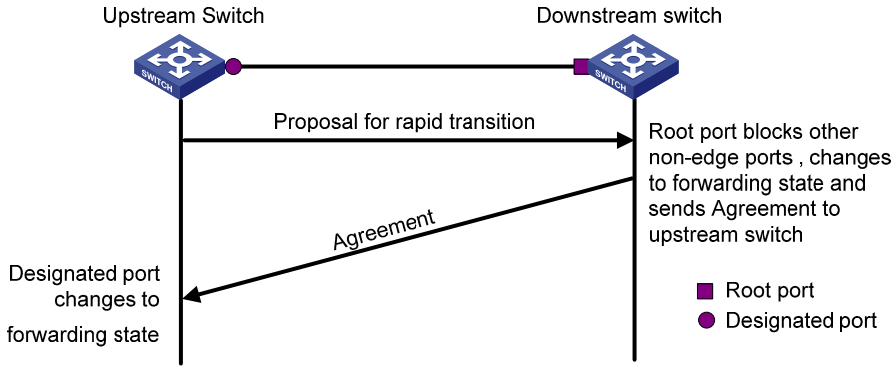


Figure 1-8 shows rapid state transition of an RSTP designated port.

Figure 1-8 Rapid state transition of an RSTP designated port



If the upstream device comes from another vendor, the rapid state transition implementation may be limited. For example, when the upstream device uses a rapid transition mechanism similar to that of RSTP, and the downstream device adopts MSTP and does not work in RSTP mode, the root port on the downstream device receives no agreement packet from the upstream device and thus sends no agreement packets to the upstream device. As a result, the designated port of the upstream device fails to transit rapidly and can only change to the forwarding state after a period twice the Forward Delay.

In this case, you can enable the No Agreement Check feature on the downstream device's port to enable the designated port of the upstream device to transit its state rapidly.

Configuration Prerequisites

- A device is connected to an upstream device supporting MSTP via a point-to-point link.
- Configure the same region name, revision level and VLAN-to-MSTI mappings on the two devices, thus assigning them to the same region.

Configuration Procedure

Follow these steps to configure No Agreement Check:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable No Agreement Check		**stp no-agreement-check**	Required Not enabled by default

Note

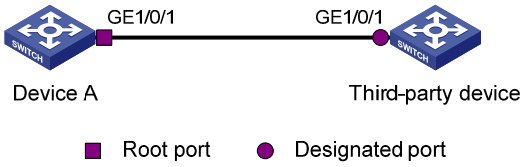
To make the No Agreement Check feature take effect, enable it on the root port.

Configuration Example

Network requirements

- Device A connects to a third-party's device that has different MSTP implementation. Both devices are in the same region.
- Another vendor's device is the regional root bridge, and Device A is the downstream device.

Figure 1-9 No Agreement Check configuration



Configuration procedure

Enable No Agreement Check on GigabitEthernet 1/0/1 of Device A.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] stp no-agreement-check
```

Configuring Protection Functions

An MSTP-enabled device supports the following protection functions:

- BPDU guard
- Root guard

- Loop guard
- TC-BPDU attack guard

 **Note**

Among loop guard, root guard and edge port settings, only one function can take effect on the same port at the same time.

Configuration prerequisites

MSTP has been correctly configured on the device.

Enabling BPDU Guard

 **Note**

We recommend that you enable BPDU guard if your device supports this function.

For access layer devices, the access ports generally connect directly with user terminals (such as PCs) or file servers. In this case, the access ports are configured as edge ports to allow rapid transition. When these ports receive configuration BPDUs, the system will automatically set these ports as non-edge ports and start a new spanning tree calculation process. This will cause a change of network topology. Under normal conditions, these ports should not receive configuration BPDUs. However, if someone forges configuration BPDUs maliciously to attack the devices, network instability will occur.

MSTP provides the BPDU guard function to protect the system against such attacks. With the BPDU guard function enabled on the devices, when edge ports receive configuration BPDUs, MSTP will close these ports and notify the NMS that these ports have been closed by MSTP. Those ports closed thereby can be restored only by the network administers.

Follow these steps to enable BPDU guard:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable the BPDU guard function for the device	**stp bpdu-protection**	Required Disabled by default

---

 **Note**

BPDU Guard does not take effect on loopback test-enabled ports. For information about loopback test, refer to *Ethernet Interface Configuration* in the *Access Volume*.

Enabling Root Guard

Note

We recommend that you enable root guard if your device supports this function.

The root bridge and secondary root bridge of a panning tree should be located in the same MST region. Especially for the CIST, the root bridge and secondary root bridge are generally put in a high-bandwidth core region during network design. However, due to possible configuration errors or malicious attacks in the network, the legal root bridge may receive a configuration BPDU with a higher priority. In this case, the current legal root bridge will be superseded by another device, causing an undesired change of the network topology. As a result, the traffic that should go over high-speed links is switched to low-speed links, resulting in network congestion.

To prevent this situation from happening, MSTP provides the root guard function to protect the root bridge. If the root guard function is enabled on a port of a root bridge, this port will keep playing the role of designated port on all MSTIs. Once this port receives a configuration BPDU with a higher priority from an MSTI, it immediately sets that port to the listening state in the MSTI, without forwarding the packet (this is equivalent to disconnecting the link connected with this port in the MSTI). If the port receives no BPDUs with a higher priority within twice the forwarding delay, the port will revert to its original state.

Follow these steps to enable root guard:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable the root guard function for the port(s)		**stp root-protection**	Required Disabled by default

Enabling Loop Guard

Note

We recommend that you enable loop guard if your device supports this function.

By keeping receiving BPDUs from the upstream device, a device can maintain the state of the root port and blocked ports. However, due to link congestion or unidirectional link failures, these ports may fail to receive BPDUs from the upstream devices. In this case, the downstream device will reselect the port roles: those ports in forwarding state that failed to receive upstream BPDUs will become designated ports, and the blocked ports will transition to the forwarding state, resulting in loops in the switched network. The loop guard function can suppress the occurrence of such loops.

If a loop guard–enabled port fails to receive BPDUs from the upstream device, and if the port took part in STP calculation, all the instances on the port, no matter what roles the port plays, will be set to, and stay in, the Discarding state.

Follow these steps to enable loop guard:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter Ethernet interface view or Layer-2 aggregate interface view	**interface** *interface-type interface-number*	Required Use either command. Configurations made in interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable the loop guard function for the port(s)		**stp loop-protection**	Required Disabled by default

Enabling TC-BPDU Attack Guard

When receiving a TC-BPDU (a BPDU used as notification of a topology change), the device will refresh the forwarding address entries. If someone forges TC-BPDUs to attack the device, the device will receive a larger number of TC-BPDUs within a short time, and frequent refresh operations bring a big burden to the device and hazard network stability.

With the TC-BPDU guard function enabled, the device limits the maximum number of times of immediately refreshing forwarding address entries within 10 seconds after it receives the first TC-BPDUs to the value set with the **stp tc-protection threshold** command (assume the value is X). At the same time, the system monitors whether the number of TC-BPDUs received within that period of time is larger than X. If so, the device will perform another refresh operation after that period of time elapses. This prevents frequent refreshing of forwarding address entries.

Follow these steps to enable TC-BPDU attack guard:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable the TC-BPDU attack guard function	**stp tc-protection enable**	Optional Enabled by default
Configure the maximum number of times the device refreshes forwarding address entries within a certain period of time immediately after it receives the first TC-BPDU	**stp tc-protection threshold** *number*	Optional 6 by default

Note

We recommend that you keep this feature enabled.

Displaying and Maintaining MSTP

To do...	Use the command...	Remarks
View information about abnormally blocked ports	**display stp abnormal-port**	Available in any view
View information about ports blocked by STP protection functions	**display stp down-port**	Available in any view
View the information of port role calculation history for the specified MSTI or all MSTIs	**display stp** [**instance** *instance-id*] **history** [**slot** *slot-number*]	Available in any view
View the statistics of TC/TCN BPDUs sent and received by all ports in the specified MSTI or all MSTIs	**display stp** [**instance** *instance-id*] **tc** [**slot** *slot-number*]	Available in any view
View the status information and statistics information of MSTP	**display stp** [**instance** *instance-id*] [**interface** *interface-list* | **slot** *slot-number*] [**brief**]	Available in any view
View the MST region configuration information that has taken effect	**display stp region-configuration**	Available in any view
View the root bridge information of all MSTIs	**display stp root**	Available in any view
Clear the statistics information of MSTP	**reset stp** [**interface** *interface-list*]	Available in user view

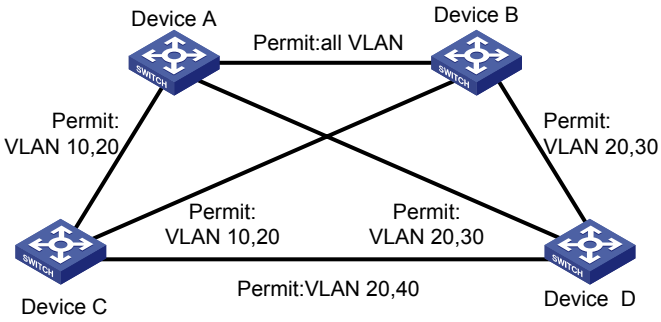
MSTP Configuration Example

Network requirements

Configure MSTP so that packets of different VLANs are forwarded along different spanning trees. The specific configuration requirements are as follows:

- All devices on the network are in the same MST region.
- Packets of VLAN 10 are forwarded along MSTI 1, those of VLAN 30 are forwarded along MSTI 3, those of VLAN 40 are forwarded along MSTI 4, and those of VLAN 20 are forwarded along MSTI 0.
- Device A and Device B are distribution layer devices, while Device C and Device D are access layer devices. VLAN 10 and VLAN 30 are terminated on the distribution layer devices, and VLAN 40 is terminated on the access layer devices, so the root bridges of MSTI 1 and MSTI 3 are Device A and Device B respectively, while the root bridge of MSTI 4 is Device C.

Figure 1-10 Network diagram for MSTP configuration



Note

"Permit:" beside each link in the figure is followed by the VLANs the packets of which are permitted to pass this link.

Configuration procedure

1) Configuration on Device A

Enter MST region view.

```
<DeviceA> system-view
[DeviceA] stp region-configuration
```

Configure the region name, VLAN-to-MSTI mappings and revision level of the MST region.

```
[DeviceA-mst-region] region-name example
[DeviceA-mst-region] instance 1 vlan 10
[DeviceA-mst-region] instance 3 vlan 30
[DeviceA-mst-region] instance 4 vlan 40
[DeviceA-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceA-mst-region] active region-configuration
[DeviceA-mst-region] quit
```

Define Device A as the root bridge of MSTI 1.

```
[DeviceA] stp instance 1 root primary
```

Enable MSTP globally.

```
[DeviceA] stp enable
```

View the MST region configuration information that has taken effect.

```
[DeviceA] display stp region-configuration
 Oper configuration
   Format selector    :0
   Region name        :example
   Revision level     :0

   Instance   Vlans Mapped
```

```
      0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
      1       10
      3       30
      4       40
```

2) Configuration on Device B

Enter MST region view.

```
<DeviceB> system-view
[DeviceB] stp region-configuration
```

Configure the region name, VLAN-to-MSTI mappings and revision level of the MST region.

```
[DeviceB-mst-region] region-name example
[DeviceB-mst-region] instance 1 vlan 10
[DeviceB-mst-region] instance 3 vlan 30
[DeviceB-mst-region] instance 4 vlan 40
[DeviceB-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Define Device B as the root bridge of MSTI 3.

```
[DeviceB] stp instance 3 root primary
```

Enable MSTP globally.

```
[DeviceB] stp enable
```

View the MST region configuration information that has taken effect.

```
[DeviceB] display stp region-configuration
 Oper configuration
   Format selector    :0
   Region name        :example
   Revision level     :0

   Instance   Vlans Mapped
      0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
      1       10
      3       30
      4       40
```

3) Configuration on Device C.

Enter MST region view.

```
<DeviceC> system-view
[DeviceC] stp region-configuration
[DeviceC-mst-region] region-name example
```

Configure the region name, VLAN-to-MSTI mappings and revision level of the MST region.

```
[DeviceC-mst-region] instance 1 vlan 10
[DeviceC-mst-region] instance 3 vlan 30
[DeviceC-mst-region] instance 4 vlan 40
[DeviceC-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Define Device C as the root bridge of MSTI 4.

```
[DeviceC] stp instance 4 root primary
```

Enable MSTP globally.

```
[DeviceC] stp enable
```

View the MST region configuration information that has taken effect.

```
[DeviceC] display stp region-configuration
 Oper configuration
   Format selector    :0
   Region name        :example
   Revision level     :0

   Instance   Vlans Mapped
      0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
      1       10
      3       30
      4       40
```

4) Configuration on Device D.

Enter MST region view.

```
<DeviceD> system-view
[DeviceD] stp region-configuration
[DeviceD-mst-region] region-name example
```

Configure the region name, VLAN-to-MSTI mappings and revision level of the MST region.

```
[DeviceD-mst-region] instance 1 vlan 10
[DeviceD-mst-region] instance 3 vlan 30
[DeviceD-mst-region] instance 4 vlan 40
[DeviceD-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Enable MSTP globally.

```
[DeviceD] stp enable
```

View the MST region configuration information that has taken effect.

```
[DeviceD] display stp region-configuration
 Oper configuration
   Format selector    :0
   Region name        :example
   Revision level     :0

   Instance   Vlans Mapped
      0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
      1       10
      3       30
```

```
4       40
```

Table of Contents

1 RRPP Configuration

When configuring RRPP, go to these sections for information you are interested in:

RRPP Overview

The Rapid Ring Protection Protocol (RRPP) is a link layer protocol designed for Ethernet rings. RRPP can prevent broadcast storms caused by data loops when an Ethernet ring is healthy, and rapidly restore the communication paths between the nodes in the event that a link is disconnected on the ring.

Compared with the IEEE spanning tree protocols, RRPP features the following:

- Fast topology convergence
- Convergence time independent of Ethernet ring size

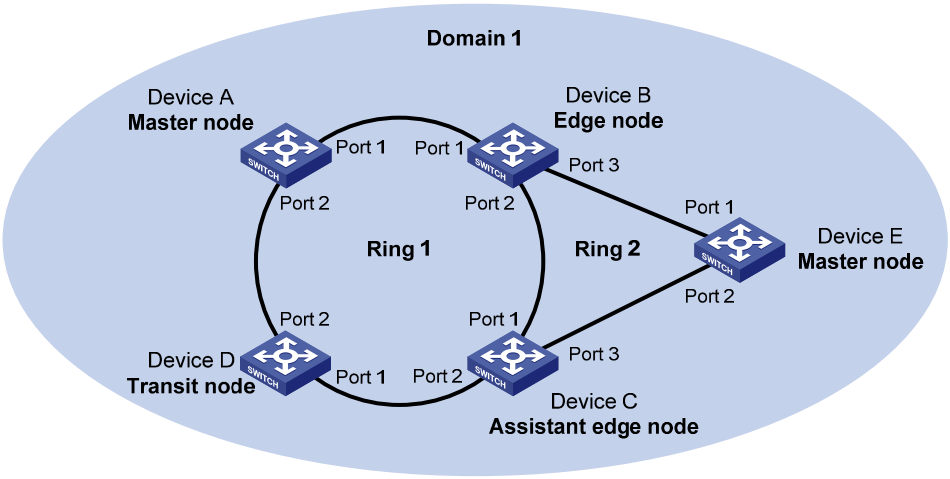
Background

Metropolitan area networks (MANs) and enterprise networks usually use the ring structure to improve reliability. However, services will be interrupted if any node in the ring network fails. A ring network usually uses Resilient Packet Ring (RPR) or Ethernet rings. RPR is high in cost as it needs dedicated hardware. Contrarily, the Ethernet ring technology is more mature and economical, so it is more and more widely used in MANs and enterprise networks.

Currently, both Spanning Tree Protocol (STP) and RRPP can be used to eliminate Layer-2 loops. STP is mature; however, it takes several seconds to converge. RRPP is an Ethernet ring-specific data link layer protocol, and converges faster than STP. Additionally, the convergence time of RRPP is independent of the number of nodes in the Ethernet ring, and therefore, RRPP can be applied to large-diameter networks.

Basic Concepts in RRPP

Figure 1-1 RRPP networking diagram



RRPP domain

The interconnected devices with the same domain ID and control VLANs constitute an RRPP domain. An RRPP domain contains the following elements: primary ring, subring, control VLAN, master node, transit node, primary port, secondary port, common port, and edge port.

As shown in Figure 1-1, Domain 1 is an RRPP domain, including two RRPP rings: Ring 1 and Ring 2. All the nodes on the two RRPP rings belong to the RRPP domain.

RRPP ring

A ring-shaped Ethernet topology is called an RRPP ring. RRPP rings fall into two types: primary ring and subring. You can configure a ring as either the primary ring or a subring by specifying its ring level. The primary ring is of level 0, while a subring is of level 1. An RRPP domain contains multiple RRPP rings, one serving as the primary ring and the others serving as subrings.

As shown in Figure 1-1, Domain 1 contains two RRPP rings: Ring 1 and Ring 2. The level of Ring 1 is set to 0, that is, Ring 1 is configured as the primary ring; the level of Ring 2 is set to 1, that is, Ring 2 is configured as a subring.

A ring can be in one of the following two states:

- Health state: All the physical links on the Ethernet ring are connected.
- Disconnect state: Some physical links on the Ethernet ring are broken.

Control VLAN and data VLAN

In an RRPP domain, a control VLAN is a VLAN dedicated to transferring RRPP packets.

On a device, the ports accessing an RRPP ring belong to the control VLANs of the ring, and only such ports can join the control VLANs.

An RRPP domain is configured with two control VLANs: one primary control VLAN, which is the control VLAN for the primary ring; one secondary control VLAN, which is the control VLAN for subrings. All subrings in the same RRPP domain share the same secondary control VLAN. After you specify a VLAN as the primary control VLAN, the system automatically configures the VLAN whose ID is the primary control VLAN ID plus one as the secondary control VLAN.

IP address configuration is prohibited on the control VLAN interfaces.

A data VLAN is a VLAN dedicated to transferring data packets. Both RRPP ports and non-RRPP ports can be assigned to a data VLAN.

Node

Each device on an RRPP ring is referred to as a node. The role of a node is configurable. There are the following node roles:

- Master node: Each ring has one and only one master node. The master node initiates the polling mechanism and determines the operations to be performed after a change in topology.
- Transit node: Transit nodes include all the nodes except the master node on the primary ring and all the nodes on subrings except the master nodes and the nodes where the primary ring intersects with the subrings. A transit node monitors the state of its directly-connected RRPP links and notifies the master node of the link state changes, if any. Based on the link state changes, the master node decides the operations to be performed.
- Edge node: A node residing on both the primary ring and a subring at the same time. An edge node is a special transit node that serves as a transit node on the primary ring and an edge node on the subring.
- Assistant-edge node: A node residing on both the primary ring and a subring at the same time. An assistant-edge node is a special transit node that serves as a transit node on the primary ring and an assistant-edge node on the subring. This node works in conjunction with the edge node to detect the integrity of the primary ring and perform loop guard.

As shown in Figure 1-1, Ring 1 is the primary ring and Ring 2 is a subring. Device A is the master node of Ring 1, Device B, Device C and Device D are the transit nodes of Ring 1. Device E is the master node of Ring 2, Device B is the edge node of Ring 2, and Device C is the assistant-edge node of Ring 2.

Primary port and secondary port

Each master node or transit node has two ports connected to an RRPP ring, one serving as the primary port and the other serving as the secondary port. You can determine the role of a port.

1) In terms of functionality, the difference between the primary port and the secondary port of a master node is:

- The primary port and the secondary port are designed to play the role of sending and receiving loop-detect packets respectively.
- When an RRPP ring is in Health state, the secondary port of the master node will logically deny data VLANs and permit only the packets of the control VLANs.
- When an RRPP ring is in Disconnect state, the secondary port of the master node will permit data VLANs, that is, forward packets of data VLANs.

2) In terms of functionality, there is no difference between the primary port and the secondary port of a transit node. Both are designed for transferring protocol packets and data packets over an RRPP ring.

As shown in Figure 1-1, Device A is the master node of Ring 1. Port 1 and Port 2 are the primary port and the secondary port of the master node on Ring 1 respectively. Device B, Device C, and Device D are the transit nodes of Ring 1. Their Port 1 and Port 2 are the primary port and the secondary port on Ring 1 respectively.

Common port and edge port

The ports connecting the edge node and assistant-edge node to the primary ring are common ports. The ports connecting the edge node and assistant-edge node only to the subrings are edge ports.

As shown in Figure 1-1, Device B and Device C lie on Ring 1 and Ring 2. Device B's Port 1 and Port 2 and Device C's Port 1 and Port 2 access the primary ring, so they are common ports. Device B's Port 3 and Device C's Port 3 access only the subring, so they are edge ports.

RRPP ring group

To reduce Edge-Hello traffic, you can configure a group of subrings on the edge node or assistant-edge node. For information about Edge-Hello packets, refer to RRPP Packets. You must configure a device as the edge node of these subrings, and another device as the assistant-edge node of these subrings. Additionally, the subrings of the edge node and assistant-edge node must connect to the same subring packet tunnels in major ring (SRPTs), so that Edge-Hello packets of the edge node of these subrings travel to the assistant-edge node of these subrings over the same link.

A ring group configured on the edge node is called an edge node ring group, and a ring group configured on an assistant-edge node is called an assistant-edge node ring group. Up to one subring in an edge node ring group is allowed to send Edge-Hello packets.

RRPP Packets

Table 1-1 shows the types of RRPP packets and their functions.

Table 1-1 RRPP packet types and their functions

Type	Description
Hello	The master node initiates Hello packets to detect the integrity of a ring in a network.
Link-Down	The transit node, the edge node or the assistant-edge node initiates Link-Down packets to notify the master node of the disappearance of a ring in case of a link failure.
Common-Flush-FDB	The master node initiates Common-Flush-FDB packets to instruct the transit nodes to update their own MAC entries and ARP/ND entries when an RRPP ring transits to Disconnect state.
Complete-Flush-FDB	The master node initiates Complete-Flush-FDB packets to instruct the transit nodes to update their own MAC entries and ARP/ND entries, and release blocked ports from being blocked temporarily when an RRPP ring transits to Health state.
Edge-Hello	The edge node initiates Edge-Hello packets to examine the links of the primary ring between the edge node and the assistant-edge node.
Major-Fault	The assistant-edge node initiates Major-Fault packets to notify the edge node of a failure when a link of primary ring between edge node and assistant-edge node is torn down.

Hello and Fail Timers

When RRPP checks the link state of an Ethernet ring, the master node sends Hello packets out the primary port according to the Hello timer and determines whether its secondary port receives the Hello packets based on the Fail timer.

- The Hello timer specifies the interval at which the master node sends Hello packets out the primary port.
- The Fail timer specifies the maximum delay between the master node sending Hello packets out the primary port and the secondary port receiving the Hello packets from the primary port. If the

secondary port receives the Hello packets sent by the local master node before the Fail timer expires, the overall ring is in Health state. Otherwise, the ring transits into Disconnect state.

Note

- In an RRPP domain, a transit node learns the Hello timer value and the Fail timer value on the master node through the received Hello packets, ensuring that all nodes in the ring network are consistent in the two timer settings.
- The Fail timer value must be greater than or equal to three times of the Hello timer value.
- In a dual-homed-ring network, to avoid temporary loops when the primary ring fails, ensure that the difference between the Fail timer value on the master node of the subring and that on the master node of the primary ring is greater than twice the Hello timer value of the master node of the subring.

How RRPP Works

Polling mechanism

The polling mechanism is used by the master node of an RRPP ring to check the Health state of the ring network.

The master node sends Hello packets out its primary port periodically, and these Hello packets travel through each transit node on the ring in turn.

- If the ring is complete, the secondary port of the master node will receive Hello packets before the Fail timer expires and the master node will keep the secondary port blocked.
- If the ring is torn down, the secondary port of the master node will fail to receive Hello packets before the Fail timer expires. The master node will release the secondary port from blocking data VLANs while sending Common-Flush-FDB packets to instruct all transit nodes to update their own MAC entries and ARP/ND entries.

Link down alarm mechanism

The transit node, the edge node or the assistant-edge node sends Link-Down packets to the master node immediately when they find any of its own ports belonging to an RRPP domain is down. Upon the receipt of a Link-Down packet, the master node releases the secondary port from blocking data VLANs while sending Common-Flush-FDB packet to instruct all the transit nodes, the edge nodes and the assistant-edge nodes to update their own MAC entries and ARP/ND entries. After each node updates its own entries, traffic is switched to the normal link.

Ring recovery

The master node may find the ring is restored after a period of time after the ports belonging to the RRPP domain on the transit nodes, the edge nodes, or the assistant-edge nodes are brought up again. A temporary loop may arise in the data VLAN during this period. As a result, broadcast storm occurs.

To prevent temporary loops, non-master nodes block them immediately (and permit only the packets of the control VLAN to pass through) when they find their ports accessing the ring are brought up again. The blocked ports are activated only when the nodes are sure that no loop will be brought forth by these ports.

Broadcast storm suppression mechanism in a multi-homed subring in case of SRPT failure

As shown in Figure 1-5, Ring 1 is the primary ring, and Ring 2 and Ring 3 are subrings. When the two SRPTs between the edge node and the assistant-edge node are down, the master nodes of Ring 2 and Ring 3 will open their respective secondary ports, and thus a loop among Device B, Device C, Device E, and Device F is generated. As a result, broadcast storm occurs.

In this case, to prevent generating this loop, the edge node will block the edge port temporarily. The blocked edge port is activated only when the edge node is sure that no loop will be brought forth when the edge port is activated.

Load balancing

In a ring network, maybe traffic of multiple VLANs is transmitted at the same time. RRPP can implement load balancing for the traffic by transmitting traffic of different VLANs along different paths.

By configuring an individual RRPP domain for transmitting the traffic of the specified VLANs (referred to as protected VLANs) in a ring network, traffic of different VLANs can be transmitted according to different topologies in the ring network. In this way, load balancing is achieved.

As shown in Figure 1-6, Ring 1 is configured as the primary ring of Domain 1 and Domain 2, which are configured with different protected VLANs. Device A is the master node of Ring 1 in Domain 1; Device B is the master node of Ring 1 in Domain 2. With such configurations, traffic of different VLANs can be transmitted on different links, and thus, load balancing is achieved in a single-ring network.

RRPP ring group

In an edge node ring group, only an activated subring with the lowest domain ID and ring ID can send Edge-Hello packets. In an assistant-edge node ring group, any activated subring that has received Edge-Hello packets will forward these packets to the other activated subrings. With an edge node ring group and an assistant-edge node group configured, only one subring sends and receives Edge-Hello packets, thus reducing CPU workload.

As shown in Figure 1-5, Device B is the edge node of Ring 2 and Ring 3, and Device C is the assistant-edge node of Ring 2 and Ring 3. Device B and Device C need to send or receive Edge-Hello packets frequently. If more subrings are configured or load balancing is configured for more multiple domains, Device B and Device C will send or receive a mass of Edge-Hello packets.

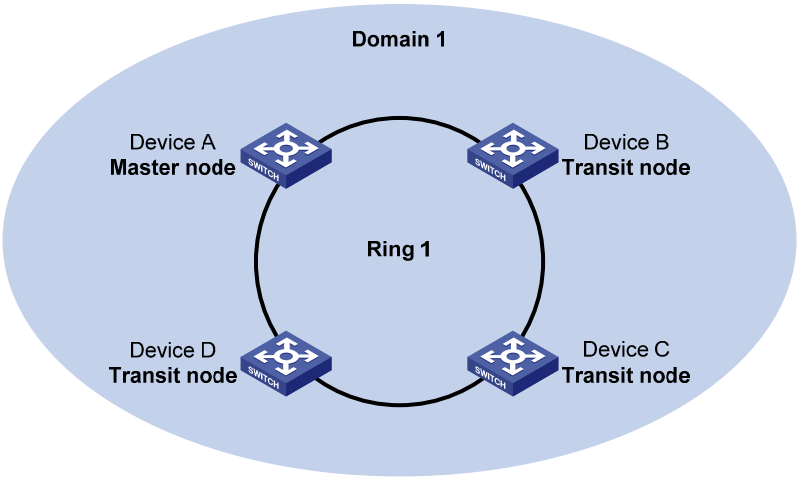
To reduce Edge-Hello traffic, you can assign Ring 2 and Ring 3 to a ring group configured on the edge node Device B, and assign Ring 2 and Ring 3 to a ring group configured on Device C. After such configurations, if all rings are activated, only Ring 2 on Device B sends Edge-Hello packets.

Typical RRPP Networking

Here are several typical networking applications.

Single ring

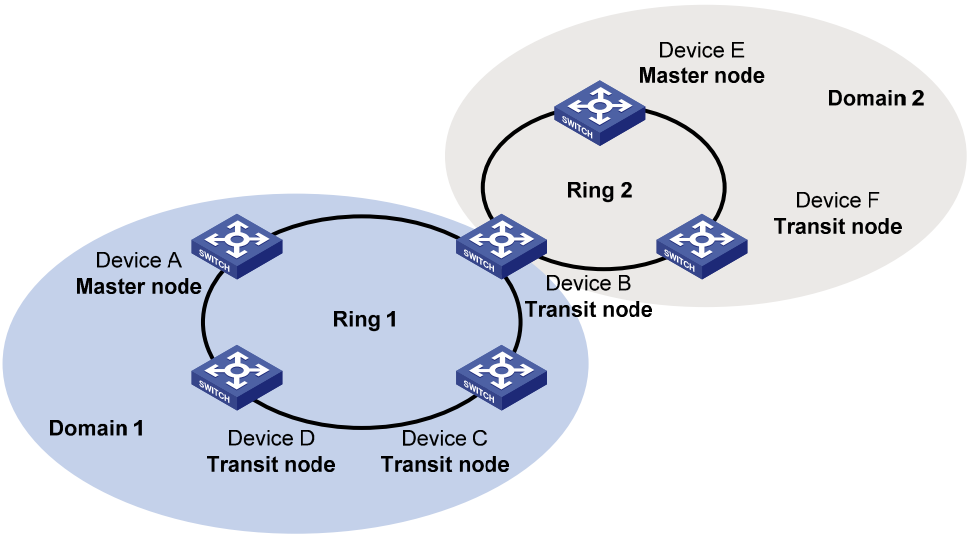
Figure 1-2 Single ring



There is only a single ring in the network topology. In this case, you only need to define an RRPP domain.

Tangent rings

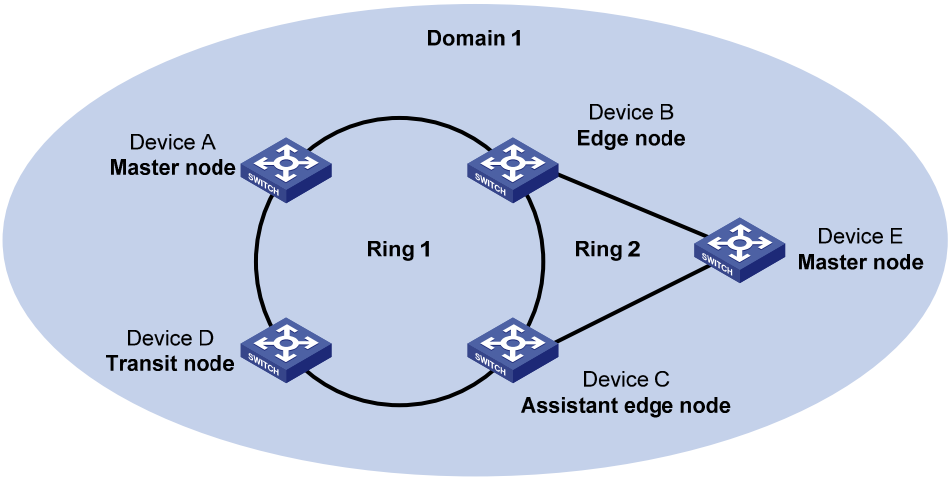
Figure 1-3 Tangent rings



There are two or more rings in the network topology and only one common node between rings. In this case, you need to define an RRPP domain for each ring.

Intersecting rings

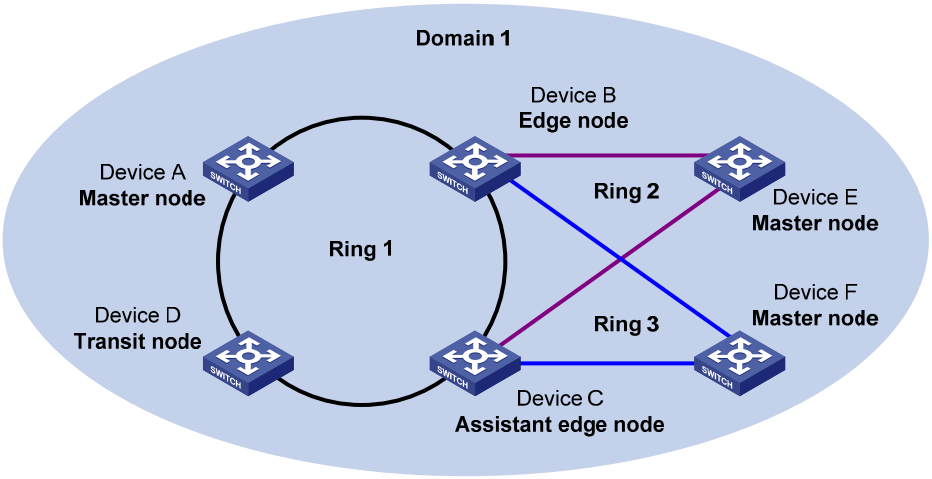
Figure 1-4 Intersecting rings



There are two or more rings in the network topology and two common nodes between rings. In this case, you only need to define an RRPP domain, and set one ring as the primary ring and the other rings as subrings.

Dual homed rings

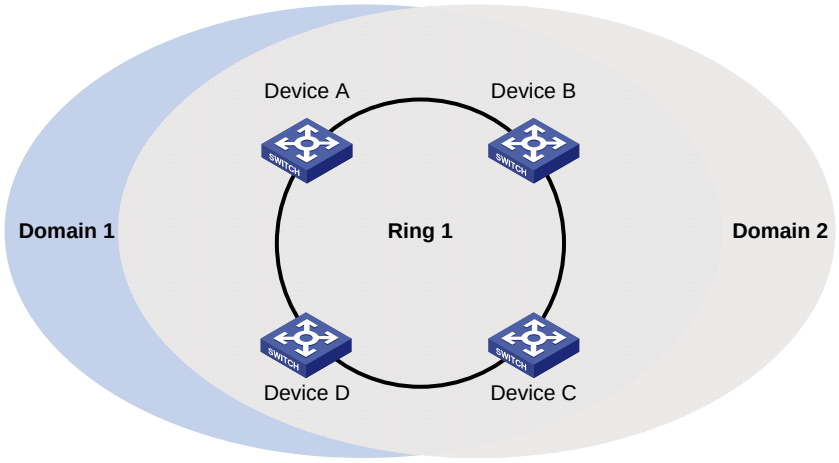
Figure 1-5 Dual homed rings



There are two or more rings in the network topology and two similar common nodes between rings. In this case, you only need to define an RRPP domain, and set one ring as the primary ring and the other rings as subrings.

Single-ring load balancing

Figure 1-6 Network diagram for single-ring load balancing

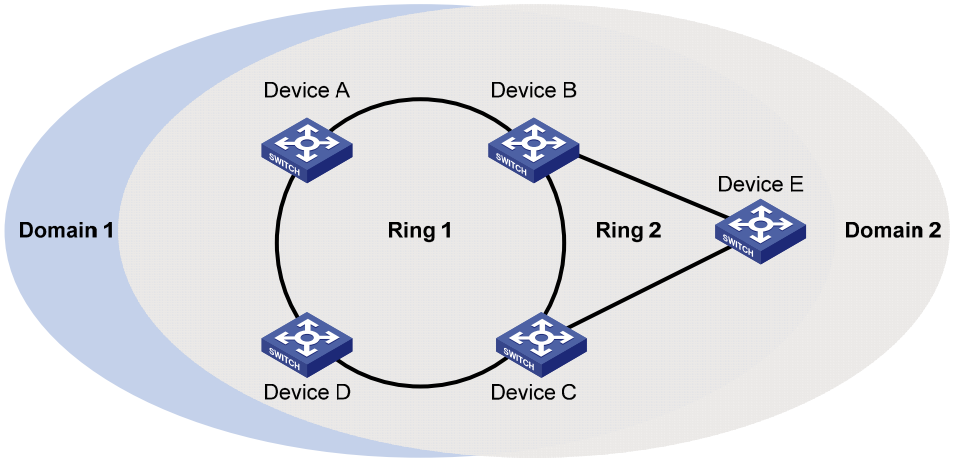


In a single-ring network, you can achieve load balancing by configuring multiple domains.

As shown in Figure 1-6, Ring 1 is configured as the primary ring of both Domain 1 and Domain 2. In Domain 1, Device A is configured as the master node of Ring 1; in Domain 2, Device B is configured as the master node of Ring 1. Such configurations enable the ring to block different links based on VLANs, thus achieving single-ring load balancing.

Intersecting-ring load balancing

Figure 1-7 Network diagram for intersecting-ring load balancing



In an intersecting-ring network, you can also achieve load balancing by configuring multiple domains.

As shown in Figure 1-7, Ring 1 is the primary ring and Ring 2 is the subring in both Domain 1 and Domain 2. Domain 1 and Domain 2 are configured with different protected VLANs. Device A is configured as the master node of Ring 1 in Domain 1; Device D is configured as the master node of Ring 1 in Domain 2. Device E is configured as the master node of Ring 2 in both Domain 1 and Domain 2. However, different ports on Device E are blocked in Domain 1 and Domain 2. After such configurations, you can enable traffic of different VLANs to travel over different paths in the subring and primary ring, thus achieving intersecting-ring load balancing.

Protocols and Standards

RFC 3619 *Extreme Networks' Ethernet Automatic Protection Switching (EAPS) Version 1* is related to RRPP.

RRPP Configuration Task List

Caution

- RRPP does not have an auto election mechanism, so you must configure each node in the ring network properly for RRPP to monitor and protect the ring network.
- Before configuring RRPP, you need to construct a ring-shaped Ethernet topology physically.

You can create RRPP domains based on service planning, specify control VLANs and data VLANs for each RRPP domain, and then determine the ring roles and node roles based on the traffic paths in each RRPP domain. You can configure devices through the following configurations.

Complete the following tasks to configure RRPP:

Task	Description
Configuring Master Node	Required
Configuring Transit Node	Optional
Configuring Edge Node	Optional
Configuring Assistant Edge Node	Optional
Configuring Ring Group	Optional To reduce Edge-Hello traffic, you can adopt the ring group mechanism, that is, assign subrings with the same edge node/assistant-edge node to a ring group.

Caution

- It is recommended to configure the primary ring first and then the subring when you configure an RRPP domain. Moreover, a Ring ID cannot be applied to more than one RRPP ring in one RRPP domain.
- If a device lies on multiple RRPP rings in an RRPP domain, only one primary ring exists. The device serves as either an edge node or an assistant-edge node on the subrings.
- The total number of rings configured on a device in all RRPP domains cannot be greater than 16.
- Modification of node mode, port role and ring level of an RRPP ring is prohibited after configuration. If needed, you must first delete the existing configuration.
- During load balancing configuration, different protected VLANs must be configured for different domains.

Ports connected to an RRPP ring must meet the following conditions:

- The link type of these ports must be trunk.
- They must be Layer 2 GE ports.
- They must not be member ports of any aggregation group, service loopback group, or smart link group.
- STP is disabled on them.
- The 802.1p priority of trusted packets on the ports is configured, so that RRPP packets take higher precedence than data packets when passing through the ports.
- Do not enable OAM remote loopback function on an RRPP port. Otherwise, this may cause temporary broadcast storm.
- You are recommended not to configure physical-link-state change suppression time on a port accessing an RRPP ring to accelerate topology convergence. For details, refer to *Ethernet Interface Configuration* in the *Access Volume*.

Note

- If you need to transparently transmit RRPP packets on a device without enabling RRPP, you must ensure only the two ports accessing an RRPP ring permit the packets of the control VLAN. Otherwise, the packets from other VLANs may go into the control VLAN in transparent transmission mode and strike the RRPP ring. Meantime, you must configure the 802.1p priority for trusted packets on the two ports accessing the RRPP ring.
- Do not configure the default VLAN of a port accessing an RRPP ring as the primary control VLAN or the secondary control VLAN, ensuring proper receiving/sending of RRPP packets.
- Do not enable QinQ or VLAN mapping on the control VLAN. Otherwise, RRPPDUs cannot be forwarded properly.
- You can still assign ports to or remove ports from the aggregation group corresponding to a Layer 2 aggregate interface configured as an RRPP port.

Configuring Master Node

Follow these steps to configure master node:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an RRPP domain and enter its view	**rrpp domain** *domain-id*	Required
Specify control VLAN for the RRPP domain	**control-vlan** *vlan-id*	Required
Specify the protected VLANs for the RRPP domain	**protected-vlan reference-instance** *instance-id-list*	Required No protected VLAN is specified for an RRPP domain by default.

To do…	Use the command…	Remarks
Specify the current device as the master node of the ring, and specify the primary port and the secondary port	**ring** *ring-id* **node-mode master** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*	Required
Configure the timer for the RRPP domain	**timer hello-timer** *hello-value* **fail-timer** *fail-value*	Optional By default, the Hello timer value is 1 second and the Fail timer value is 3 seconds.
Enable the RRPP ring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Return to system view	**quit**	—
Enable RRPP	**rrpp enable**	Required By default, RRPP is disabled.

Caution

- Before specifying RRPP rings for an RRPP domain, you must specify protected VLANs for the domain.
- Before specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain; after specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain, however, you cannot delete all the protected VLANs configured for the domain.
- Deleting an RRPP domain deletes its protected VLANs at the same time.
- The **protected-vlan** command configures protected VLANs for an RRPP domain by referencing MSTIs to which the protected VLANs are mapped. You can use the **display stp region-configuration** command to view the VLAN-to-MSTI mappings. For detailed information about VLAN-to-MSTI mapping configuration, refer to *MSTP Configuration* in the *Access Volume*.
- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.
- Before removing or modifying the control VLAN of an RRPP domain, make sure that the RRPP domain is not configured with any RRPP ring.

Configuring Transit Node

Follow these steps to configure transit node:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an RRPP domain and enter its view	**rrpp domain** *domain-id*	Required

To do…	Use the command…	Remarks
Specify a control VLAN for the RRPP domain	**control-vlan** *vlan-id*	Required
Specify protected VLANs for the RRPP domain	**protected-vlan reference-instance** *instance-id-list*	Required No protected VLAN is specified for an RRPP domain by default.
Specify the current device as the transit node of the ring, and specify the primary port and the secondary port	**ring** *ring-id* **node-mode transit** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*	Required
Enable the RRPP ring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Return to system view	**quit**	—
Enable RRPP	**rrpp enable**	Required By default, RRPP is disabled.

Caution

- Before specifying RRPP rings for an RRPP domain, you must specify protected VLANs for the domain.
- Before specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain; after specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain, however, you cannot delete all the protected VLANs configured for the domain.
- Deleting an RRPP domain deletes its protected VLANs at the same time.
- The **protected-vlan** command configures protected VLANs for an RRPP domain by referencing MSTIs to which the protected VLANs are mapped. You can use the **display stp region-configuration** command to view the VLAN-to-MSTI mappings. For detailed information about VLAN-to-MSTI mapping configuration, refer to *MSTP Configuration* in the *Access Volume*.
- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.
- Before removing or modifying the control VLAN of an RRPP domain, make sure that the RRPP domain is not configured with any RRPP ring.

Configuring Edge Node

Follow these steps to configure edge node:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an RRPP domain and enter its view	**rrpp domain** *domain-id*	Required
Specify a control VLAN for the RRPP domain	**control-vlan** *vlan-id*	Required
Specify protected VLANs for the RRPP domain	**protected-vlan reference-instance** *instance-id-list*	Required No protected VLAN is specified for an RRPP domain by default.
Specify the current device as the transit node of the primary ring, and specify the primary port and the secondary port	**ring** *ring-id* **node-mode transit** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*	Required
Specify the current device as the edge node of a subring, and specify the edge port	**ring** *ring-id* **node-mode edge** [**edge-port** *interface-type interface-number*]	Required
Enable the primary ring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Enable the subring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Return to system view	**quit**	—
Enable RRPP	**rrpp enable**	Required By default, RRPP is disabled.

Caution

- Before specifying RRPP rings for an RRPP domain, you must specify protected VLANs for the domain.
- Before specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain; after specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain, however, you cannot delete all the protected VLANs configured for the domain.
- Deleting an RRPP domain deletes its protected VLANs at the same time.
- The **protected-vlan** command configures protected VLANs for an RRPP domain by referencing MSTIs to which the protected VLANs are mapped. You can use the **display stp region-configuration** command to view the VLAN-to-MSTI mappings. For detailed information about VLAN-to-MSTI mapping configuration, refer to *MSTP Configuration* in the *Access Volume*.
- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- A Ring ID cannot be applied to more than one RRPP ring in an RRPP domain.
- You must first configure the primary ring and then the subring when configuring an edge node. Moreover, you must remove all subring configurations before deleting the primary ring configuration of an edge node. However, the RRPP ring enabled cannot be deleted.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.
- Before removing or modifying the control VLAN of an RRPP domain, make sure that the RRPP domain is not configured with any RRPP ring.

Configuring Assistant Edge Node

Follow these steps to configure assistant-edge node:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an RRPP domain and enter its view	**rrpp domain** *domain-id*	Required
Specify a control VLAN for the RRPP domain	**control-vlan** *vlan-id*	Required
Specify protected VLANs for the RRPP domain	**protected-vlan reference-instance** *instance-id-list*	Required No protected VLAN is specified for an RRPP domain by default.
Specify the current device as the transit node of the primary ring, and specify the primary port and the secondary port	**ring** *ring-id* **node-mode transit** [**primary-port** *interface-type interface-number*] [**secondary-port** *interface-type interface-number*] **level** *level-value*	Required

To do…	Use the command…	Remarks
Specify the current device as the assistant-edge node of the subring, and specify an edge port	**ring** *ring-id* **node-mode assistant-edge** [**edge-port** *interface-type interface-number*]	Required
Enable the primary ring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Enable the subring	**ring** *ring-id* **enable**	Required By default, the RRPP ring is disabled.
Return to system view	**quit**	—
Enable RRPP	**rrpp enable**	Required By default, RRPP is disabled.

Caution

- Before specifying RRPP rings for an RRPP domain, you must specify protected VLANs for the domain.
- Before specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain; after specifying rings for an RRPP domain, you can delete or modify the protected VLANs configured for the RRPP domain, however, you cannot delete all the protected VLANs configured for the domain.
- Deleting an RRPP domain deletes its protected VLANs at the same time.
- The **protected-vlan** command configures protected VLANs for an RRPP domain by referencing MSTIs to which the protected VLANs are mapped. You can use the **display stp region-configuration** command to view the VLAN-to-MSTI mappings. For detailed information about VLAN-to-MSTI mapping configuration, refer to *MSTP Configuration* in the *Access Volume*.
- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- A Ring ID cannot be applied to more than one RRPP ring in an RRPP domain.
- You must first configure the primary ring and then the subring when configuring an edge node. Moreover, you must remove all subring configurations before deleting the primary ring configuration of an edge node. However, the RRPP ring enabled cannot be deleted.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.
- Before removing or modifying the control VLAN of an RRPP domain, make sure that the RRPP domain is not configured with any RRPP ring.

Configuring Ring Group

To reduce Edge-Hello traffic, you can adopt the ring group mechanism, that is, assign subrings with the same edge node/assistant-edge node to a ring group.

You need to configure ring groups on both the edge node and the assistant-edge node at the same time. The two ring groups must be configured with the same subrings. Otherwise, the ring groups cannot operate properly.

Configuration Prerequisites

- The RRPP domain, control VLANs, protected VLANs, the primary ring, and the subrings have been configured on the edge node device.
- The RRPP domain, control VLANs, protected VLANs, the primary ring, and the subrings have been configured on the assistant-edge node device.

Configuring Ring Group

Follow these steps to configure a ring group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a ring group and enter ring group view	**rrpp ring-group** *ring-group-id*	Required
Assign the specified subrings to the ring group	**domain** *domain-id* **ring** *ring-id-list*	Required

Note

- To add an activated ring to a ring group, first add the ring to the assistant-edge node ring group and then to the edge node ring group.
- To remove a ring from a ring group, first remove the ring from the edge node ring group and then from the assistant-edge node group.
- To remove a ring group, first remove the edge node ring group and then the assistant-edge node ring group.
- To activate the rings in a ring group, first activate the rings in the assistant-edge node ring group and then the rings in the edge node ring group.
- To deactivate the rings in a ring group, first deactivate the rings in the edge node ring group and then the rings in the assistant-edge node ring group.
- If you do not following the orders above, the assistant-edge node may take the primary ring as failed because the assistant-edge node cannot receive Edge-Hello packets.

Displaying and Maintaining RRPP

To do…	Use the command…	Remarks
Display brief information about RRPP configuration	**display rrpp brief**	Available in any view
Display detailed information about RRPP configuration	**display rrpp verbose domain** *domain-id* [**ring** *ring-id*]	
Display RRPP statistics	**display rrpp statistics domain** *domain-id* [**ring** *ring-id*]	

To do…	Use the command…	Remarks
Clear RRPP statistics	**reset rrpp statistics domain** *domain-id* [**ring** *ring-id*]	Available in user view

RRPP Typical Configuration Examples

Configuring Single Ring Topology

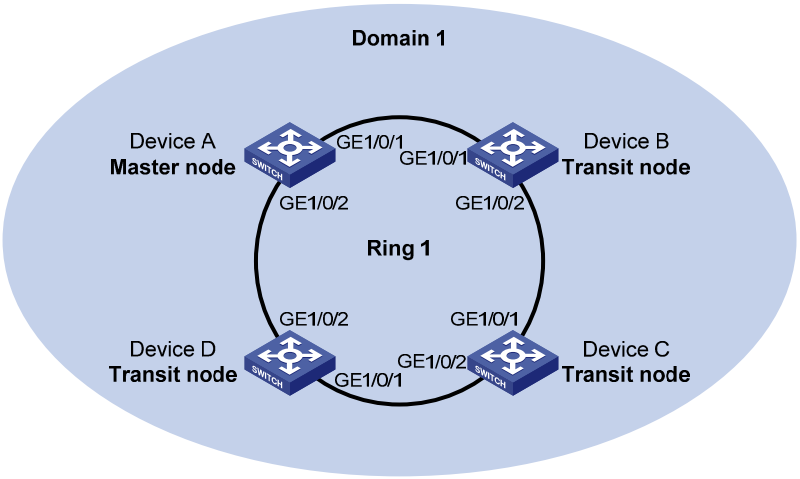
Networking requirements

- Device A, Device B, Device C, and Device D constitute RRPP domain 1, specify the primary control VLAN of RRPP domain 1 as VLAN 4092, and RPPP domain 1 protects all VLANs;
- Device A, Device B, Device C and Device D constitute primary ring 1;
- Specify Device A as the master node of primary ring 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- Specify Device B, Device C and Device D as the transit nodes of primary ring 1, their GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- The timers of the primary ring adopt the default value.

First, determine the node mode of a device in an RRPP ring, and then perform the following configurations on a per-device basis:

- Disable STP on all ports accessing RRPP rings on these devices and configure these ports to permit the traffic of all VLANs to pass through.
- Configure the 802.1p priority for trusted packets on ports accessing RRPP rings on each device.
- Create an RRPP domain.
- Specify the control VLAN for the RRPP domain.
- Configure the MSTIs referenced by the protected VLANs. The MSTI ID ranges from 0 to 32.
- Specify the node mode of a device on the primary ring and the ports accessing the RRPP ring on the device.
- Enable the RRPP ring.
- Enable RRPP.

Figure 1-8 Network diagram for single ring networking configuration



Configuration procedure

1) Perform the following configuration on Device A:

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceA-GigabitEthernet1/0/1] qos trust dot1p
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceA-GigabitEthernet1/0/2] qos trust dot1p
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure the VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceA] rrpp domain 1
[DeviceA-rrpp-domain1] control-vlan 4092
[DeviceA-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2) Perform the following configuration on Device B:

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] undo stp enable
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/1] qos trust dot1p
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/2] qos trust dot1p
[DeviceB-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure the VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceB] rrpp domain 1
[DeviceB-rrpp-domain1] control-vlan 4092
[DeviceB-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device B as the transit node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
[DeviceB-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3) Perform the following configuration on Device C:

The configuration on Device C is similar to that on Device B and thus omitted here.

4) Perform the following configuration on Device D:

The configuration on Device D is similar to that on Device B and thus omitted here.

5) Verification

After the above configuration, you can use the **display** command to view RRPP configuration on each device.

Configuring Single-Domain Intersecting Ring Topology

Networking requirements

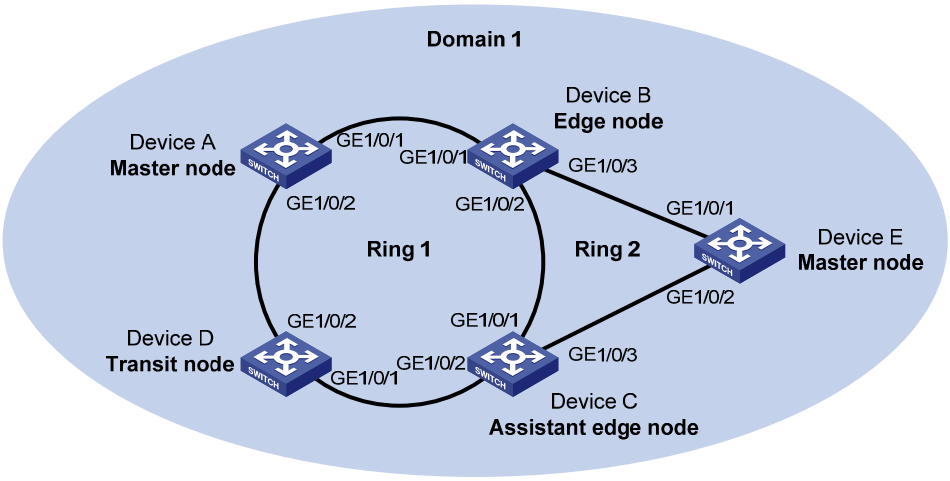
- Device A, Device B, Device C and Device D constitute RRPP domain 1, VLAN 4092 is the primary control VLAN of RRPP domain 1, and RRPP domain 1 protects all the VLANs;
- Device A, Device B, Device C and Device D constitute primary ring 1;
- Device B, Device C and Device E constitute subring 2;
- Device A is the master node of primary ring 1, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device E is the master node of subring 2, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device B is the transit node of primary ring 1 and the edge node of subring 2, and GigabitEthernet 1/0/3 is the edge port;
- Device C is the transit node of primary ring 1 and the assistant-edge node of subring 1, and GigabitEthernet 1/0/3 is the edge port;
- Device D is the transit node of primary ring 1, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- The timers of both the primary ring and the subring adopt the default value.

First, determine the primary ring and subring in an RRPP domain, node mode of a device on each RRPP ring, and then perform the following configuration on a per-device basis:

- Disable STP on all ports accessing RRPP rings on these devices and configure these ports to permit the traffic of all VLANs to pass through.
- Configure the 802.1p priority for trusted packets on ports accessing RRPP rings on each device.
- Create an RRPP domain.

- Specify the control VLAN for the RRPP domain.
- Configure the protected VLANs to reference all MSTIs. The MSTI ID ranges from 0 to 32.
- Specify the node mode of a device on an RRPP ring and the ports accessing the RRPP ring on the device.
- Enable these two RRPP rings.
- Enable RRPP

Figure 1-9 Network diagram for intersecting rings configuration



Configuration procedure

1) Configuration on Device A

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
<DeviceA> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceA-GigabitEthernet1/0/1] qos trust dot1p
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceA-GigabitEthernet1/0/2] qos trust dot1p
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure the VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceA] rrpp domain 1
[DeviceA-rrpp-domain1] control-vlan 4092
[DeviceA-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
```

```
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2) Configuration on Device B

Configure RRPP ports GigabitEthernet1/0/1, GigabitEthernet1/0/2 and GigabitEthernet1/0/3.

```
<DeviceB> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] undo stp enable
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/1] qos trust dot1p
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/2] qos trust dot1p
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] undo stp enable
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan all
[DeviceB-GigabitEthernet1/0/3] qos trust dot1p
[DeviceB-GigabitEthernet1/0/3] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure the VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceB] rrpp domain 1
[DeviceB-rrpp-domain1] control-vlan 4092
[DeviceB-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device B as a transit node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0//2 as the secondary port, and enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
```

Configure Device B as the edge node of subring 2, with GigabitEthernet1/0/3 as the edge port, and enable ring 2.

```
[DeviceB-rrpp-domain1] ring 2 node-mode edge edge-port gigabitethernet 1/0/3
[DeviceB-rrpp-domain1] ring 2 enable
[DeviceB-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3) Configuration on Device C

Configure RRPP ports GigabitEthernet1/0/1, GigabitEthernet1/0/2 and GigabitEthernet1/0/3.

```
<DeviceC> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo stp enable
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceC-GigabitEthernet1/0/1] qos trust dot1p
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceC-GigabitEthernet1/0/2] qos trust dot1p
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] undo stp enable
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan all
[DeviceC-GigabitEthernet1/0/3] qos trust dot1p
[DeviceC-GigabitEthernet1/0/3] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceC] rrpp domain 1
[DeviceC-rrpp-domain1] control-vlan 4092
[DeviceC-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device C as a transit node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceC-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain1] ring 1 enable
```

Configure Device C as the assistant-edge node of subring 2, with GigabitEthernet1/0/3 as the edge port, and enable ring 2.

```
[DeviceC-rrpp-domain1] ring 2 node-mode assistant-edge edge-port gigabitethernet 1/0/3
[DeviceC-rrpp-domain1] ring 2 enable
[DeviceC-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceC] rrpp enable
```

4) Configuration on Device D

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
<DeviceD> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] undo stp enable
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceD-GigabitEthernet1/0/1] qos trust dot1p
[DeviceD-GigabitEthernet1/0/1] quit
```

```
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceD-GigabitEthernet1/0/2] qos trust dot1p
[DeviceD-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceD] rrpp domain 1
[DeviceD-rrpp-domain1] control-vlan 4092
[DeviceD-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device D as the transit node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceD-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain1] ring 1 enable
[DeviceD-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceD] rrpp enable
```

5) Configuration on Device E

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
<DeviceE> system-view
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] undo stp enable
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan all
[DeviceE-GigabitEthernet1/0/1] qos trust dot1p
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] undo stp enable
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan all
[DeviceE-GigabitEthernet1/0/2] qos trust dot1p
[DeviceE-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 4092 as the primary control VLAN of RRPP domain 1, and configure VLANs mapped to MSTIs 0 through 32 as the protected VLANs of RRPP domain 1.

```
[DeviceE] rrpp domain 1
[DeviceE-rrpp-domain1] control-vlan 4092
[DeviceE-rrpp-domain1] protected-vlan reference-instance 0 to 32
```

Configure Device E as the master node of subring 2, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 2.

```
[DeviceE-rrpp-domain1] ring 2 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 1
[DeviceE-rrpp-domain1] ring 2 enable
[DeviceE-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceE] rrpp enable
```

6) Verification

After the configuration, you can use the **display** command to view RRPP configuration result on each device.

Configuring Intersecting-Ring Load Balancing

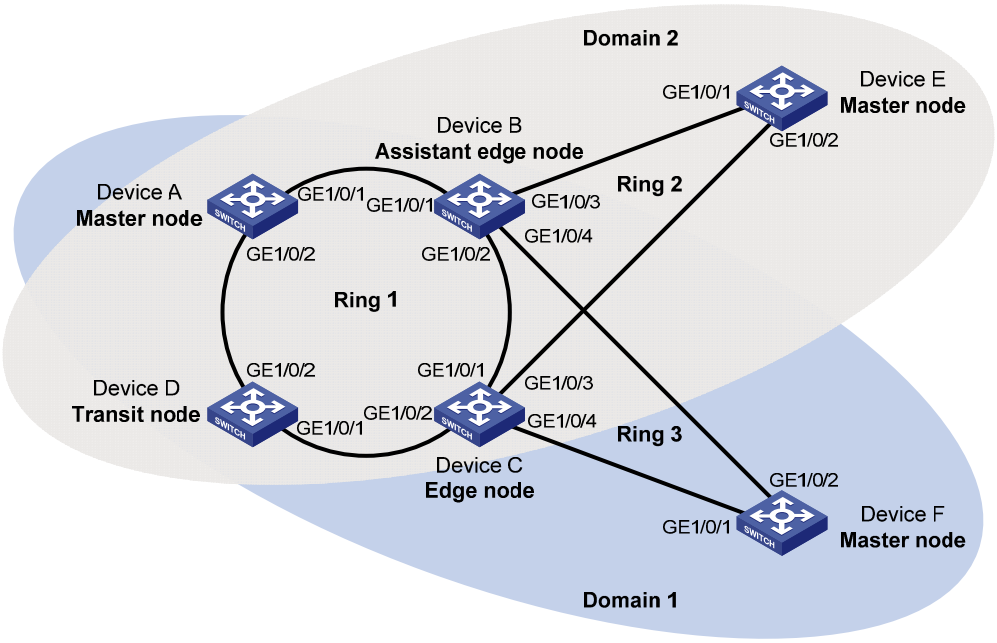
Networking requirements

- Device A, Device B, Device C, Device D, and Device F constitute RRPP domain 1, and VLAN 100 is the primary control VLAN of the RRPP domain. Device A is the master node of the primary ring Ring 1; Device D is the transit node of the primary ring Ring 1; Device F is the master node of the subring Ring 3; Device C is the edge node of the subring Ring 3; Device B is the assistant-edge node of the subring Ring 3.
- Device A, Device B, Device C, Device D, and Device E constitute RRPP domain 2, and VLAN 105 is the primary control VLAN of the RRPP domain. Device A is the master node of the primary ring Ring 1; Device D is the transit node of the primary ring Ring 1; Device E is the master node of the subring Ring 2; Device C is the edge node of the subring Ring 2; Device B is the assistant-edge node of the subring Ring 2.
- Specify VLAN 10 as the protected VLAN of domain 1, and VLAN 20 as the protected VLAN of domain 2. Thus, you can achieve VLAN-based load balancing on the primary ring.
- As the edge node and assistant-edge node of subring Ring 2 is the same as those of subring Ring 3, and the two subrings have the same SRPTs, you can add subrings Ring 2 and Ring 3 to the RRPP ring group to reduce Edge-Hello traffic.

According to the diagram as shown Figure 1-10, perform the following configurations:

- Create data VLANs, and map the VLANs to be protected in each RRPP domain to different MSTIs.
- Disable STP on all ports accessing RRPP rings on these devices and configure the VLANs whose traffic is permitted to pass through.
- Configure the 802.1p priority for trusted packets on ports accessing RRPP rings on each device.
- Create RRPP domains.
- Specify control VLANs for RRPP domains.
- Specify protected VLANs for each domain by specifying MSTIs.
- Specify the roles of devices in these RRPP rings and the ports accessing RRPP rings.
- Enable RRPP rings.
- Enable the RRPP protocol.
- Configure a ring group on the edge nodes and assistant-edge nodes.

Figure 1-10 Network diagram for intersecting-ring load balancing configuration



Configuration procedure

1) Configure Device A as the master node of the primary ring

Create VLANs 10 and 20, and map VLAN 10 to MSTI 1 and VLAN 20 to MSTI 2.

```
<DeviceA> system-view
[DeviceA] vlan 10
[DeviceA-vlan10] quit
[DeviceA] vlan 20
[DeviceA-vlan20] quit
[DeviceA] stp region-configuration
[DeviceA-mst-region] instance 1 vlan 10
[DeviceA-mst-region] instance 2 vlan 20
[DeviceA-mst-region] active region-configuration
[DeviceA-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo stp enable
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 10 20
[DeviceA-GigabitEthernet1/0/1] qos trust dot1p
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] undo stp enable
[DeviceA-GigabitEthernet1/0/2] port link-type trunk
[DeviceA-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceA-GigabitEthernet1/0/2] port trunk permit vlan 10 20
[DeviceA-GigabitEthernet1/0/2] qos trust dot1p
```

```
[DeviceA-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 100 as the primary control VLAN of RRPP domain 1, and configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceA] rrpp domain 1
[DeviceA-rrpp-domain1] control-vlan 100
[DeviceA-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceA-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[DeviceA-rrpp-domain1] ring 1 enable
[DeviceA-rrpp-domain1] quit
```

Create RRPP domain 2, configure VLAN 105 as the primary control VLAN of RRPP domain 2, and configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceA] rrpp domain 2
[DeviceA-rrpp-domain2] control-vlan 105
[DeviceA-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device A as the master node of primary ring 1, with GigabitEthernet1/0/2 as the master port and GigabitEthernet1/0/1 as the secondary port, and enable ring 1.

```
[DeviceA-rrpp-domain2] ring 1 node-mode master primary-port gigabitethernet 1/0/2 secondary-port gigabitethernet 1/0/1 level 0
[DeviceA-rrpp-domain2] ring 1 enable
[DeviceA-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceA] rrpp enable
```

2) Configure Device B as the assistant-edge node of subrings Ring 2 and Ring 3

Create VLANs 10 and 20, and map VLAN 10 to MSTI 1 and VLAN 20 to MSTI 2.

```
<DeviceB> system-view
[DeviceB] vlan 10
[DeviceB-vlan10] quit
[DeviceB] vlan 20
[DeviceB-vlan20] quit
[DeviceB] stp region-configuration
[DeviceB-mst-region] instance 1 vlan 10
[DeviceB-mst-region] instance 2 vlan 20
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1, GigabitEthernet1/0/2, GigabitEthernet1/0/3, and GigabitEthernet1/0/4.

```
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] undo stp enable
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan 10 20
```

```
[DeviceB-GigabitEthernet1/0/1] qos trust dot1p
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] undo stp enable
[DeviceB-GigabitEthernet1/0/2] port link-type trunk
[DeviceB-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/2] port trunk permit vlan 10 20
[DeviceB-GigabitEthernet1/0/2] qos trust dot1p
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] undo stp enable
[DeviceB-GigabitEthernet1/0/3] port link-type trunk
[DeviceB-GigabitEthernet1/0/3] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/3] port trunk permit vlan 20
[DeviceB-GigabitEthernet1/0/3] qos trust dot1p
[DeviceB-GigabitEthernet1/0/3] quit
[DeviceB] interface gigabitethernet 1/0/4
[DeviceB-GigabitEthernet1/0/4] undo stp enable
[DeviceB-GigabitEthernet1/0/4] port link-type trunk
[DeviceB-GigabitEthernet1/0/4] undo port trunk permit vlan 1
[DeviceB-GigabitEthernet1/0/4] port trunk permit vlan 10
[DeviceB-GigabitEthernet1/0/4] qos trust dot1p
[DeviceB-GigabitEthernet1/0/4] quit
```

Create RRPP domain 1, configure VLAN 100 as the primary control VLAN of RRPP domain 1, and configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceB] rrpp domain 1
[DeviceB-rrpp-domain1] control-vlan 100
[DeviceB-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device B as a transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceB-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain1] ring 1 enable
```

Configure Device B as the assistant-edge node of subring 3 in RRPP domain 1, with GigabitEthernet1/0/4 as the edge port, and enable subring 3.

```
[DeviceB-rrpp-domain1] ring 3 node-mode assistant-edge edge-port gigabitethernet 1/0/4
[DeviceB-rrpp-domain1] ring 3 enable
[DeviceB-rrpp-domain1] quit
```

Create RRPP domain 2, configure VLAN 105 as the primary control VLAN of RRPP domain 2, and configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceB] rrpp domain 2
[DeviceB-rrpp-domain2] control-vlan 105
[DeviceB-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device B as the transit node of primary ring 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceB-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[DeviceB-rrpp-domain2] ring 1 enable
```

Configure Device B as the assistant-edge node of subring 2 in RRPP domain 2, with GigabitEthernet1/0/3 as the edge port, and enable subring 2.

```
[DeviceB-rrpp-domain2] ring 2 node-mode assistant-edge edge-port gigabitethernet 1/0/3
[DeviceB-rrpp-domain2] ring 2 enable
[DeviceB-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceB] rrpp enable
```

3) Configure Device C as the edge node of subrings Ring 2 and Ring 3

Create VLANs 10 and 20, and map VLAN 10 to MSTI 1 and VLAN 20 to MSTI 2.

```
<DeviceC> system-view
[DeviceC] vlan 10
[DeviceC-vlan10] quit
[DeviceC] vlan 20
[DeviceC-vlan20] quit
[DeviceC] stp region-configuration
[DeviceC-mst-region] instance 1 vlan 10
[DeviceC-mst-region] instance 2 vlan 20
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1, GigabitEthernet1/0/2, GigabitEthernet1/0/3, and GigabitEthernet1/0/4.

```
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] undo stp enable
[DeviceC-GigabitEthernet1/0/1] port link-type trunk
[DeviceC-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/1] port trunk permit vlan 10 20
[DeviceC-GigabitEthernet1/0/1] qos trust dot1p
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] undo stp enable
[DeviceC-GigabitEthernet1/0/2] port link-type trunk
[DeviceC-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/2] port trunk permit vlan 10 20
[DeviceC-GigabitEthernet1/0/2] qos trust dot1p
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] undo stp enable
[DeviceC-GigabitEthernet1/0/3] port link-type trunk
[DeviceC-GigabitEthernet1/0/3] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/3] port trunk permit vlan 20
[DeviceC-GigabitEthernet1/0/3] qos trust dot1p
[DeviceC-GigabitEthernet1/0/3] quit
[DeviceC] interface gigabitethernet 1/0/4
```

```
[DeviceC-GigabitEthernet1/0/4] undo stp enable
[DeviceC-GigabitEthernet1/0/4] port link-type trunk
[DeviceC-GigabitEthernet1/0/4] undo port trunk permit vlan 1
[DeviceC-GigabitEthernet1/0/4] port trunk permit vlan 10
[DeviceC-GigabitEthernet1/0/4] qos trust dot1p
[DeviceC-GigabitEthernet1/0/4] quit
```

Create RRPP domain 1, configure VLAN 10 as the primary control VLAN of RRPP domain 1, and configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceC] rrpp domain 1
[DeviceC-rrpp-domain1] control-vlan 100
[DeviceC-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device C as the transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceC-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain1] ring 1 enable
```

Configure Device C as the edge node of subring 3 in RRPP domain 1, with GigabitEthernet1/0/4 as the edge port, and enable subring 3.

```
[DeviceC-rrpp-domain1] ring 3 node-mode edge edge-port gigabitethernet 1/0/4
[DeviceC-rrpp-domain1] ring 3 enable
[DeviceC-rrpp-domain1] quit
```

Create RRPP domain 2, configure VLAN 105 as the primary control VLAN of RRPP domain 2, and configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceC] rrpp domain 2
[DeviceC-rrpp-domain2] control-vlan 105
[DeviceC-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device C as the transit node of primary ring 1 in RRPP domain 2, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceC-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[DeviceC-rrpp-domain2] ring 1 enable
```

Configure Device C as the edge node of subring 2 in RRPP domain 2, with GigabitEthernet1/0/3 as the edge port, and enable subring 2.

```
[DeviceC-rrpp-domain2] ring 2 node-mode edge edge-port gigabitethernet 1/0/3
[DeviceC-rrpp-domain2] ring 2 enable
[DeviceC-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceC] rrpp enable
```

4) Configure Device D as a transit node of the primary ring

Create VLANs 10 and 20, and map VLAN 10 to MSTI 1 and VLAN 20 to MSTI 2.

```
<DeviceD> system-view
[DeviceD] vlan 10
[DeviceD-vlan10] quit
[DeviceD] vlan 20
```

```
[DeviceD-vlan20] quit
[DeviceD] stp region-configuration
[DeviceD-mst-region] instance 1 vlan 10
[DeviceD-mst-region] instance 2 vlan 20
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] undo stp enable
[DeviceD-GigabitEthernet1/0/1] port link-type trunk
[DeviceD-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceD-GigabitEthernet1/0/1] port trunk permit vlan 10 20
[DeviceD-GigabitEthernet1/0/1] qos trust dot1p
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] undo stp enable
[DeviceD-GigabitEthernet1/0/2] port link-type trunk
[DeviceD-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceD-GigabitEthernet1/0/2] port trunk permit vlan 10 20
[DeviceD-GigabitEthernet1/0/2] qos trust dot1p
[DeviceD-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 100 as the primary control VLAN of RRPP domain 1, and configure the VLAN mapped to MSTI 1 as the protected VLAN of RRPP domain 1.

```
[DeviceD] rrpp domain 1
[DeviceD-rrpp-domain1] control-vlan 100
[DeviceD-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device D as the transit node of primary ring 1 in RRPP domain 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceD-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain1] ring 1 enable
[DeviceD-rrpp-domain1] quit
```

Create RRPP domain 2, configure VLAN 105 as the primary control VLAN of RPPP domain 2, and configure the VLAN mapped to MSTI 2 as the protected VLAN of RRPP domain 2.

```
[DeviceD] rrpp domain 2
[DeviceD-rrpp-domain2] control-vlan 105
[DeviceD-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device D as the transit node of primary ring 1 in RRPP domain 2, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable ring 1.

```
[DeviceD-rrpp-domain2] ring 1 node-mode transit primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 0
[DeviceD-rrpp-domain2] ring 1 enable
[DeviceD-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceD] rrpp enable
```

5) Configure Device E as the master node of subring Ring 2 in domain 2

Create VLAN 20, and map VLAN 20 to MSTI 2.

```
<DeviceE> system-view
[DeviceE] vlan 20
[DeviceE-vlan20] quit
[DeviceE] stp region-configuration
[DeviceE-mst-region] instance 2 vlan 20
[DeviceE-mst-region] active region-configuration
[DeviceE-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] undo stp enable
[DeviceE-GigabitEthernet1/0/1] port link-type trunk
[DeviceE-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceE-GigabitEthernet1/0/1] port trunk permit vlan 20
[DeviceE-GigabitEthernet1/0/1] qos trust dot1p
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] undo stp enable
[DeviceE-GigabitEthernet1/0/2] port link-type trunk
[DeviceE-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceE-GigabitEthernet1/0/2] port trunk permit vlan 20
[DeviceE-GigabitEthernet1/0/2] qos trust dot1p
[DeviceE-GigabitEthernet1/0/2] quit
```

Create RRPP domain 2, configure VLAN 105 as the primary control VLAN, and configure the VLAN mapped to MSTI 2 as the protected VLAN.

```
[DeviceE] rrpp domain 2
[DeviceE-rrpp-domain2] control-vlan 105
[DeviceE-rrpp-domain2] protected-vlan reference-instance 2
```

Configure Device E as the master mode of subring 2 in RRPP domain 2, with GigabitEthernet1/0/2 as the primary port and GigabitEthernet1/0/1 as the secondary port, and enable ring 2.

```
[DeviceE-rrpp-domain2] ring 2 node-mode master primary-port gigabitethernet 1/0/2
secondary-port gigabitethernet 1/0/1 level 1
[DeviceE-rrpp-domain2] ring 2 enable
[DeviceE-rrpp-domain2] quit
```

Enable RRPP.

```
[DeviceE] rrpp enable
```

6) Configure Device F as the master node of subring Ring 3 in domain 1

Create VLAN 10, and map VLAN 10 to MSTI 1.

```
<DeviceF> system-view
[DeviceF] vlan 10
[DeviceF-vlan10] quit
[DeviceF] stp region-configuration
[DeviceF-mst-region] instance 1 vlan 10
[DeviceF-mst-region] active region-configuration
```

```
[DeviceF-mst-region] quit
```

Configure RRPP ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2.

```
[DeviceF] interface gigabitethernet 1/0/1
[DeviceF-GigabitEthernet1/0/1] undo stp enable
[DeviceF-GigabitEthernet1/0/1] port link-type trunk
[DeviceF-GigabitEthernet1/0/1] undo port trunk permit vlan 1
[DeviceF-GigabitEthernet1/0/1] port trunk permit vlan 10
[DeviceF-GigabitEthernet1/0/1] qos trust dot1p
[DeviceF-GigabitEthernet1/0/1] quit
[DeviceF] interface gigabitethernet 1/0/2
[DeviceF-GigabitEthernet1/0/2] undo stp enable
[DeviceF-GigabitEthernet1/0/2] port link-type trunk
[DeviceF-GigabitEthernet1/0/2] undo port trunk permit vlan 1
[DeviceF-GigabitEthernet1/0/2] port trunk permit vlan 10
[DeviceF-GigabitEthernet1/0/2] qos trust dot1p
[DeviceF-GigabitEthernet1/0/2] quit
```

Create RRPP domain 1, configure VLAN 100 as the primary control VLAN, and configure the VLAN mapped to MSTI 1 as the protected VLAN.

```
[DeviceF] rrpp domain 1
[DeviceF-rrpp-domain1] control-vlan 100
[DeviceF-rrpp-domain1] protected-vlan reference-instance 1
```

Configure Device F as the master node of subring 3 in RRPP domain 1, with GigabitEthernet1/0/1 as the primary port and GigabitEthernet1/0/2 as the secondary port, and enable subring 3.

```
[DeviceF-rrpp-domain1] ring 3 node-mode master primary-port gigabitethernet 1/0/1
secondary-port gigabitethernet 1/0/2 level 1
[DeviceF-rrpp-domain1] ring 3 enable
[DeviceF-rrpp-domain1] quit
```

Enable RRPP.

```
[DeviceF] rrpp enable
```

7) Configure a ring group on Device B and Device C after the configurations above

Create RRPP ring group 1 on Device B, and add subrings 2 and 3 to the RRPP ring group.

```
[DeviceB] rrpp ring-group 1
[DeviceB-rrpp-ring-group1] domain 2 ring 2
[DeviceB-rrpp-ring-group1] domain 1 ring 3
```

Create RRPP ring group 1 on Device C, and add subrings 2 and 3 to the RRPP ring group.

```
[DeviceC] rrpp ring-group 1
[DeviceC-rrpp-ring-group1] domain 2 ring 2
[DeviceC-rrpp-ring-group1] domain 1 ring 3
```

8) Verification

After the configuration, you can use the **display** command to view RRPP configuration result on each device.

Troubleshooting

Symptom:

When the link state is normal, the master node cannot receive Hello packets, and the master node unblocks the secondary port.

Analysis:

The reasons may be:

- RRPP is not enabled on some nodes in the RRPP ring.
- The domain ID or primary control VLAN ID is not the same for the nodes in the same RRPP ring.
- Some ports are abnormal.

Solution:

- Use the **display rrpp brief** command to check whether RRPP is enabled for all nodes. If not, use the **rrpp enable** command and the **ring enable** command to enable RRPP and RRPP rings for all nodes.
- Use the **display rrpp brief** command to check whether the domain ID and primary control VLAN ID are the same for all nodes. If not, set the same domain ID and primary control VLAN ID for the nodes.
- Use the **display rrpp verbose** command to check the link state of each port in each ring.
- Use the **debugging rrpp** command on each node to check whether a port receives or transmits Hello packets. If not, Hello packets are lost.

Table of Contents

1 Port Mirroring Configuration

When configuring port mirroring, go to these sections for information you are interested in:

Introduction to Port Mirroring

Port mirroring is to copy the packets passing through a port (called a mirroring port) to another port (called the monitor port) connected with a monitoring device for packet analysis.

You can select to port-mirror inbound, outbound, or bidirectional traffic on a port/VLAN as needed.

Classification of Port Mirroring

Port mirroring can be local or remote.

- In local port mirroring, the mirroring port or ports and the monitor port are located on the same device.
- In remote port mirroring, the mirroring port or ports and the monitor port can be located on the same device or different devices. Currently, remote port mirroring can be implemented only at Layer 2.

Note

As a monitor port can monitor multiple ports, it may receive multiple duplicates of a packet in some cases. Suppose that port P 1 is monitoring bidirectional traffic on ports P 2 and P 3 on the same device. If a packet travels from P 2 to P 3, two duplicates of the packet will be received on P 1.

Implementing Port Mirroring

Port mirroring is implemented through port mirroring groups. There are three types of mirroring groups: local, remote source, and remote destination.

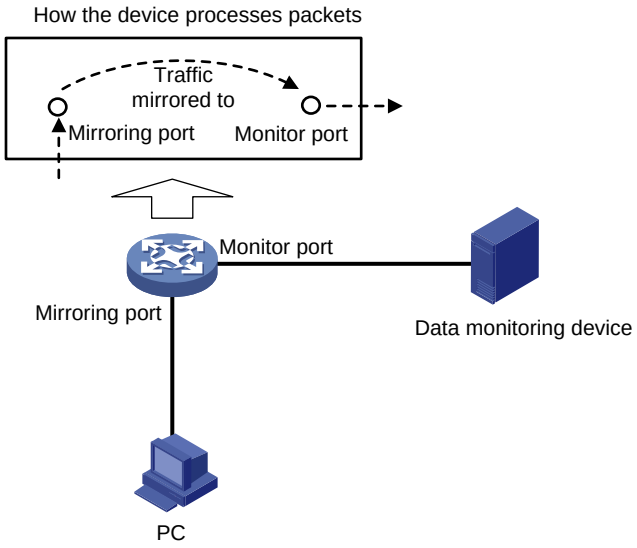
The following subsections describe how local port mirroring and remote port mirroring are implemented.

Local port mirroring

In local port mirroring, all packets passing through a port can be mirrored. Local port mirroring is implemented through a local mirroring group.

As shown in Figure 1-1, packets on the mirroring port are mirrored to the monitor port for the data monitoring device to analyze.

Figure 1-1 Local port mirroring implementation

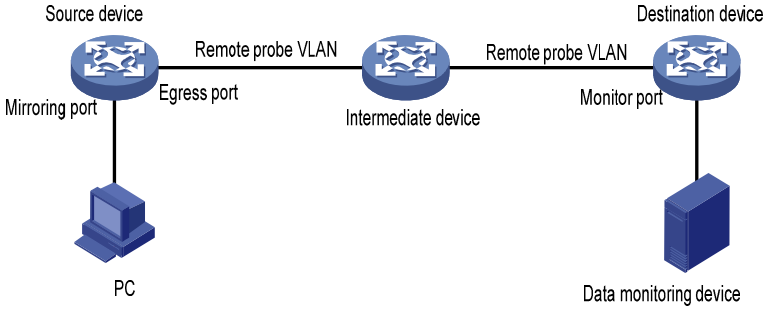


Remote port mirroring

Remote port mirroring can mirror all packets but protocol packets.

Remote port mirroring is implemented through the cooperation of a remote source mirroring group and a remote destination mirroring group as shown Figure 1-2.

Figure 1-2 Remote port mirroring implementation (with an egress port)



Remote mirroring involves the following device roles:

- Source device

The source device is the device where the mirroring ports are located. On it, you must create a remote source mirroring group to hold the mirroring ports.

The source device copies the packets passing through the mirroring ports, broadcasts the packets in the remote probe VLAN for remote mirroring, and transmits the packets to the next device, which could be an intermediate device (if any) or the destination device.

- Intermediate device

Intermediate devices are devices located in between the source device and the destination device.

An intermediate device forwards mirrored packets to the next intermediate device (if any) or the destination device.

You must ensure that the source device and the destination device can communicate at Layer 2 in the remote probe VLAN.

- Destination device

The destination device is the device where the monitor port is located. On it, you must create the remote destination mirroring group.

When receiving a packet, the destination device compares the VLAN ID carried in the packet with the ID of the probe VLAN configured in the remote destination mirroring group. If they are the same, the device forwards the packet to the monitoring device through the monitor port.

Note

To make the port mirroring function work properly, before configuring bidirectional traffic mirroring on a port in a mirroring group, you need to use the **mac-address mac-learning disable** command on the source device, intermediate devices, and destination device to disable the MAC address learning function for the remote port mirroring VLAN. For more information about the **mac-address mac-learning disable** command, refer to *MAC Address Table Management Commands* in the *System Volume*.

Configuring Local Port Mirroring

Configuring local port mirroring is to configure local mirroring groups.

A local mirroring group comprises one or multiple mirroring ports and one monitor port. These ports must not have been assigned to any other mirroring group.

Follow these steps to configure a local mirroring group:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a local mirroring group		**mirroring-group** *groupid* **local**	Required
Configure mirroring ports	In system view	**mirroring-group** *groupid* **mirroring-port** *mirroring-port-list* { **both** | **inbound** | **outbound** }	Required You can configure mirroring ports in a mirroring group. In system view, you can configure a list of mirroring ports to the mirroring group at a time. In interface view, you can assign only the current port to the mirroring group. To monitor multiple ports, repeat the step.
	In interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **mirroring-port** { **both** | **inbound** | **outbound** }	
		quit	
Configure the monitor port	In system view	**mirroring-group** *groupid* **monitor-port** *monitor-port-id*	Required Use either approach.
	In interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **monitor-port**	

Note

- A local port mirroring group takes effect only after its mirroring and monitor ports are configured.
- To ensure operation of your device, do not enable STP, MSTP, or RSTP on the monitor port.
- A port mirroring group can have multiple mirroring ports, but only one monitor port.
- A mirroring or monitor port to be configured cannot belong to an existing port mirroring group.
- You are recommended to use a monitor port only for port mirroring. This is to ensure that the data monitoring device receives and analyzes only the mirrored traffic rather than a mix of mirrored traffic and normally forwarded traffic.

Configuring Remote Port Mirroring

Configuring remote port mirroring is to configure remote mirroring groups. When doing that, configure the remote source mirroring group on the source device and the cooperating remote destination mirroring group on the destination device.

Note

If GVRP is enabled, GVRP may register the remote probe VLAN to unexpected ports, resulting in undesired duplicates. For information on GVRP, refer to *GVRP Configuration* in the *Access Volume*.

Configuration Prerequisites

Create a static VLAN for the probe VLAN on the source and destination device. To ensure correct packet handling, ensure that the VLANs you created on the two devices use the same ID and function only for remote port mirroring.

Configuring a Remote Source Mirroring Group (on the Source Device)

A remote source mirroring group comprises the following:

- One or multiple mirroring ports.
- A remote probe VLAN.
- An egress port.

After you assign a port to a mirroring group either as a mirroring port or as a monitor port, you cannot assign it to any other mirroring group. The same is true of probe VLANs.

Configuring a remote source mirroring group with an egress port

Follow these steps to configure a remote port mirroring group with an egress port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a remote source mirroring group	**mirroring-group** *groupid* **remote-source**	Required

To do…		Use the command…	Remarks
Configure mirroring ports	In system view	**mirroring-group** *groupid* **mirroring-port** *mirroring-port-list* { **both** | **inbound** | **outbound** }	Required You configure multiple mirroring ports in a mirroring group. In system view, you can assign a list of mirroring ports to the mirroring group at a time. In interface view, you can assign only the current interface to the mirroring group. To monitor multiple ports, repeat the step.
	In interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **mirroring-port** { **both** | **inbound** | **outbound** }	
		quit	
Configure the egress port	In system view	**mirroring-group** *groupid* **monitor-egress** *monitor-egress-port-id*	Required Use either approach.
	In interface view	**interface** *interface-type interface-number*	
		mirroring-group *groupid* **monitor-egress**	
		quit	
Configure the probe VLAN		**mirroring-group** *groupid* **remote-probe vlan** *rprobe-vlan-id*	Required

Note

When configuring the mirroring ports, note that:

- The mirroring ports and the egress port must be located on the same device.
- To ensure device performance, do not assign the mirroring ports to the remote probe VLAN.

Note

When configuring the egress port, note that:

- The port must not be a mirroring port in the mirroring group.
- To ensure operation of the device, disable these functions on the port: STP, MSTP, RSTP, 802.1X, IGMP Snooping, static ARP, and MAC address learning.
- A remote port mirroring group can have only one egress port.

Note

To remove the VLAN configured as a remote probe VLAN, you must remove the remote probe VLAN with **undo mirroring-group remote-probe vlan** command first. Removing the probe VLAN can invalidate the remote source mirroring group.

Configuring a Remote Destination Mirroring Group (on the Destination Device)

A remote destination mirroring group comprises a remote probe VLAN and a monitor port. You must ensure that the remote probe VLAN is the same as the one configured in the remote source mirroring group.

Follow these steps to configure a remote destination port mirroring group:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a remote destination mirroring group		**mirroring-group** *groupid* **remote-destination**	Required
Configure the remote probe VLAN		**mirroring-group** *groupid* **remote-probe vlan** *rprobe-vlan-id*	Required
Configure the monitor port	In system view	**mirroring-group** *groupid* **monitor-port** *monitor-port-id*	Required Use either approach.
	In interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **monitor-port**	
		quit	
Enter the interface view of the monitor port		**interface** *interface-type interface-number*	—
Assign the port to the probe VLAN	For an access port	**port access vlan** *rprobe-vlan-id*	Required Use one of the commands depending on the link type of the monitor port.
	For a trunk port	**port trunk permit vlan** *rprobe-vlan-id*	
	For a hybrid port	**port hybrid vlan** *rprobe-vlan-id* { **tagged** | **untagged** }	

Note

When configuring the probe VLAN, use the following guidelines:

- A VLAN can be the remote probe VLAN of only one port mirroring group.
- To remove the VLAN configured as the remote probe VLAN, you must remove the remote probe VLAN with **undo mirroring-group remote-probe vlan** command first. Removing the probe VLAN can invalidate the remote source mirroring group.

Note

When configuring the monitor port, use the following guidelines:

- The port can belong to only the current mirroring group.
- To ensure operation of your device, do not assign the monitor port to a mirroring VLAN.
- Disable these functions on the port: STP, MSTP, and RSTP.
- You are recommended to use a monitor port only for port mirroring. This is to ensure that the data monitoring device receives and analyzes only the mirrored traffic rather than a mix of mirrored traffic and normally forwarded traffic.

Displaying and Maintaining Port Mirroring

To do…	Use the command…	Remarks
Display the configuration of port mirroring groups	**display mirroring-group** { *groupid* | **all** | **local** | **remote-destination** | **remote-source** }	Available in any view

Port Mirroring Configuration Examples

Local Port Mirroring Configuration Example

Network requirements

The departments of a company connect to each other through Ethernet switches:

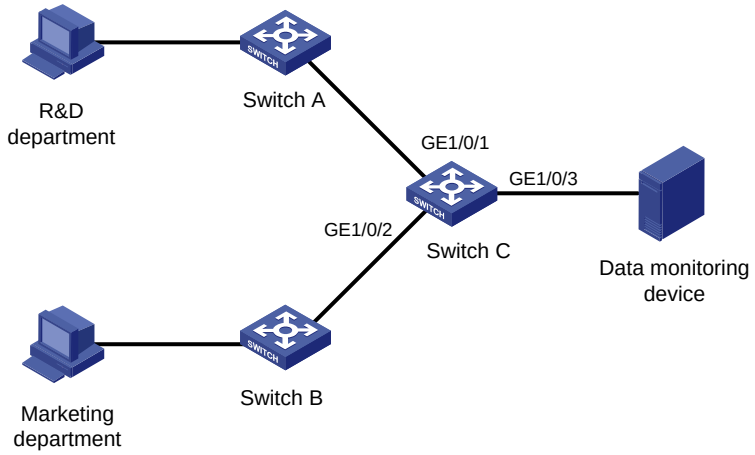
- Research and Development (R&D) department is connected to Switch C through GigabitEthernet 1/0/1.
- Marketing department is connected to Switch C through GigabitEthernet 1/0/2.
- Data monitoring device is connected to Switch C through GigabitEthernet 1/0/3

As shown in Figure 1-3, the administrator wants to monitor the packets received on and sent from the R&D department and the marketing department through the data monitoring device.

Use the local port mirroring function to meet the requirement. Perform the following configurations on Switch C.

- Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as mirroring source ports.
- Configure GigabitEthernet 1/0/3 as the mirroring destination port.

Figure 1-3 Network diagram for local port mirroring configuration



Configuration procedure

Configure Switch C.

Create a local port mirroring group.

```
<SwitchC> system-view
[SwitchC] mirroring-group 1 local
```

Add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the port mirroring group as source ports. Add port GigabitEthernet 1/0/3 to the port mirroring group as the destination port.

```
[SwitchC] mirroring-group 1 mirroring-port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2 both
[SwitchC] mirroring-group 1 monitor-port GigabitEthernet 1/0/3
```

Display the configuration of all the port mirroring groups.

```
[SwitchC] display mirroring-group all
mirroring-group 1:
    type: local
    status: active
    mirroring port:
        GigabitEthernet1/0/1  both
        GigabitEthernet1/0/2  both
    monitor port: GigabitEthernet1/0/3
```

After finishing the configuration, you can monitor all the packets received and sent by R&D department and Marketing department on the Data monitoring device.

Remote Port Mirroring Configuration Example

Network requirements

The departments of a company connect to each other through Ethernet switches:

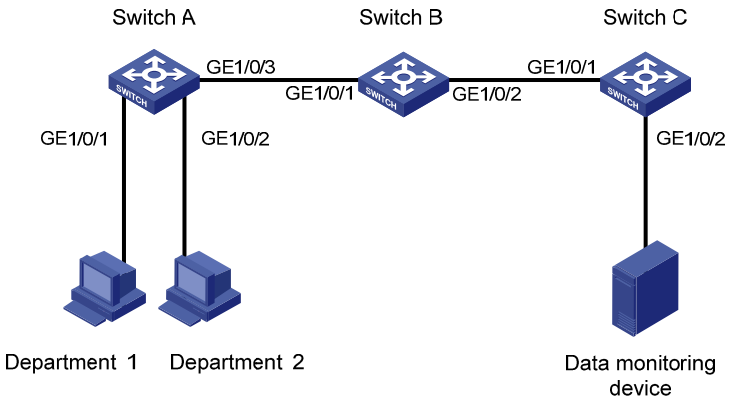
- Department 1 is connected to GigabitEthernet 1/0/1 of Switch A.
- Department 2 is connected to GigabitEthernet 1/0/2 of Switch A.
- GigabitEthernet 1/0/3 of Switch A connects to GigabitEthernet 1/0/1 of Switch B.
- GigabitEthernet 1/0/2 of Switch B connects to GigabitEthernet 1/0/1 of Switch C.
- The data monitoring device is connected to GigabitEthernet 1/0/2 of Switch C.

As shown in Figure 1-4, the administrator wants to monitor the packets sent from Department 1 and 2 through the data monitoring device.

Use the remote port mirroring function to meet the requirement. Perform the following configurations:

- Use Switch A as the source device, Switch B as the intermediate device, and Switch C as the destination device.
- On Switch A, create a remote source mirroring group; create VLAN 2 and configure it as the remote port mirroring VLAN; add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the port mirroring group as two source ports. Configure port GigabitEthernet 1/0/3 as the outbound mirroring port.
- Configure port GigabitEthernet 1/0/3 of Switch A, port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of Switch B, and port GigabitEthernet 1/0/1 of Switch C as trunk ports and configure them to permit packets of VLAN 2.
- Create a remote destination mirroring group on Switch C. Configure VLAN 2 as the remote port mirroring VLAN and port GigabitEthernet 1/0/2, to which the data monitoring device is connected, as the destination port.

Figure 1-4 Network diagram for remote port mirroring configuration



Configuration procedure

1) Configure Switch A (the source device).

Create a remote source port mirroring group.

```
<SwitchA> system-view
[SwitchA] mirroring-group 1 remote-source
```

Create VLAN 2.

```
[SwitchA] vlan 2
[SwitchA-vlan2] quit
```

Configure VLAN 2 as the remote port mirroring VLAN of the remote port mirroring group. Add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the remote port mirroring group as source ports. Configure port GigabitEthernet 1/0/3 as the outbound mirroring port.

```
[SwitchA] mirroring-group 1 remote-probe vlan 2
[SwitchA] mirroring-group 1 mirroring-port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2
inbound
[SwitchA] mirroring-group 1 monitor-egress GigabitEthernet 1/0/3
```

Configure port GigabitEthernet 1/0/3 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[SwitchA] interface GigabitEthernet 1/0/3
```

```
[SwitchA-GigabitEthernet1/0/3] port link-type trunk
[SwitchA-GigabitEthernet1/0/3] port trunk permit vlan 2
```

2) Configure Switch B (the intermediate device).

Configure port GigabitEthernet 1/0/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
<SwitchB> system-view
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-type trunk
[SwitchB-GigabitEthernet1/0/1] port trunk permit vlan 2
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure port GigabitEthernet 1/0/2 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[SwitchB] interface GigabitEthernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type trunk
[SwitchB-GigabitEthernet1/0/2] port trunk permit vlan 2
```

3) Configure Switch C (the destination device).

Configure port GigabitEthernet 1/0/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
<SwitchC> system-view
[SwitchC] interface GigabitEthernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] port link-type trunk
[SwitchC-GigabitEthernet1/0/1] port trunk permit vlan 2
[SwitchC-GigabitEthernet1/0/1] quit
```

Create a remote destination port mirroring group.

```
[SwitchC] mirroring-group 1 remote-destination
```

Create VLAN 2.

```
[SwitchC] vlan 2
[SwitchC-vlan2] quit
```

Configure VLAN 2 as the remote port mirroring VLAN of the remote destination port mirroring group. Add port GigabitEthernet 1/0/2 to the remote destination port mirroring group as the destination port.

```
[SwitchC] mirroring-group 1 remote-probe vlan 2
[SwitchC] mirroring-group 1 monitor-port GigabitEthernet 1/0/2
[SwitchC] interface GigabitEthernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] port access vlan 2
```

After finishing the configuration, you can monitor all the packets sent by Department 1 and Department 2 on the Data monitoring device.

IP Services Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The IP Services Volume is organized as follows:

Features	Description
IP Address	An IP address is a 32-bit address allocated to a network interface on a device that is attached to the Internet. This document introduces the commands for IP address configuration
ARP	Address Resolution Protocol (ARP) is used to resolve an IP address into a data link layer address. This document introduces the commands for: • Configuring ARP • Configuring Gratuitous ARP • Proxy ARP and Local Proxy ARP configuration • ARP Attack Defense configuration
DHCP	DHCP is built on a client-server model, in which the client sends a configuration request and then the server returns a reply to send configuration parameters such as an IP address to the client. This document introduces the commands for: • DHCP server configuration • DHCP relay agent configuration • DHCP Client configuration • DHCP Snooping configuration • BOOTP Client configuration
DNS	Used in the TCP/IP application, Domain Name System (DNS) is a distributed database which provides the translation between domain name and the IP address. This document introduces the commands for: • Configuring the DNS Client • Configuring the DNS Proxy
IP Performance	In some network environments, you need to adjust the IP parameters to achieve best network performance. This document introduces the commands for: • Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network • Configuring TCP Attributes • Configuring ICMP to Send Error Packets

Features	Description
UDP Helper	UDP Helper functions as a relay agent that converts UDP broadcast packets into unicast packets and forwards them to a specified server. This document introduces the commands for UDP Helper configuration
URPF	Unicast Reverse Path Forwarding (URPF) protects a network against source address spoofing attacks. This document introduces the commands for URPF configuration
IPv6 Basics	Internet protocol version 6 (IPv6), also called IP next generation (IPng), was designed by the Internet Engineering Task Force (IETF) as the successor to Internet protocol version 4 (IPv4). This document introduces the commands for: • IPv6 overview • Basic IPv6 functions configuration • IPv6 NDP configuration • PMTU discovery configuration • IPv6 TCP properties configuration • ICMPv6 packet sending configuration • IPv6 DNS Client configuration
Dual Stack	A network node that supports both IPv4 and IPv6 is called a dual stack node. A dual stack node configured with an IPv4 address and an IPv6 address can have both IPv4 and IPv6 packets transmitted. This document describes: • Dual stack overview • Dual stack configuration
Tunneling	Tunneling is an encapsulation technique, which utilizes one network transport protocol to encapsulate packets of another network transport protocol and transfer them over the network. This document introduces the commands for: • IPv6 manually tunnel configuration • 6to4 tunnel configuration • ISATAP tunnel configuration
sFlow	Based on packet sampling, Sampled Flow (sFlow) is a traffic monitoring technology mainly used to collect and analyze traffic statistics. This document introduces the commands for sFlow Configuration

Table of Contents

1 IP Addressing Configuration

When assigning IP addresses to interfaces on your device, go to these sections for information you are interested in:

- IP Addressing Overview
- Configuring IP Addresses
- Displaying and Maintaining IP Addressing

IP Addressing Overview

This section covers these topics:

- IP Address Classes
- Special IP Addresses

IP Address Classes

On an IP network, a 32-bit address is used to identify a host. An example is 01010000100000001000000010000000 in binary. To make IP addresses in 32-bit form easier to read, they are written in dotted decimal notation, each being four octets in length, for example, 10.1.1.1 for the address just mentioned.

Each IP address breaks down into two parts:

- Net ID: The first several bits of the IP address defining a network, also known as class bits.
- Host-id: Identifies a host on a network.

IP addresses are divided into five classes, as shown in the following figure (in which the blue parts represent the address class).

Figure 1-1 IP address classes

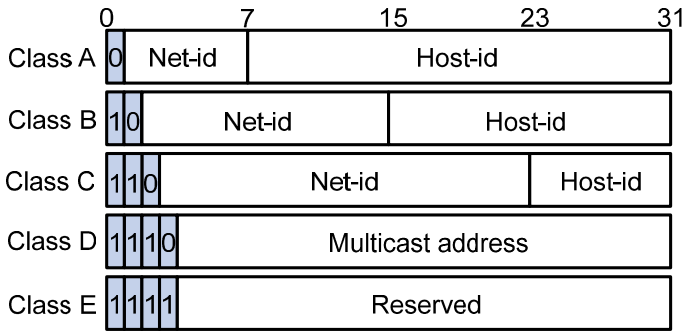


Table 1-1 describes the address ranges of these five classes.

Table 1-1 IP address classes and ranges

Class	Address range	Remarks
A	0.0.0.0 to 127.255.255.255	The IP address 0.0.0.0 is used by a host at bootstrap for temporary communication. This address is never a valid destination address. Addresses starting with 127 are reserved for loopback test. Packets destined to these addresses are processed locally as input packets rather than sent to the link.
B	128.0.0.0 to 191.255.255.255	—
C	192.0.0.0 to 223.255.255.255	—
D	224.0.0.0 to 239.255.255.255	Multicast addresses.
E	240.0.0.0 to 255.255.255.255	Reserved for future use except for the broadcast address 255.255.255.255.

Special IP Addresses

The following IP addresses are for special use, and they cannot be used as host IP addresses:

- IP address with an all-zero net ID: Identifies a host on the local network. For example, IP address 0.0.0.16 indicates the host with a host ID of 16 on the local network.
- IP address with an all-zero host ID: Identifies a network.
- IP address with an all-one host ID: Identifies a directed broadcast address. For example, a packet with the destination address of 192.168.1.255 will be broadcasted to all the hosts on the network 192.168.1.0.

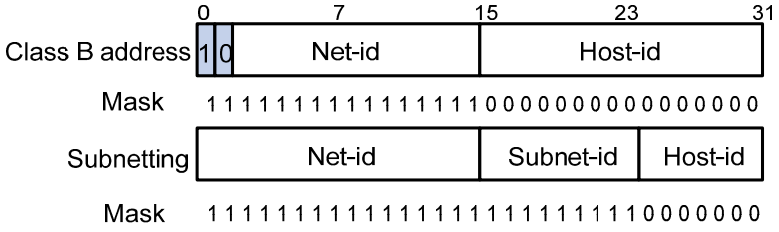
Subnetting and Masking

Subnetting was developed to address the risk of IP address exhaustion resulting from fast expansion of the Internet. The idea is to break a network down into smaller networks called subnets by using some bits of the host ID to create a subnet ID. To identify the boundary between the host ID and the combination of net ID and subnet ID, masking is used.

Each subnet mask comprises 32 bits related to the corresponding bits in an IP address. In a subnet mask, the part containing consecutive ones identifies the combination of net ID and subnet ID whereas the part containing consecutive zeros identifies the host ID.

Figure 1-2 shows how a Class B network is subnetted.

Figure 1-2 Subnet a Class B network



In the absence of subnetting, some special addresses such as the addresses with the net ID of all zeros and the addresses with the host ID of all ones, are not assignable to hosts. The same is true for subnetting. When designing your network, you should note that subnetting is somewhat a tradeoff between subnets and accommodated hosts. For example, a Class B network can accommodate 65,534 (2^{16} – 2. Of the two deducted Class B addresses, one with an all-one host ID is the broadcast address and the other with an all-zero host ID is the network address) hosts before being subnetted. After you break it down into 512 (2^9) subnets by using the first 9 bits of the host ID for the subnet, you have only 7 bits for the host ID and thus have only 126 (2^7 – 2) hosts in each subnet. The maximum number of hosts is thus 64,512 (512 × 126), 1022 less after the network is subnetted.

Class A, B, and C networks, before being subnetted, use these default masks (also called natural masks): 255.0.0.0, 255.255.0.0, and 255.255.255.0 respectively.

Configuring IP Addresses

An interface needs an IP address to communicate with other devices. You can assign an IP address to a VLAN interface or a loopback interface on a switch. Besides directly assigning an IP address to the VLAN interface, you may configure the VLAN interface to obtain one through BOOTP, or DHCP as alternatives. If you change the way an interface obtains an IP address, from manual assignment to BOOTP for example, the IP address obtained from BOOTP will overwrite the old one manually assigned.

Note

This chapter only covers how to assign an IP address manually. For the other two approaches, refer to *DHCP Configuration* in the *IP Services Volume*.

This section includes:

Assigning an IP Address to an Interface

You may assign an interface on the S4800G series switch multiple IP addresses, one primary and multiple secondaries, to connect multiple logical subnets on the same physical subnet. You can assign up to nine secondary IP addresses to an interface.

Follow these steps to assign an IP address to an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Assign an IP address to the interface	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Required No IP address is assigned by default.

Caution

- The primary IP address you assigned to the interface can overwrite the old one if there is any.
- You cannot assign secondary IP addresses to an interface that has BOOTP or DHCP configured.
- The primary and secondary IP addresses you assign to the interface can be located on the same network segment. However, this should not violate the rule that different physical interfaces on your device must reside on different network segments.

IP Addressing Configuration Example

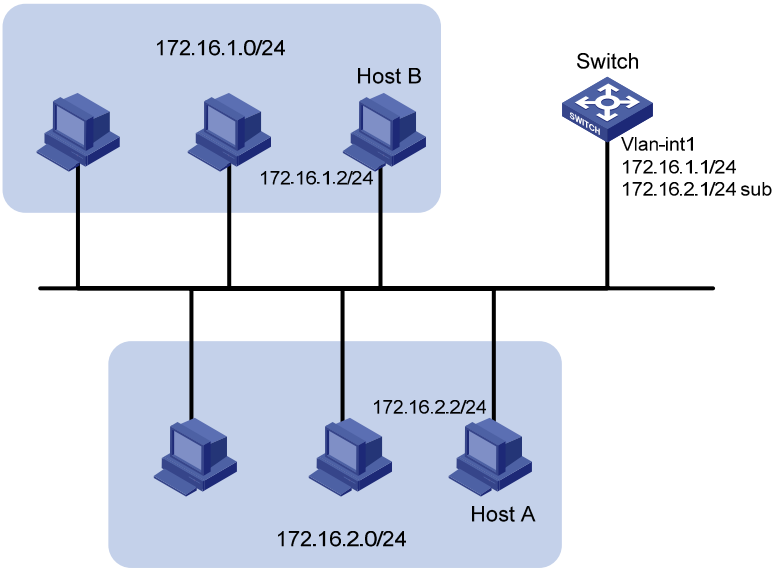
Network requirements

As shown in Figure 1-3, a port in VLAN 1 on a switch is connected to a LAN comprising two segments: 172.16.1.0/24 and 172.16.2.0/24.

To enable the hosts on the two network segments to communicate with the external network through the switch, and the hosts on the LAN can communicate with each other, do the following:

- Assign two IP addresses to VLAN-interface 1 on the switch.
- Set the switch as the gateway on all PCs in the two networks.

Figure 1-3 Network diagram for IP addressing configuration



Configuration procedure

Assign a primary IP address and a secondary IP address to VLAN-interface 1.

```
<Switch> system-view
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 172.16.1.1 255.255.255.0
[Switch-Vlan-interface1] ip address 172.16.2.1 255.255.255.0 sub
```

Set the gateway address to 172.16.1.1 on the PCs attached to subnet 172.16.1.0/24, and to 172.16.2.1 on the PCs attached to subnet 172.16.2.0/24.

Ping a host on subnet 172.16.1.0/24 from the switch to check the connectivity.

```
<Switch> ping 172.16.1.2
  PING 172.16.1.2: 56  data bytes, press CTRL_C to break
    Reply from 172.16.1.2: bytes=56 Sequence=1 ttl=255 time=25 ms
    Reply from 172.16.1.2: bytes=56 Sequence=2 ttl=255 time=27 ms
    Reply from 172.16.1.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 172.16.1.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 172.16.1.2: bytes=56 Sequence=5 ttl=255 time=26 ms

  --- 172.16.1.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 25/26/27 ms
```

The output information shows that the switch can communicate with the hosts on subnet 172.16.1.0/24.

Ping a host on subnet 172.16.2.0/24 from the switch to check the connectivity.

```
<Switch> ping 172.16.2.2
  PING 172.16.2.2: 56  data bytes, press CTRL_C to break
    Reply from 172.16.2.2: bytes=56 Sequence=1 ttl=255 time=25 ms
    Reply from 172.16.2.2: bytes=56 Sequence=2 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=3 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=4 ttl=255 time=26 ms
    Reply from 172.16.2.2: bytes=56 Sequence=5 ttl=255 time=26 ms

  --- 172.16.2.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 25/25/26 ms
```

The output information shows that the switch can communicate with the hosts on subnet 172.16.2.0/24.

Ping a host on subnet 172.16.1.0/24 from a host on subnet 172.16.2.0/24 to check the connectivity. Host B can be successfully pinged from Host A.

Displaying and Maintaining IP Addressing

To do…	Use the command…	Remarks
Display information about a specified or all Layer 3 interfaces	**display ip interface** [*interface-type interface-number*]	Available in any view
Display brief information about a specified or all Layer 3 interfaces	**display ip interface brief** [*interface-type* [*interface-number*]]	

Table of Contents

This document is organized as follows:

1 ARP Configuration

When configuring ARP, go to these sections for information you are interested in:

ARP Overview

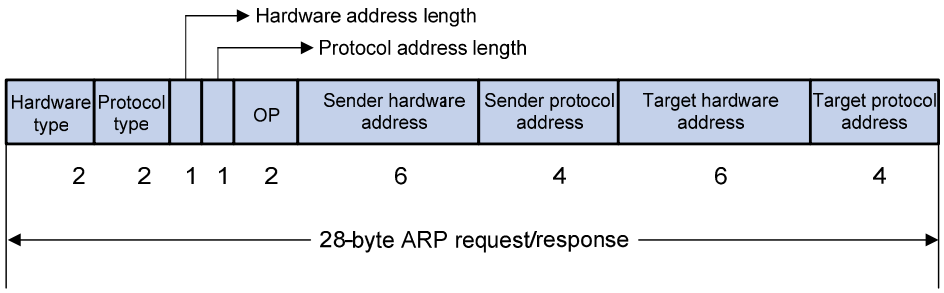
ARP Function

The Address Resolution Protocol (ARP) is used to resolve an IP address into an Ethernet MAC address (or physical address).

In a LAN, when a host or other network device is to send data to another host or device, the sending host or device must know the network layer address (that is, the IP address) of the destination host or device. Because IP datagrams must be encapsulated within Ethernet frames before they can be transmitted over physical networks, the sending host or device also needs to know the physical address of the destination host or device. Therefore, a mapping between the IP address and the physical address is needed. ARP is the protocol to implement the mapping function.

ARP Message Format

Figure 1-1 ARP message format



The following explains the fields in Figure 1-1.

- Hardware type: This field specifies the hardware address type. The value “1” represents Ethernet.
- Protocol type: This field specifies the type of the protocol address to be mapped. The hexadecimal value “0x0800” represents IP.

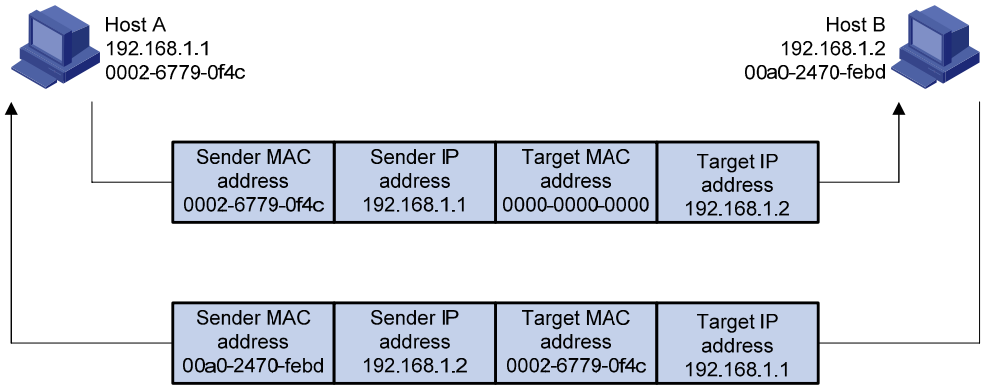
- Hardware address length and protocol address length: They respectively specify the length of a hardware address and a protocol address, in bytes. For an Ethernet address, the value of the hardware address length field is "6". For an IP(v4) address, the value of the protocol address length field is "4".
- OP: Operation code. This field specifies the type of ARP message. The value "1" represents an ARP request and "2" represents an ARP reply.
- Sender hardware address: This field specifies the hardware address of the device sending the message.
- Sender protocol address: This field specifies the protocol address of the device sending the message.
- Target hardware address: This field specifies the hardware address of the device the message is being sent to.
- Target protocol address: This field specifies the protocol address of the device the message is being sent to.

ARP Address Resolution Process

Suppose that Host A and Host B are on the same subnet and Host A sends a packet to Host B, as shown in Figure 1-2. The resolution process is as follows:

1) Host A looks into its ARP table to see whether there is an ARP entry for Host B. If yes, Host A uses the MAC address in the entry to encapsulate the IP packet into a data link layer frame and sends the frame to Host B.
2) If Host A finds no entry for Host B, Host A buffers the packet and broadcasts an ARP request, in which the sender IP address and the sender MAC address are the IP address and the MAC address of Host A respectively, and the target IP address and the target MAC address are the IP address of Host B and an all-zero MAC address respectively. Because the ARP request is a broadcast, all hosts on this subnet can receive the request, but only the requested host (namely, Host B) will respond to the request.
3) Host B compares its own IP address with the destination IP address in the ARP request. If they are the same, Host B saves the source IP address and source MAC address in its ARP table, encapsulates its MAC address into an ARP reply, and unicasts the reply to Host A.
4) After receiving the ARP reply, Host A adds the MAC address of Host B to its ARP table. Meanwhile, Host A encapsulates the IP packet and sends it out.

Figure 1-2 ARP address resolution process



If Host A is not on the same subnet with Host B, Host A first sends an ARP request to the gateway. The target IP address in the ARP request is the IP address of the gateway. After obtaining the MAC address

of the gateway from an ARP reply, Host A sends the packet to the gateway. If the gateway maintains the ARP entry of Host B, it forwards the packet to Host B directly; if not, it broadcasts an ARP request, in which the target IP address is the IP address of Host B. After obtaining the MAC address of Host B, the gateway sends the packet to Host B.

ARP Table

After obtaining the MAC address for the destination host, the device puts the IP-to-MAC mapping into its own ARP table. This mapping is used for forwarding packets with the same destination in future.

An ARP table contains ARP entries, which fall into one of two categories: dynamic or static.

Dynamic ARP entry

A dynamic entry is automatically created and maintained by ARP. It can get aged, be updated by a new ARP packet, or be overwritten by a static ARP entry. When the aging timer expires or the interface goes down, the corresponding dynamic ARP entry will be removed.

Static ARP entry

A static ARP entry is manually configured and maintained. It cannot get aged or be overwritten by a dynamic ARP entry.

Using static ARP entries enhances communication security. You can configure a static ARP entry to restrict an IP address to communicate with the specified MAC address only. After that, attack packets cannot modify the IP-to-MAC mapping specified in the static ARP entry. Thus, communications between the protected device and the specified device are ensured.

Static ARP entries can be classified into permanent or non-permanent.

- A permanent static ARP entry can be directly used to forward packets. When configuring a permanent static ARP entry, you must configure a VLAN and an outbound interface for the entry besides the IP address and the MAC address.
- A non-permanent static ARP entry has only an IP address and a MAC address configured. If a non-permanent static ARP entry matches an IP packet to be forwarded, the device sends an ARP request first. If the sender IP and MAC addresses in the received ARP reply are the same as those in the non-permanent static ARP entry, the device adds the interface receiving the ARP reply to the non-permanent static ARP entry. Then the entry can be used for forwarding IP packets.

 **Note**

Usually ARP dynamically resolves IP addresses to MAC addresses, without manual intervention.

Configuring ARP

Configuring a Static ARP Entry

A static ARP entry is effective when the device works normally. However, when a VLAN or VLAN interface to which a static ARP entry corresponds is deleted, the entry, if permanent, will be deleted, and if non-permanent and resolved, will become unresolved.

Follow these steps to configure a static ARP entry:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a permanent static ARP entry	**arp static** *ip-address mac-address vlan-id interface-type interface-number* [**vpn-instance** *vpn-instance-name*]	Required No permanent static ARP entry is configured by default.
Configure a non-permanent static ARP entry	**arp static** *ip-address mac-address* [**vpn-instance** *vpn-instance-name*]	Required No non-permanent static ARP entry is configured by default.

Caution

The *vlan-id* argument must be the ID of an existing VLAN which corresponds to the ARP entries. In addition, the Ethernet interface following the argument must belong to that VLAN. A VLAN interface must be created for the VLAN.

Configuring the Maximum Number of ARP Entries for a VLAN Interface

Follow these steps to set the maximum number of dynamic ARP entries that a VLAN interface can learn:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VLAN interface view	**interface** *interface-type interface-number*	—
Set the maximum number of dynamic ARP entries that a VLAN interface can learn	**arp max-learning-num** *number*	Optional 8192 by default.

Setting the Aging Time for Dynamic ARP Entries

To keep pace with the network changes, the ARP table is refreshed. Each dynamic ARP entry in the ARP table has a limited lifetime rather than is always valid. Dynamic ARP entries that are not refreshed before expiring are deleted from the ARP table. The lifetime is called the aging time. The aging time is reset each time the dynamic ARP entry is used within the lifetime. You can adjust the aging time for dynamic ARP entries according to the actual network condition.

Follow these steps to set the aging time for dynamic ARP entries:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the aging time for dynamic ARP entries	**arp timer aging** *aging-time*	Optional 20 minutes by default.

Enabling the ARP Entry Check

The ARP entry check function disables the device from learning multicast MAC addresses. With the ARP entry check enabled, the device cannot learn any ARP entry with a multicast MAC address, and configuring such a static ARP entry is not allowed; otherwise, the system displays error messages.

After the ARP entry check is disabled, the device can learn the ARP entry with a multicast MAC address, and you can also configure such a static ARP entry on the device.

Follow these steps to enable the ARP entry check:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the ARP entry check	**arp check enable**	Optional By default, the device is disabled from learning multicast MAC addresses.

ARP Configuration Example

Network requirements

- Enable the ARP entry check.
- Set the aging time for dynamic ARP entries to 10 minutes.
- Set the maximum number of dynamic ARP entries that VLAN-interface 10 can learn to 1,000.
- Add a static ARP entry, with the IP address being 192.168.1.1/24, the MAC address being 000f-e201-0000, and the outbound interface being GigabitEthernet 1/0/1 of VLAN 10.

Configuration procedure

```
<Sysname> system-view
[Sysname] arp check enable
[Sysname] arp timer aging 10
[Sysname] vlan 10
[Sysname-vlan10] quit
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] port access vlan 10
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface vlan-interface 10
[Sysname-vlan-interface10] arp max-learning-num 1000
[Sysname-vlan-interface10] quit
[Sysname] arp static 192.168.1.1 000f-e201-0000 10 gigabitethernet 1/0/1
```

Configuring Gratuitous ARP

Introduction to Gratuitous ARP

A gratuitous ARP packet is a special ARP packet, in which the sender IP address and the target IP address are both the IP address of the sender, the sender MAC address is the MAC address of the sender, and the target MAC address is the broadcast address ff:ff:ff:ff:ff:ff.

A device implements the following functions by sending gratuitous ARP packets:

- Determining whether its IP address is already used by another device.
- Informing other devices of its MAC address change so that they can update their ARP entries.

A device receiving a gratuitous ARP packet adds the information carried in the packet to its own dynamic ARP table if it finds no corresponding ARP entry for the ARP packet in the cache.

Configuring Gratuitous ARP

Follow these steps to configure gratuitous ARP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the device to send gratuitous ARP packets when receiving ARP requests from another network segment	**gratuitous-arp-sending enable**	Required By default, a device cannot send gratuitous ARP packets when receiving ARP requests from another network segment.
Enable the gratuitous ARP packet learning function	**gratuitous-arp-learning enable**	Optional Enabled by default.

Displaying and Maintaining ARP

To do…	Use the command…	Remarks
Display ARP entries in the ARP table	**display arp** [[**all** | **dynamic** | **static**] [**slot** *slot-number*] | **vlan** *vlan-id* | **interface** *interface-type interface-number*] [[**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*] | **count**]	Available in any view
Display the ARP entry for a specified IP address	**display arp** *ip-address* [**slot** *slot-number*] [**verbose**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]	Available in any view
Display the ARP entries for a specified VPN instance	**display arp vpn-instance** *vpn-instance-name* [**|** { **begin** | **exclude** | **include** } *regular-expression* | **count**]	Available in any view
Display the aging time for dynamic ARP entries	**display arp timer aging**	Available in any view
Clear ARP entries from the ARP table For distributed devices	**reset arp** { **all** | **dynamic** | **static** | **slot** *slot-number* | **interface** *interface-type interface-number* }	Available in user view

2 Proxy ARP Configuration

When configuring proxy ARP, go to these sections for information you are interested in:

- Proxy ARP Overview
- Enabling Proxy ARP
- Displaying and Maintaining Proxy ARP

Proxy ARP Overview

If a host sends an ARP request for the MAC address of another host that actually resides on another network (but the sending host considers the requested host is on the same network) or that is isolated from the sending host at Layer 2, the device in between must be able to respond to the request with the MAC address of the receiving interface to allow Layer 3 communication between the two hosts. This is achieved by proxy ARP. Proxy ARP hides the physical details of the network.

Proxy ARP involves common proxy ARP and local proxy ARP, which are described in the following sections.

Note

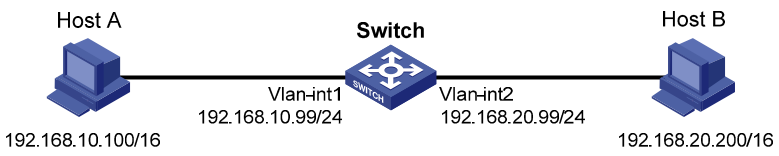
The term proxy ARP in the following sections of this chapter refers to common proxy ARP unless otherwise specified.

Proxy ARP

A proxy ARP enabled device allows hosts that reside on different subnets to communicate.

As shown in Figure 2-1, Switch connects to two subnets through VLAN-interface 1 and VLAN-interface 2. The IP addresses of the two interfaces are 192.168.10.99/24 and 192.168.20.99/24. Host A and Host B have the same prefix 192.168.0.0 assigned and connect to VLAN-interface 1 and VLAN-interface 2, respectively.

Figure 2-1 Application environment of proxy ARP



Because Host A considers that Host B is on the same network, it directly sends an ARP request for the MAC address of Host B. Host B, however, cannot receive this request because it locates in a different broadcast domain.

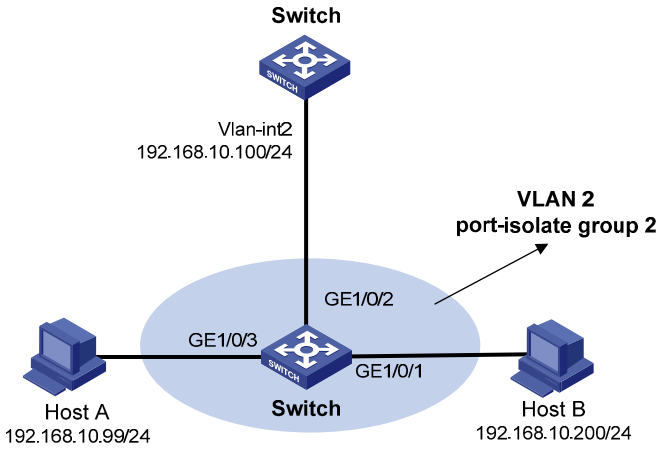
You can solve the problem by enabling proxy ARP on Switch. After that, Switch can reply to the ARP request from Host A with the MAC address of VLAN-interface 1, and forward packets sent from Host A to Host B. In this case, Switch seems to be a proxy of Host B.

A main advantage of proxy ARP is that it is added on a single router without disturbing routing tables of other routers in the network. Proxy ARP acts as the gateway for IP hosts that are not configured with a default gateway or do not have routing capability.

Local Proxy ARP

As shown in Figure 2-2, Host A and Host B belong to VLAN 2, but are isolated at Layer 2. Host A connects to GigabitEthernet 1/0/3 while Host B connects to GigabitEthernet 1/0/1. Enable local proxy ARP on Switch to allow Layer 3 communication between the two hosts.

Figure 2-2 Application environment of local proxy ARP



In one of the following cases, you need to enable local proxy ARP:

- Hosts connecting to different isolated Layer 2 ports in the same VLAN need to communicate at Layer 3.
- If an isolate-user-vlan is configured, hosts in different secondary VLANs of the isolate-user-vlan need to communicate at Layer 3.

Enabling Proxy ARP

Follow these steps to enable proxy ARP in VLAN interface view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Required
Enable proxy ARP	**proxy-arp enable**	Required Disabled by default.

Follow these steps to enable local proxy ARP in VLAN interface view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Required

To do…	Use the command…	Remarks
Enable local proxy ARP	**local-proxy-arp enable**	Required Disabled by default.

Displaying and Maintaining Proxy ARP

To do…	Use the command…	Remarks
Display whether proxy ARP is enabled	**display proxy-arp** [**interface vlan-interface** *vlan-id*]	Available in any view
Display whether local proxy ARP is enabled	**display local-proxy-arp** [**interface vlan-interface** *vlan-id*]	Available in any view

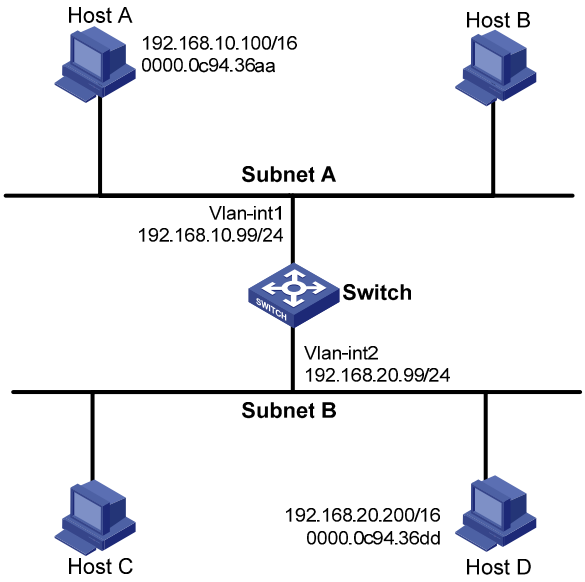
Proxy ARP Configuration Examples

Proxy ARP Configuration Example

Network requirements

Host A and Host D have the same IP prefix and mask. Host A belongs to VLAN 1; Host D belongs to VLAN 2. Configure proxy ARP on the switch to enable the communication between the two hosts.

Figure 2-3 Network diagram for proxy ARP



Configuration procedure

Configure Proxy ARP on Switch to enable the communication between Host A and Host D.

```
<Switch> system-view
[Switch] vlan 2
[Switch-vlan2] quit
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.10.99 255.255.255.0
```

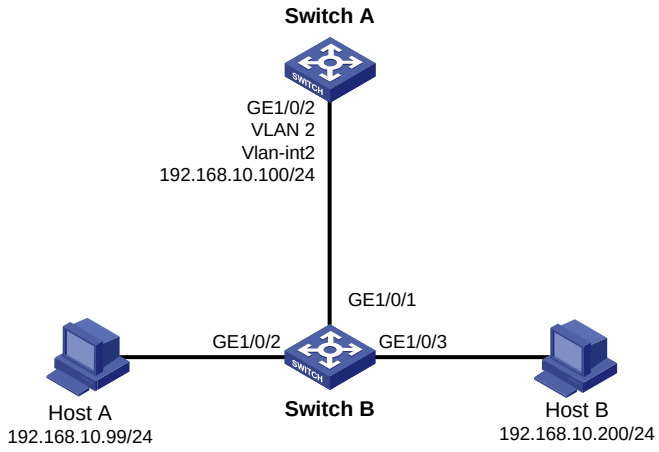
```
[Switch-Vlan-interface1] proxy-arp enable
[Switch-Vlan-interface1] quit
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.20.99 255.255.255.0
[Switch-Vlan-interface2] proxy-arp enable
[Switch-Vlan-interface2] quit
```

Local Proxy ARP Configuration Example in Case of Port Isolation

Network requirements

- Host A and Host B belong to the same VLAN, and connect to Switch B via GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3, respectively.
- Switch B connects to Switch A via GigabitEthernet 1/0/1.
- On Switch B, Layer 2 and Layer 3 port isolation are configured on GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3. Enable proxy ARP on Switch A to allow communication between Host A and Host B.

Figure 2-4 Network diagram for local proxy ARP between isolated ports



Configuration procedure

1) Configure Switch B

Add GigabitEthernet 1/0/3, GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 2. Host A and Host B are isolated and unable to exchange Layer 2 packets.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] port gigabitethernet 1/0/2
[SwitchB-vlan2] port gigabitethernet 1/0/3
[SwitchB-vlan2] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port-isolate enable
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port-isolate enable
[SwitchB-GigabitEthernet1/0/3] quit
```

2) Configure Switch A

Configure an IP address of VLAN-interface 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/2
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.10.100 255.255.0.0
```

The ping operation from Host A to Host B is unsuccessful because they are isolated at Layer 2.

Configure local proxy ARP to let Host A and Host B communicate at Layer 3.

```
[SwitchA-Vlan-interface2] local-proxy-arp enable
[SwitchA-Vlan-interface2] quit
```

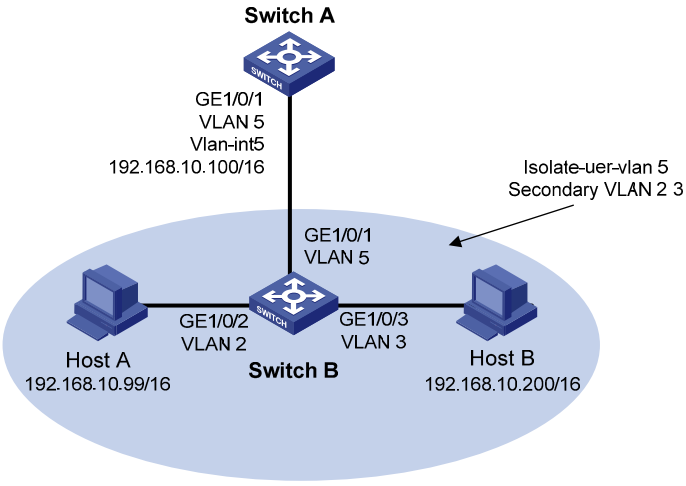
The ping operation from Host A to Host B is successful after the configuration.

Local Proxy ARP Configuration Example in Isolate-user-vlan

Network requirements

- Switch A is attached to Switch B through GigabitEthernet 1/0/1.
- VLAN 5 on Switch B is an isolate-user-vlan, which includes uplink port GigabitEthernet 1/0/1 and two secondary VLANs (VLAN 2 and VLAN 3). GigabitEthernet 1/0/2 belongs to VLAN 2, and GigabitEthernet 1/0/3 belongs to VLAN 3.
- Configure local proxy ARP on Switch A to implement Layer 3 communication between VLAN 2 and VLAN 3.

Figure 2-5 Network diagram for local proxy ARP configuration in isolate-user-vlan



Configuration procedure

1) Configure Switch B

Create VLAN 2, VLAN 3, and VLAN 5 on Switch B. Add GigabitEthernet 1/0/2 to VLAN 2, GigabitEthernet 1/0/3 to VLAN 3, and GigabitEthernet 1/0/1 to VLAN 5. Configure VLAN 5 as the isolate-user-vlan, and VLAN 2 and VLAN 3 as secondary VLANs. Configure the mappings between isolate-user-vlan and the secondary VLANs.

```
<SwitchB> system-view
[SwitchB] vlan 2
```

```
[SwitchB-vlan2] port gigabitethernet 1/0/2
[SwitchB-vlan2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/3
[SwitchB-vlan3] quit
[SwitchB] vlan 5
[SwitchB-vlan5] port gigabitethernet 1/0/1
[SwitchB-vlan5] isolate-user-vlan enable
[SwitchB-vlan5] quit
[SwitchB] isolate-user-vlan 5 secondary 2 3
```

2) Configure Switch A

Create VLAN 5 and add GigabitEthernet 1/0/1 to it.

```
<SwitchA> system-view
[SwitchA] vlan 5
[SwitchA-vlan5] port gigabitethernet 1/0/1
[SwitchA-vlan5] interface vlan-interface 5
[SwitchA-Vlan-interface5] ip address 192.168.10.100 255.255.0.0
```

The ping operation from Host A to Host B is unsuccessful because they are isolated at Layer 2.

Configure local proxy ARP to implement communication between VLAN 2 and VLAN 3.

```
[SwitchA-Vlan-interface5] local-proxy-arp enable
[SwitchA-Vlan-interface5] quit
```

The ping operation from Host A to Host B is successful after the configuration.

3 ARP Attack Defense Configuration

When configuring ARP attack defense, go to these sections for information you are interested in:

Although ARP is easy to implement, it provides no security mechanism and thus is prone to network attacks. Currently, ARP attacks and viruses are threatening LAN security. The device can provide multiple features to detect and prevent such attacks. This chapter mainly introduces these features.

Configuring ARP Source Suppression

Introduction to ARP Source Suppression

If a device receives large numbers of IP packets from a host to unreachable destinations,

- The device sends large numbers of ARP requests to the destination subnets, which increases the load of the destination subnets.
- The device continuously resolves destination IP addresses, which increases the load of the CPU.

To protect the device from such attacks, you can enable the ARP source suppression function. With the function enabled, whenever the number of packets with unresolvable destination IP addresses from a host within five seconds exceeds a specified threshold, the device suppresses the sending host from triggering any ARP requests within the following five seconds.

Configuring ARP Source Suppression

Follow these steps to configure ARP source suppression:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable ARP source suppression	**arp source-suppression enable**	Required Disabled by default.
Set the maximum number of packets with the same source IP address but unresolvable destination IP addresses that the device can receive in five consecutive seconds	**arp source-suppression limit** *limit-value*	Optional 10 by default.

Displaying and Maintaining ARP Source Suppression

To do…	Use the command…	Remarks
Display the ARP source suppression configuration information	**display arp source-suppression**	Available in any view

Configuring ARP Defense Against IP Packet Attacks

Introduction to ARP Defense Against IP Packet Attacks

When forwarding an IP packet, a device depends on ARP to resolve the MAC address of the next hop. If the address resolution is successful, the forwarding chip forwards the packet directly. Otherwise, the device runs software for further processing. If the device cannot resolve the next hops for large numbers of incoming packets, the CPU of the device will be exhausted. This is called IP packet attacks.

To protect a device against IP packet attacks, you can enable the ARP defense against IP packet attacks function. After receiving an IP packet whose next hop cannot be resolved by ARP, a device with this function enabled creates a black hole route immediately and the forwarding chip simply drops all packets matching the next hop during the age time of the black hole route.

Enabling ARP Defense Against IP Packet Attacks

The ARP defense against IP packet attack function applies to packets to be forwarded and those originated by the device.

Follow these steps to configure ARP defense against IP packet attacks:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable ARP defense against IP packet attacks	**arp resolving-route enable**	Optional Enabled by default.

Configuring ARP Active Acknowledgement

Introduction

Typically, the ARP active acknowledgement feature is configured on gateway devices to identify invalid ARP packets.

With this feature enabled, the gateway, upon receiving an ARP packet with a different source MAC address from that in the corresponding ARP entry, checks whether the ARP entry has been updated within the last minute:

- If yes, the gateway does not update the ARP entry;
- If not, the gateway unicasts an ARP request to the source MAC address of the ARP entry.

Then,

- If an ARP reply is received within five seconds, the ARP packet is ignored;
- If not, the gateway unicasts an ARP request to the MAC address of the ARP packet.

Then,

- If an ARP reply is received within five seconds, the gateway updates the ARP entry;
- If not, the ARP entry is not updated.

Configuring the ARP Active Acknowledgement Function

Follow these steps to configure ARP active acknowledgement:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the ARP active acknowledgement function	**arp anti-attack active-ack enable**	Required Disabled by default.

Configuring Source MAC Address Based ARP Attack Detection

Introduction

This feature allows the device to check the source MAC address of ARP packets. If the number of ARP packets sent from a MAC address within five seconds exceeds the specified value, the device considers this an attack.

Only the ARP packets delivered to the CPU are detected.

Configuration Procedure

Enabling source MAC address based ARP attack detection

After this feature is enabled for a device, if the number of ARP packets it receives from a MAC address within five seconds exceeds the specified value, it generates an alarm and filters out ARP packets sourced from that MAC address (in **filter** mode), or only generates an alarm (in **monitor** mode).

Follow these steps to configure source MAC address based ARP attack detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable source MAC address based ARP attack detection and specify the detection mode	**arp anti-attack source-mac** { **filter** | **monitor** }	Required Disabled by default.

Configuring protected MAC addresses

A protected MAC address is excluded from ARP attack detection even though it is an attacker. You can specify certain MAC addresses, such as that of a gateway or important servers, as protected MAC addresses.

Follow these steps to configure protected MAC addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure protected MAC addresses	**arp anti-attack source-mac exclude-mac** *mac-address*&<1-n>	Optional Not configured by default.

Configuring the aging timer for protected MAC addresses

Follow these steps to configure the aging timer for protected MAC addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure aging timer for protected MAC addresses	**arp anti-attack source-mac aging-time** *time*	Optional Five minutes by default.

Configuring the threshold

Follow these steps to configure the threshold:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the threshold	**arp anti-attack source-mac threshold** *threshold-value*	Optional 50 by default.

Displaying and Maintaining Source MAC Address Based ARP Attack Detection

To do…	Use the command…	Remarks
Display attacking entries detected	**display arp anti-attack source-mac** { **slot** *slot-number* **| interface** *interface-type interface-number* }	Available in any view

Note

A protected MAC address is no longer excluded from detection after the specified aging time expires.

Configuring ARP Packet Source MAC Address Consistency Check

Introduction

This feature enables a gateway device to filter out ARP packets with the source MAC address in the Ethernet header different from the sender MAC address in the ARP message, so that the gateway device can learn correct ARP entries.

ARP detection also checks source MAC address consistency of ARP packets, but it is enabled on an access device to detect only ARP packets sent to it.

Configuring ARP Packet Source MAC Address Consistency Check

Follow these steps to enable ARP packet source MAC address consistency check:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable ARP packet source MAC address consistency check	**arp anti-attack valid-check enable**	Required Disabled by default.

Configuring ARP Packet Rate Limit

Introduction

This feature allows you to limit the rate of ARP packets to be delivered to the CPU.

Configuring the ARP Packet Rate Limit Function

Follow these steps to configure ARP packet rate limit:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Configure ARP packet rate limit	**arp rate-limit** { **disable | rate** *pps* **drop** }	Required By default, the ARP packet rate limit is enabled and is 100 pps.

Configuring ARP Detection

Note

- For information about DHCP snooping, refer to *DHCP Configuration* in the *IP Services Volume*.
- For information about 802.1X, refer to *802.1X Configuration* in the *Security Volume*.

Introduction to ARP Detection

The ARP detection feature allows only the ARP packets of legal clients to be forwarded.

Enabling ARP Detection Based on DHCP Snooping Entries/802.1x Security Entries/Static IP-to-MAC Bindings

With this feature enabled, the device compares the source IP and MAC addresses of an ARP packet received from the VLAN against the DHCP snooping entries, 802.1X security entries, or static IP-to-MAC binding entries. You can specify a detection type or types as needed. If all the detection types are specified, the system uses DHCP snooping entries first, then 802.1X security entries, and then IP-to-MAC bindings.

1) After you enable ARP detection based on DHCP snooping entries for a VLAN,

- Upon receiving an ARP packet from an ARP untrusted port, the device compares the ARP packet against the DHCP snooping entries. If a match is found, that is, the parameters (such as IP address, MAC addresses, port index, and VLAN ID) are consistent, the ARP packet passes the check; if not, the ARP packet cannot pass the check.
- Upon receiving an ARP packet from an ARP trusted port, the device does not check the ARP packet.
- If ARP detection is not enabled for the VLAN, the ARP packet is not checked even if it is received from an ARP untrusted port.

Note

ARP detection based on DHCP snooping entries involves both dynamic DHCP snooping entries and static IP Source Guard binding entries. Dynamic DHCP snooping entries are automatically generated through the DHCP snooping function. For details, refer to *DHCP Configuration* in the *IP Service Volume*. Static IP Source Guard binding entries are created by using the **user-bind** command. For details, refer to *IP Source Guard Configuration* in the *Security Volume*.

---

2) After you enable ARP detection based on 802.1X security entries, the device, upon receiving an ARP packet from an ARP untrusted port, compares the ARP packet against the 802.1X security entries.

- If an entry with matching source IP and MAC addresses, port index, and VLAN ID is found, the ARP packet is considered valid.
- If an entry with no matching IP address but with a matching OUI MAC address is found, the ARP packet is considered valid.

Otherwise, the packet is considered invalid and discarded.

3) After you enable ARP detection based on static IP-to-MAC bindings, the device, upon receiving an ARP packet from an ARP trusted/untrusted port, compares the source IP and MAC addresses of the ARP packet against the static IP-to-MAC bindings.

- If an entry with a matching IP address but a different MAC address is found, the ARP packet is considered invalid and discarded.
- If an entry with both matching IP and MAC addresses is found, the ARP packet is considered valid and can pass the detection.
- If no match is found, the ARP packet is considered valid and can pass the detection.

Follow these steps to enable ARP detection for a VLAN and specify a trusted port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Enable ARP detection for the VLAN	**arp detection enable**	Required Disabled by default. That is, the ARP packets received on all the ports in the VLAN will not be checked.
Return to system view	**quit**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Configure the port as a trusted port	**arp detection trust**	Optional The port is an untrusted port by default.
Return to system view	**quit**	—
Specify an ARP attack detection mode	**arp detection mode** { **dhcp-snooping** | **dot1x** | **static-bind** } *	Required No ARP attack detection mode is specified by default; that is, ARP detection based on DHCP snooping entries/802.1x security entries/static IP-to-MAC bindings is not enabled by default.
Configure a static IP-to-MAC binding for ARP detection	**arp detection static-bind** *ip-address mac-address*	Optional Not configured by default. If the ARP attack detection mode is **static-bind**, you need to configure static IP-to-MAC bindings for ARP detection.

Caution

During the DHCP assignment process, when the client receives the DHCP-ACK message from the DHCP server, it broadcasts a gratuitous ARP packet to detect address conflicts. If no response is received in a pre-defined time period, the client uses the assigned IP address. If the client is enabled with ARP detection based on 802.1X security entries, the IP address is not uploaded to the 802.1X device before the client uses the IP address. As a result, the gratuitous ARP packet is considered to be an attack packet and is discarded, and thus cannot detect conflicts. After the client uploads its IP address to the 802.1X device, subsequent ARP packets sent by the client are considered to be valid and are allowed to travel through.

Configuring ARP Detection Based on Specified Objects

You can also specify objects in ARP packets to be detected. The objects involve:

- src-mac: Checks whether the sender MAC address of an ARP packet is identical to the source MAC address in the Ethernet header. If they are identical, the packet is forwarded; otherwise, the packet is discarded.

- dst-mac: Checks the target MAC address of ARP replies. If the target MAC address is all-zero, all-one, or inconsistent with the destination MAC address in the Ethernet header, the packet is considered invalid and discarded.
- ip: Checks both the source and destination IP addresses in an ARP packet. The all-zero, all-one or multicast IP addresses are considered invalid and the corresponding packets are discarded. With this object specified, the source and destination IP addresses of ARP replies, and the source IP address of ARP requests are checked.

Before performing the following configuration, make sure you have configured the **arp detection enable** command.

Follow these steps to configure ARP detection based on specified objects:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify objects for ARP detection	**arp detection validate** { **dst-mac** | **ip** | **src-mac** } *	Required Not specified by default.

Note

- If both the ARP detection based on specified objects and the ARP detection based on snooping entries/802.1X security entries/static IP-to-MAC bindings are enabled, the former one applies first, and then the latter applies.
- Before enabling ARP detection based on DHCP snooping entries, make sure that DHCP snooping is enabled.
- Before enabling ARP detection based on 802.1X security entries, make sure that 802.1X is enabled and the 802.1X clients are configured to upload IP addresses.

Displaying and Maintaining ARP Detection

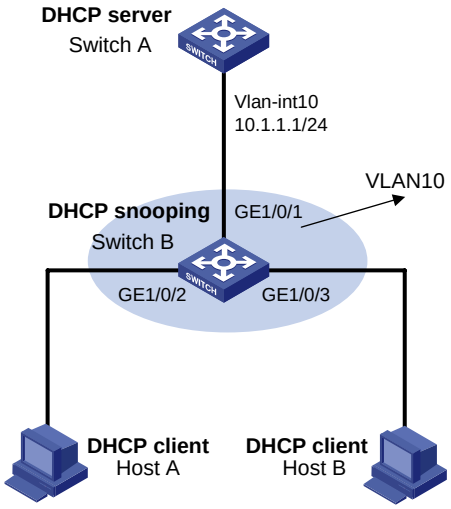
To do…	Use the command…	Remarks
Display the VLANs enabled with ARP detection	**display arp detection**	Available in any view
Display the ARP detection statistics	**display arp detection statistics** [**interface** *interface-type interface-number*]	Available in any view
Clear the ARP detection statistics	**reset arp detection statistics** [**interface** *interface-type interface-number*]	Available in user view

ARP Detection Configuration Example I

Network requirements

- Configure Switch A as a DHCP server and enable DHCP snooping on Switch B. Enable ARP detection for VLAN 10 to allow only packets from valid clients to pass.
- Configure Host A and Host B as DHCP clients.

Figure 3-1 Network diagram for ARP detection configuration



Configuration procedure

1) Add all the ports on Switch B into VLAN 10, and configure the IP address of VLAN-interface 10 on Switch A (the configuration procedure is omitted).
2) Configure Switch A as a DHCP server

Configure DHCP address pool 0.

```
<SwitchA> system-view
[SwitchA] dhcp enable
[SwitchA] dhcp server ip-pool 0
[SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[SwitchA-dhcp-pool-0] display interface vlan-interface 10
Vlan-interface10 current state: UP
Line protocol current state: UP
Description: Vlan-interface10 Interface
The Maximum Transmit Unit is 1500
Internet Address is 10.1.1.1/24 Primary
IP Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 000f-e249-8050
IPv6 Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 000f-e249-8050
    Last 300 seconds input:  0 bytes/sec 0 packets/sec
    Last 300 seconds output:  0 bytes/sec 0 packets/sec
    0 packets input, 0 bytes, 0 drops
0 packets output, 0 bytes, 0 drops
```

From the above information, you can see that the MAC address of VLAN-interface 10 is 000f-e249-8050.

3) Configure Host A and Host B as DHCP clients (the configuration procedure is omitted).
4) Configure Switch B

Enable DHCP snooping.

```
<SwitchB> system-view
[SwitchB] dhcp-snooping
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dhcp-snooping trust
```

```
[SwitchB-GigabitEthernet1/0/1] quit
```

Enable ARP detection for VLAN 10. Configure the upstream port as a trusted port and the downstream ports as untrusted ports (a port is an untrusted port by default).

```
[SwitchB] vlan 10
[SwitchB-vlan10] arp detection enable
[SwitchB-vlan10] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] arp detection trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure a static IP Source Guard binding entry on GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] user-bind ip-address 10.1.1.5 mac-address 0001-0203-0405 vlan 10
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure a static IP Source Guard binding entry on GigabitEthernet 1/0/3.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] user-bind ip-address 10.1.1.6 mac-address 0001-0203-0607 vlan 10
[SwitchB-GigabitEthernet1/0/3] quit
```

Enable ARP detection based on both DHCP snooping entries and static IP-to-MAC bindings.

```
[SwitchB] arp detection mode dhcp-snooping static-bind
[SwitchB] arp detection static-bind 10.1.1.1 000f-e249-8050
```

Enable the checking of the MAC addresses and IP addresses of ARP packets.

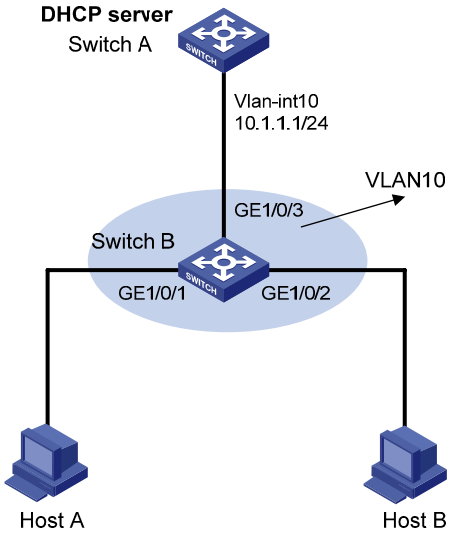
```
[SwitchB] arp detection validate dst-mac ip src-mac
```

ARP Detection Configuration Example II

Network requirements

- Configure Switch A as a DHCP server and enable 802.1X on Switch B. Enable ARP detection for VLAN 10 to allow only packets from valid clients to pass.
- Configure Host A and Host B as local 802.1X access users.

Figure 3-2 Network diagram for ARP detection configuration



Configuration procedure

1) Add all the ports on Switch B into VLAN 10, and configure the IP address of VLAN-interface 10 on Switch A (the configuration procedure is omitted).
2) Configure Switch A as a DHCP server

Configure DHCP address pool 0

```
<SwitchA> system-view
[SwitchA] dhcp enable
[SwitchA] dhcp server ip-pool 0
[SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[SwitchA-dhcp-pool-0] display interface vlan-interface 10
Vlan-interface10 current state: UP
Line protocol current state: UP
Description: Vlan-interface10 Interface
The Maximum Transmit Unit is 1500
Internet Address is 10.1.1.1/24 Primary
IP Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 000f-e249-8050
IPv6 Packet Frame Type: PKTFMT_ETHNT_2,  Hardware Address: 000f-e249-8050
    Last 300 seconds input:  0 bytes/sec 0 packets/sec
    Last 300 seconds output:  0 bytes/sec 0 packets/sec
    0 packets input, 0 bytes, 0 drops
    0 packets output, 0 bytes, 0 drops
```

From the above information, you can see that the MAC address of VLAN-interface 10 is 000f-e249-8050.

3) Configure Host A and Host B as 802.1x clients (the configuration procedure is omitted) and configure them to upload IP addresses for ARP detection.
4) Configure Switch B

Enable the 802.1x function.

```
<SwitchB> system-view
[SwitchB] dot1x
```

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dot1x
[SwitchB-GigabitEthernet1/0/1] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] dot1x
[SwitchB-GigabitEthernet1/0/2] quit
```

Add local access user **test**.

```
[SwitchB] local-user test
[SwitchB-luser-test] service-type lan-access
[SwitchB-luser-test] password simple test
[SwitchB-luser-test] quit
```

Enable ARP detection for VLAN 10. Configure the upstream port as a trusted port and the downstream ports as untrusted ports (a port is an untrusted port by default).

```
[SwitchB] vlan 10
[SwitchB-vlan10] arp detection enable
[SwitchB-vlan10] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] arp detection trust
[SwitchB-GigabitEthernet1/0/3] quit
```

Enable ARP detection based on 802.1X security entries.

```
[SwitchB] arp detection mode dot1x
```

Table of Contents

This document is organized as follows:

1 DHCP Overview

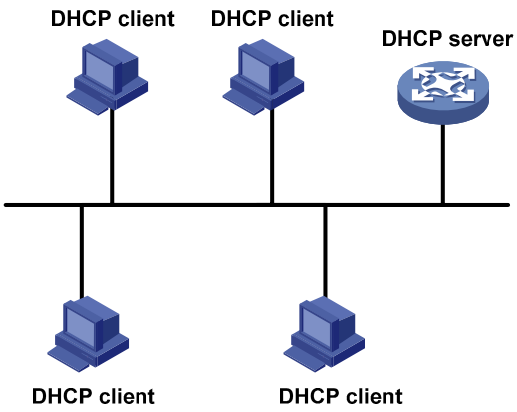
Introduction to DHCP

The fast expansion and growing complexity of networks result in scarce IP addresses assignable to hosts. Meanwhile, as many people need to take their laptops across networks, the IP addresses need to be changed accordingly. Therefore, related configurations on hosts become more complex. The Dynamic Host Configuration Protocol (DHCP) was introduced to solve these problems.

DHCP is built on a client-server model, in which a client sends a configuration request and then the server returns a reply to send configuration parameters such as an IP address to the client.

A typical DHCP application, as shown in Figure 1-1, includes a DHCP server and multiple clients (PCs and laptops).

Figure 1-1 A typical DHCP application



Note

A DHCP client can get an IP address and other configuration parameters from a DHCP server on another subnet via a DHCP relay agent. For information about the DHCP relay agent, refer to Introduction to DHCP Relay Agent.

DHCP Address Allocation

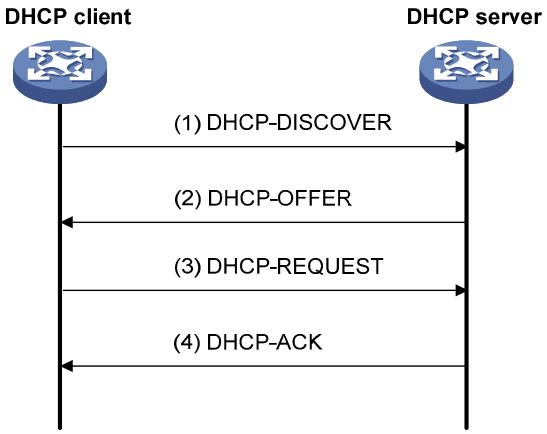
Allocation Mechanisms

DHCP supports three mechanisms for IP address allocation.

- Manual allocation: The network administrator assigns an IP address to a client like a WWW server, and DHCP conveys the assigned address to the client.
- Automatic allocation: DHCP assigns a permanent IP address to a client.
- Dynamic allocation: DHCP assigns an IP address to a client for a limited period of time, which is called a lease. Most DHCP clients obtain their addresses in this way.

Dynamic IP Address Allocation Process

Figure 1-2 Dynamic IP address allocation process



As shown in Figure 1-2, a DHCP client obtains an IP address from a DHCP server via four steps:

1) The client broadcasts a DHCP-DISCOVER message to locate a DHCP server.
2) A DHCP server offers configuration parameters including an IP address to the client in a DHCP-OFFER message. The sending mode of the DHCP-OFFER message is determined by the flag field in the DHCP-DISCOVER message. Refer to DHCP Message Format for related information.
3) If several DHCP servers send offers to the client, the client accepts the first received offer, and broadcasts it in a DHCP-REQUEST message to formally request the IP address.
4) All DHCP servers receive the DHCP-REQUEST message, but only the server from which the client accepts the offered IP address responds. The server returns a DHCP-ACK message to the client, confirming that the IP address has been allocated to the client, or a DHCP-NAK unicast message, denying the IP address allocation.

Note

- After receiving the DHCP-ACK message, the client probes whether the IP address assigned by the server is in use by broadcasting a gratuitous ARP packet. If the client receives no response within a specified time, the client can use this IP address. Otherwise, the client sends a DHCP-DECLINE message to the server and requests an IP address again.
- The IP addresses offered by other DHCP servers are still assignable to other clients.

IP Address Lease Extension

The IP address dynamically allocated by a DHCP server to a client has a lease. When the lease expires, the IP address is reclaimed by the DHCP server. If the client wants to use the IP address longer, it has to extend the lease duration.

When the half lease duration elapses, the DHCP client sends to the DHCP server a DHCP-REQUEST unicast to extend the lease duration. Upon availability of the IP address, the DHCP server returns a DHCP-ACK unicast confirming that the client's lease duration has been extended, or a DHCP-NAK unicast denying the request.

If the client receives no reply, it broadcasts another DHCP-REQUEST message for lease extension after 7/8 lease duration elapses. The DHCP server handles the request as above mentioned.

DHCP Message Format

Figure 1-3 gives the DHCP message format, which is based on the BOOTP message format and involves eight types. These types of messages have the same format except that some fields have different values. The numbers in parentheses indicate the size of each field in bytes.

Figure 1-3 DHCP message format

- op: Message type defined in option field. 1 = REQUEST, 2 = REPLY
- htype, hlen: Hardware address type and length of a DHCP client.
- hops: Number of relay agents a request message traveled.
- xid: Transaction ID, a random number chosen by the client to identify an IP address allocation.

- secs: Filled in by the client, the number of seconds elapsed since the client began address acquisition or renewal process. Currently this field is reserved and set to 0.
- flags: The leftmost bit is defined as the BROADCAST (B) flag. If this flag is set to 0, the DHCP server sent a reply back by unicast; if this flag is set to 1, the DHCP server sent a reply back by broadcast. The remaining bits of the flags field are reserved for future use.
- ciaddr: Client IP address.
- yiaddr: 'your' (client) IP address, assigned by the server.
- siaddr: Server IP address, from which the clients obtained configuration parameters.
- giaddr: IP address of the first relay agent a request message traveled.
- chaddr: Client hardware address.
- sname: Server host name, from which the client obtained configuration parameters.
- file: Bootfile name and path information, defined by the server to the client.
- options: Optional parameters field that is variable in length, which includes the message type, lease, domain name server IP address, and WINS IP address.

DHCP Options

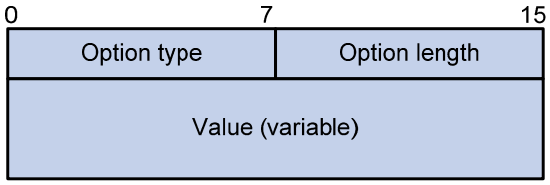
DHCP Options Overview

The DHCP message adopts the same format as the Bootstrap Protocol (BOOTP) message for compatibility, but differs from it in the option field, which identifies new features for DHCP.

DHCP uses the option field in DHCP messages to carry control information and network configuration parameters, implementing dynamic address allocation and providing more network configuration information for clients.

Figure 1-4 shows the DHCP option format.

Figure 1-4 DHCP option format



Introduction to DHCP Options

The common DHCP options are as follows:

- Option 3: Router option. It specifies the gateway address to be assigned to the client.
- Option 6: DNS server option. It specifies the DNS server IP address to be assigned to the client.
- Option 51: IP address lease option.
- Option 53: DHCP message type option. It identifies the type of the DHCP message.
- Option 55: Parameter request list option. It is used by a DHCP client to request specified configuration parameters. The option contains values that correspond to the parameters requested by the client.
- Option 66: TFTP server name option. It specifies a TFTP server to be assigned to the client.
- Option 67: Bootfile name option. It specifies the bootfile name to be assigned to the client.
- Option 150: TFTP server IP address option. It specifies the TFTP server IP address to be assigned to the client.

- Option 121: Classless route option. It specifies a list of classless static routes (the destination addresses in these static routes are classless) that the requesting client should add to its routing table.
- Option 33: Static route option. It specifies a list of classful static routes (the destination addresses in these static routes are classful) that a client should add to its routing table. If Option 121 exists, Option 33 is ignored.

For more information about DHCP options, refer to RFC 2132.

Self-Defined Options

Some options, such as Option 43, have no unified definitions in RFC 2132. The formats of some self-defined options are introduced as follows.

Vendor-specific option (Option 43)

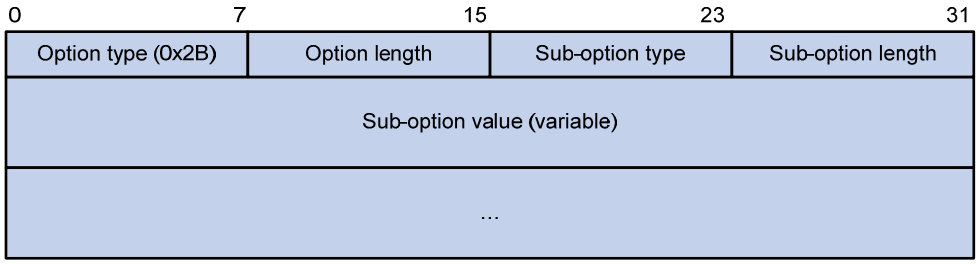
DHCP servers and clients exchange vendor-specific information through messages containing the vendor-specific option (Option 43). Upon receiving a DHCP message requesting Option 43 (in Option 55), the DHCP server returns a response message containing Option 43 to assign vendor-specific information to the DHCP client.

The DHCP client can obtain the following information through Option 43:

- Auto-Configuration Server (ACS) parameters, including the ACS URL, username, and password.
- Service provider identifier acquired by the customer premises equipment (CPE) from the DHCP server and sent to the ACS for selecting vender-specific configurations and parameters.
- Preboot Execution Environment (PXE) server address for further obtaining the bootfile or other control information from the PXE server.

1) Format of Option 43

Figure 1-5 Format of Option 43



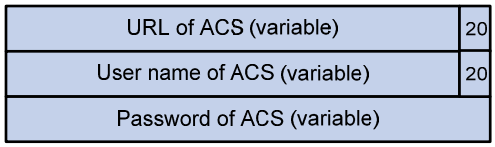
For the sake of scalability, network configuration parameters are carried in different sub-options of Option 43 so that the DHCP client can obtain more information through Option 43 as shown in Figure 1-5. The sub-option fields are described as follows:

- Sub-option type: Type of a sub-option. The field value can be 0x01, 0x02, or 0x80. 0x01 indicates an ACS parameter sub-option. 0x02 indicates a service provider identifier sub-option. 0x80 indicates a PXE server address sub-option.
- Sub-option length: Length of a sub-option excluding the sub-option type and sub-option length fields.
- Sub-option value: Value of a sub-option.

2) Format of the sub-option value field of Option 43

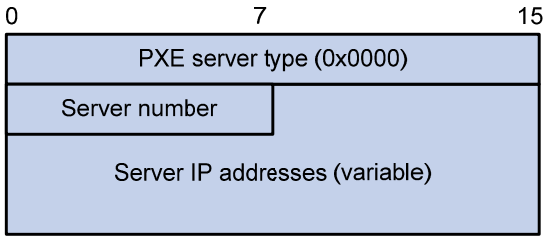
- As shown in Figure 1-6, the value field of the ACS parameter sub-option is filled in with variable ACS URL, username, and password separated with a space (0x20) in between.

Figure 1-6 Format of the value field of the ACS parameter sub-option



- The value field of the service provider identifier sub-option contains the service provider identifier.
- Figure 1-7 shows the format of the value field of the PXE server address sub-option. Currently, the value of the PXE server type can only be 0. The server number field indicates the number of PXE servers contained in the sub-option. The server IP addresses filed contains the IP addresses of the PXE servers.

Figure 1-7 Format of the value field of the PXE server address sub-option



Relay agent option (Option 82)

Option 82 is the relay agent option in the option field of the DHCP message. It records the location information of the DHCP client. When a DHCP relay agent or DHCP snooping device receives a client's request, it adds Option 82 to the request message before forwarding the message to the server.

The administrator can locate the DHCP client to further implement security control and accounting. The Option 82 supporting server can also use such information to define individual assignment policies of IP address and other parameters for the clients.

Option 82 involves at most 255 sub-options. At least one sub-option is defined. Currently the DHCP relay agent supports two sub-options: sub-option 1 (Circuit ID) and sub-option 2 (Remote ID).

Option 82 has no unified definition. Its padding formats vary with vendors.

You can use the following methods to configure Option 82:

- User-defined method: Manually specify the content of Option 82.
- Non-user-defined method: Pad Option 82 in the default normal or verbose format.

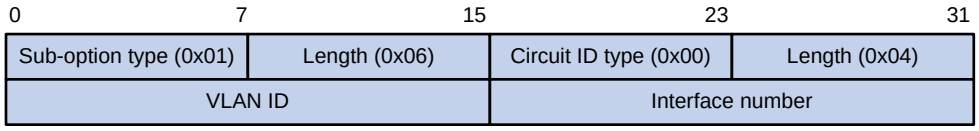
If you choose the second method, specify the code type for the sub-options as ASCII or HEX.

1) Normal padding format

The padding contents for sub-options in the normal padding format are as follows:

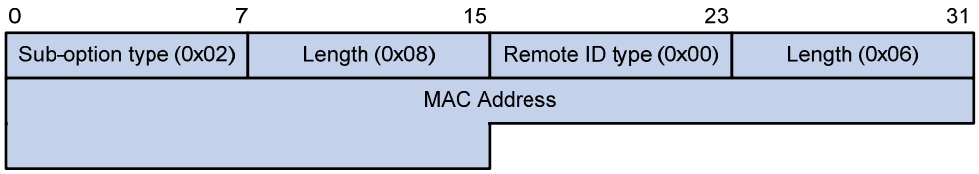
- Sub-option 1: Padded with the VLAN ID and interface number of the interface that received the client's request. The following figure gives its format. The value of the sub-option type is 1, and that of the circuit ID type is 0.

Figure 1-8 Sub-option 1 in normal padding format



- Sub-option 2: Padded with the MAC address of the DHCP relay agent interface or the MAC address of the DHCP snooping device that received the client's request. The following figure gives its format. The value of the sub-option type is 2, and that of the remote ID type is 0.

Figure 1-9 Sub-option 2 in normal padding format

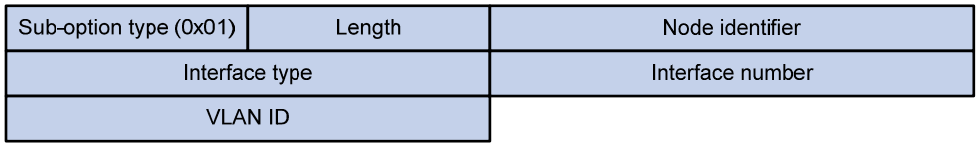


2) Verbose padding format

The padding contents for sub-options in the verbose padding format are as follows:

- Sub-option 1: Padded with the user-specified access node identifier (ID of the device that adds Option 82 in DHCP messages), and the type, number, and VLAN ID of the interface that received the client's request. Its format is shown in Figure 1-10.

Figure 1-10 Sub-option 1 in verbose padding format



Note

In Figure 1-10, except that the VLAN ID field has a fixed length of 2 bytes, all the other padding contents of sub-option 1 are length variable.

- Sub-option 2: Padded with the MAC address of the DHCP relay agent interface or the MAC address of the DHCP snooping device that received the client's request. It has the same format as that in normal padding format, as shown in Figure 1-9.

Option 184

Option 184 is a reserved option, and parameters in the option can be defined as needed. The device supports Option 184 carrying the voice related parameters, so a DHCP client with voice functions can get an IP address along with specified voice parameters from the DHCP server.

Option 184 involves the following sub-options:

- Sub-option 1: IP address of the primary network calling processor, which is a server serving as the network calling control source and providing program downloads.
- Sub-option 2: IP address of the backup network calling processor that DHCP clients will contact when the primary one is unreachable.
- Sub-option 3: Voice VLAN ID and the result whether DHCP clients take this ID as the voice VLAN or not.
- Sub-option 4: Failover route that specifies the destination IP address and the called number (SIP users use such IP addresses and numbers to communicate with each other) that a SIP user uses to reach another SIP user when both the primary and backup calling processors are unreachable.

Note

You must define the sub-option 1 to make other sub-options effective.

Protocols and Standards

- RFC 2131: Dynamic Host Configuration Protocol
- RFC 2132: DHCP Options and BOOTP Vendor Extensions
- RFC 1542: Clarifications and Extensions for the Bootstrap Protocol
- RFC 3046: DHCP Relay Agent Information Option

2 DHCP Server Configuration

When configuring the DHCP server, go to these sections for information you are interested in:

Note

- The DHCP server configuration is supported only on VLAN interfaces and loopback interfaces. The secondary IP address pool configuration is not supported on loopback interfaces.
- DHCP snooping must be disabled on the DHCP server.

Introduction to DHCP Server

Application Environment

The DHCP server is well suited to the network where:

- It is hard to implement manual configuration and centralized management.
- The hosts are more than the assignable IP addresses and it is impossible to assign a fixed IP address to each host. For example, an ISP limits the number of hosts accessing the Internet at a time, so lots of hosts need to acquire IP addresses dynamically.
- A few hosts need fixed IP addresses.

DHCP Address Pool

Address pool types

DHCP address pools can be classified into two types:

- Common address pool: Supports both static binding and dynamic allocation.
- Extended address pool: Supports dynamic allocation only.

Common address pool structure

In response to a client's request, the DHCP server selects an idle IP address from an address pool and sends it together with other parameters such as lease and DNS server address to the client.

The common address pool database is organized as a tree. The root of the tree is the address pool for natural networks, branches are address pools for subnets, and leaves are addresses statically bound to clients. For the same level address pools, a previously configured pool has a higher selection priority than a new one.

At the very beginning, subnetworks inherit network parameters and clients inherit subnetwork parameters. Therefore, common parameters, for example a DNS server address, should be configured at the highest (network or subnetwork) level of the tree.

After establishment of the inheritance relationship, the new configuration at the higher level (father) of the tree will be:

- Inherited if the lower level (child) has no such configuration, or
- Overridden if the lower level (child) has such configuration.

Note

- The extended address pool database is not organized as a tree.
- The IP address lease does not enjoy the inheritance attribute.

Principles for selecting an address pool

The DHCP server observes the following principles to select an address pool when assigning an IP address to a client:

1) If the receiving interface has an extended address pool referenced, the DHCP server will assign an IP address from this address pool. If no IP address is available in the address pool, the DHCP server will fail to assign an address to the client. For the configuration of such an address pool, refer to section Configuring Dynamic Address Allocation for an Extended Address Pool.
2) If there is an address pool where an IP address is statically bound to the MAC address or ID of the client, the DHCP server will select this address pool and assign the statically bound IP address to the client. For the configuration of this address pool, refer to section Configuring manual address allocation.
3) Otherwise, the DHCP server will select the smallest common address pool that contains the IP address of the receiving interface (if the client and the server reside on the same network segment), or the smallest common address pool that contains the IP address specified in the giaddr field of the client's request (if a DHCP relay agent is in-between). If no IP address is available in the address pool, the DHCP server will fail to assign an address to the client because it cannot assign an IP address from the father address pool to the client. For the configuration of such address pool, refer to section Configuring dynamic address allocation.

For example, two common address pools, 1.1.1.0/24 and 1.1.1.0/25, are configured on the DHCP server. If the IP address of the interface receiving DHCP requests is 1.1.1.1/25, the DHCP server will select IP addresses for clients from address pool 1.1.1.0/25. If no IP address is available in the address pool, the DHCP server will fail to assign addresses to clients. If the IP address of the interface receiving

DHCP requests is 1.1.1.130/25, the DHCP server will select IP addresses for clients from the 1.1.1.0/24 address pool.

Note

Keep the IP addresses for dynamic allocation within the subnet where the interface of the DHCP server or DHCP relay agent resides to avoid wrong IP address allocation.

IP Address Allocation Sequence

A DHCP server assigns an IP address to a client according to the following sequence:

1) The first assignable IP address found in the extended address pool referenced on the receiving interface
2) The IP address manually bound to the client's MAC address or ID
3) The IP address that was ever assigned to the client
4) The IP address designated by the Option 50 field in a DHCP-DISCOVER message
5) The first assignable IP address found in a proper common address pool
6) The IP address that was a conflict or passed its lease duration

If no IP address is assignable, the server does not respond.

Note

Option 50 is the requested IP address field in DHCP-DISCOVER messages. It is padded by the client to specify the IP address that the client wants to obtain. The contents to be padded depend on the client.

DHCP Server Configuration Task List

Complete the following tasks to configure the DHCP server:

Task	Remarks
Configuring an Address Pool for the DHCP Server	Required
Enabling DHCP	Required
Enabling the DHCP Server on an Interface	Required
Applying an Extended Address Pool on an Interface	Required by the extended address pool configuration When configuring a common address pool, ignore this task.
Configuring the DHCP Server Security Functions	Optional
Configuring the Handling Mode for Option 82	Optional

Configuring an Address Pool for the DHCP Server

Configuration Task List

Complete the following tasks to configure an address pool:

Task		Remarks
Creating a DHCP Address Pool		Required
Configuring an Address Allocation Mode for a Common Address Pool	Configuring manual address allocation	Required to configure either of the two for the common address pool configuration
	Configuring dynamic address allocation	
Configuring Dynamic Address Allocation for an Extended Address Pool		Required for the extended address pool configuration
Configuring a Domain Name Suffix for the Client		Optional
Configuring DNS Servers for the Client		
Configuring WINS Servers and NetBIOS Node Type for the Client		
Configuring the BIMS Server Information for the Client		
Configuring Gateways for the Client		
Configuring Option 184 Parameters for the Client with Voice Service		
Configuring the TFTP Server and Bootfile Name for the Client		
Configuring Self-Defined DHCP Options		

Creating a DHCP Address Pool

When creating a DHCP address pool, specify it as a common address pool or an extended address pool.

Follow these steps to create a DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a DHCP address pool and enter its view	**dhcp server ip-pool** *pool-name* [**extended**]	Required No DHCP address pool is created by default.

Note

A common address pool and an extended address pool are different in address allocation mode configuration. Configurations of other parameters (such as the domain name suffix and DNS server address) for them are the same.

Configuring an Address Allocation Mode for a Common Address Pool

Caution

You can configure either the static binding or dynamic address allocation for a common address pool as needed.

It is required to specify an address range for the dynamic address allocation. A static binding is a special address pool containing only one IP address.

Configuring manual address allocation

Some DHCP clients such as a WWW server need fixed IP addresses. You can create a static binding of a client's MAC or ID to IP address in the DHCP address pool.

When the client with the MAC address or ID requests an IP address, the DHCP server will find the IP address from the binding for the client.

A DHCP address pool now supports only one static binding, which can be a MAC-to-IP or ID-to-IP binding.

Follow these steps to configure a static binding in a common address pool:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter common address pool view		**dhcp server ip-pool** *pool-name*	—
Specify the IP address of the binding		**static-bind ip-address** *ip-address* [*mask-length* | **mask** *mask*]	Required No IP addresses are statically bound by default.
Specify the MAC address or client ID	Specify the MAC address	**static-bind mac-address** *mac-address*	Required to configure either of the two Neither is bound statically by default.
	Specify the client ID	**static-bind client-identifier** *client-identifier*	

Note

- Use the **static-bind ip-address** command together with **static-bind mac-address** or **static-bind client-identifier** to accomplish a static binding configuration.
- In a DHCP address pool, if you execute the **static-bind mac-address** command before the **static-bind client-identifier** command, the latter will overwrite the former and vice versa.
- If you use the **static-bind ip-address**, **static-bind mac-address**, or **static-bind client-identifier** command repeatedly in the DHCP address pool, the new configuration will overwrite the previous one.
- The IP address of the static binding cannot be an interface address of the DHCP server. Otherwise, an IP address conflict may occur and the bound client cannot obtain an IP address correctly.
- The ID of the static binding must be identical to the ID displayed by using the **display dhcp client verbose** command on the client. Otherwise, the client cannot obtain an IP address.
- If the interfaces on a DHCP client share the same MAC address, you need to specify the client ID, rather than MAC address, in a static binding to identify the requesting interface; otherwise, the client may fail to obtain an IP address.

Configuring dynamic address allocation

You need to specify one and only one address range using a mask for the dynamic address allocation.

To avoid address conflicts, the DHCP server excludes IP addresses used by the gateway or FTP server from dynamic allocation.

You can specify the lease duration for a DHCP address pool different from others, and a DHCP address pool can only have the same lease duration. A lease does not enjoy the inheritance attribute.

Follow these steps to configure dynamic address allocation for a common address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter common address pool view	**dhcp server ip-pool** *pool-name*	—
Specify an IP address range	**network** *network-address* [*mask-length* | **mask** *mask*]	Required Not specified by default.
Specify the address lease duration	**expired** { **day** *day* [**hour** *hour* [**minute** *minute*]] | **unlimited** }	Optional One day by default.
Return to system view	**quit**	—
Exclude IP addresses from automatic allocation	**dhcp server forbidden-ip** *low-ip-address* [*high-ip-address*]	Optional Except IP addresses of the DHCP server interfaces, all addresses in the DHCP address pool are assignable by default.

Note

- In common address pool view, using the **network** command repeatedly overwrites the previous configuration.
- After you exclude IP addresses from automatic allocation using the **dhcp server forbidden-ip** command, neither a common address pool nor an extended address pool can assign these IP addresses through dynamic address allocation.
- Using the **dhcp server forbidden-ip** command repeatedly can exclude multiple IP address ranges from allocation.

Configuring Dynamic Address Allocation for an Extended Address Pool

Extended address pools support dynamic address allocation only.

When configuring address allocation for an extended address pool, you need to specify:

- Assignable IP address range
- Mask

After the assignable IP address range and the mask are specified, the address pool is valid.

Follow these steps to configure dynamic address allocation for an extended address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter extended address pool view	**dhcp server ip-pool** *pool-name* **extended**	—
Specify the IP address range	**network ip range** *min-address max-address*	Required Not specified by default,.
Specify the IP address mask	**network mask** *mask*	Required Not specified by default.
Specify the address lease duration	**expired** { **day** *day* [**hour** *hour* [**minute** *minute*]] | **unlimited** }	Optional One day by default.
Exclude IP addresses from dynamic allocation	**forbidden-ip** *ip-address*&<1-8>	Optional Except IP addresses of the DHCP server interfaces, all addresses in the DHCP address pool are assignable by default.

Note

Excluded IP addresses specified with the **forbidden-ip** command are not assignable in the current extended address pool only, but assignable in other address pools.

Configuring a Domain Name Suffix for the Client

You can specify a domain name suffix in each DHCP address pool on the DHCP server to provide the clients with the domain name suffix. With this suffix assigned, the client only needs to input part of a domain name, and the system will add the domain name suffix for name resolution. For details about DNS, refer to *DNS Configuration* in the *IP Services Volume.*

Follow these steps to configure a domain name suffix in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify a domain name suffix for the client	**domain-name** *domain-name*	Required Not specified by default.

Configuring DNS Servers for the Client

When a DHCP client wants to access a host on the Internet via the host name, it contacts a Domain Name System (DNS) server holding host name-to-IP address mappings to get the host IP address. You can specify up to eight DNS servers in the DHCP address pool.

Follow these steps to configure DNS servers in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify DNS servers for the client	**dns-list** *ip-address*&<1-8>	Required Not specified by default.

Configuring WINS Servers and NetBIOS Node Type for the Client

A Microsoft DHCP client using NetBIOS protocol contacts a Windows Internet Naming Service (WINS) server for name resolution. Therefore, the DHCP server should assign a WINS server address when assigning an IP address to the client.

You can specify up to eight WINS servers in a DHCP address pool.

You need to specify in a DHCP address pool a NetBIOS node type for the client to approach name resolution. There are four NetBIOS node types:

- b (broadcast)-node: The b-node client sends the destination name in a broadcast message. The destination returns its IP address to the client after receiving the message.
- p (peer-to-peer)-node: The p-node client sends the destination name in a unicast message to the WINS server, and the WINS server returns the destination IP address.
- m (mixed)-node: A combination of broadcast first and peer-to-peer second. The m-node client broadcasts the destination name, if no response is received, then unicasts the destination name to the WINS server to get the destination IP address.

- h (hybrid)-node: A combination of peer-to-peer first and broadcast second. The h-node client unicasts the destination name to the WINS server, if no response is received, then broadcasts it to get the destination IP address.

Follow these steps to configure WINS servers and NetBIOS node type in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify WINS server IP addresses for the client	**nbns-list** *ip-address*&<1-8>	Required (optional for b-node) No address is specified by default.
Specify the NetBIOS node type	**netbios-type** { **b-node** | **h-node** | **m-node** | **p-node** }	Required Not specified by default.

Note

If b-node is specified for the client, you do not need to specify any WINS server address.

Configuring the BIMS Server Information for the Client

A DHCP client performs regular software update and backup using configuration files obtained from a branch intelligent management system (BIMS) server. Therefore, the DHCP server needs to offer DHCP clients the BIMS server IP address, port number, shared key from the DHCP address pool.

Follow these steps to configure the BIMS server IP address, port number, and shared key in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify the BIMS server IP address, port number, and shared key	**bims-server ip** *ip-address* [**port** *port-number*] **sharekey** *key*	Required Not specified by default.

Configuring Gateways for the Client

DHCP clients that want to access hosts outside the local subnet request gateways to forward data. You can specify gateways in each address pool for clients and the DHCP server will assign gateway addresses while assigning an IP address to the client. Up to eight gateways can be specified in a DHCP address pool.

Follow these steps to configure the gateways in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify gateways	**gateway-list** *ip-address*&<1-8>	Required No gateway is specified by default.

Configuring Option 184 Parameters for the Client with Voice Service

To assign voice calling parameters along with an IP address to DHCP clients with voice service, you need to configure Option 184 on the DHCP server. For information about Option 184, refer to Option 184.

If Option 55 in the request from a DHCP client contains Option 184, the DHCP server will return parameters specified in Option 184 to the client. The client then can initiate a call using parameters in Option 184.

Follow these steps to configure option 184 parameters in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify the IP address of the primary network calling processor	**voice-config ncp-ip** *ip-address*	Required Not specified by default.
Specify the IP address of the backup network calling processor	**voice-config as-ip** *ip-address*	Optional Not specified by default.
Configure the voice VLAN	**voice-config voice-vlan** *vlan-id* { **disable** | **enable** }	Optional Not configured by default.
Specify the failover IP address and dialer string	**voice-config fail-over** *ip-address dialer-string*	Optional No failover IP address or dialer string is specified by default.

Note

Specify an IP address for the network calling processor before performing other configuration.

Configuring the TFTP Server and Bootfile Name for the Client

This task is to specify the IP address and name of a TFTP server and the bootfile name in the DHCP address pool. The DHCP clients use these parameters to contact the TFTP server, requesting the configuration file used for system initialization, which is called auto-configuration. The request process of the client is described below:

1) When a router starts up without loading any configuration file, the system sets an active interface (such as the interface of the default VLAN) as the DHCP client to request from the DHCP server for parameters, such as an IP address and name of a TFTP server, and the bootfile name.
2) After getting related parameters, the DHCP client will send a TFTP request to obtain the configuration file from the specified TFTP server for system initialization. If the client cannot get such parameters, it will perform system initialization without loading any configuration file.

To implement auto-configuration, you need to specify the IP address or name of a TFTP server and the bootfile name in the DHCP address pool on the DHCP server, but you do not need to perform any configuration on the DHCP client.

When Option 55 in the client's request contains parameters of Option 66, Option 67, or Option 150, the DHCP server will return the IP address or name of the specified TFTP server, and bootfile name to the client.

Follow these steps to configure the IP address and name of the TFTP server and the bootfile name in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—
Specify the TFTP server	**tftp-server ip-address** *ip-address*	Required to use either command Not specified by default.
Specify the name of the TFTP server	**tftp-server domain-name** *domain-name*	
Specify the bootfile name	**bootfile-name** *bootfile-name*	Required Not specified by default.

Configuring Self-Defined DHCP Options

By configuring self-defined DHCP options, you can

- Define new DHCP options. New configuration options will come out with DHCP development. To support these new options, you can add them into the attribute list of the DHCP server.
- Define existing DHCP options. Some options have no unified definitions in RFC 2132; however, vendors can define such options as Option 43 as needed. The self-defined DHCP option enables DHCP clients to obtain vendor-specific information.
- Extend existing DHCP options. When the current DHCP options cannot meet the customers' requirements (for example, you cannot use the **dns-list** command to configure more than eight DNS server addresses), you can configure a self-defined option for extension.

Follow these steps to configure a self-defined DHCP option in the DHCP address pool:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name* [**extended**]	—

To do…	Use the command…	Remarks
Configure a self-defined DHCP option	**option** *code* { **ascii** *ascii-string* | **hex** *hex-string*&<1-16> | **ip-address** *ip-address*&<1-8> }	Required No DHCP option is configured by default.

Table 2-1 Description of common options

Option	Option name	Corresponding command	Command parameter
3	Router Option	**gateway-list**	**ip-address**
6	Domain Name Server Option	**dns-list**	**ip-address**
15	Domain Name	**domain-name**	**ascii**
44	NetBIOS over TCP/IP Name Server Option	**nbns-list**	**ip-address**
46	NetBIOS over TCP/IP Node Type Option	**netbios-type**	**hex**
66	TFTP server name	**tftp-server**	**ascii**
67	Bootfile name	**bootfile-name**	**ascii**
43	Vendor Specific Information	—	**hex**

Caution

Be cautious when configuring self-defined DHCP options because such configuration may affect the DHCP operation process.

Enabling DHCP

Enable DHCP before performing other configurations.

Follow these steps to enable DHCP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable DHCP	**dhcp enable**	Required Disabled by default.

Enabling the DHCP Server on an Interface

With the DHCP server enabled on an interface, upon receiving a client's request, the DHCP server will assign an IP address from its address pool to the DHCP client.

Follow these steps to enable the DHCP server on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the DHCP server on an interface	**dhcp select server global-pool** [**subaddress**]	Optional Enabled by default.

Note

If a DHCP relay agent exists between the DHCP server and client, the DHCP server, regardless of whether the **subaddress** keyword is used, will select an IP address from the address pool containing the primary IP address of the DHCP relay agent's interface (connected to the client) for a requesting client.

When the DHCP server and client are on the same subnet:

- With the keyword **subaddress** specified, the DHCP server will assign an IP address from the address pool containing the secondary IP address of the server interface (connected to the client); if the interface has multiple secondary IP addresses, the address pool containing the first secondary IP address is selected. If the interface has no secondary IP addresses, the server is unable to assign an IP address to the client.
- Without the keyword **subaddress** specified, the DHCP server will assign an IP address from the address pool containing the primary IP address of the server interface (connected to the client).

Applying an Extended Address Pool on an Interface

After you create an extended address pool and apply it on an interface, the DHCP server, upon receiving a client's request on the interface, will assign an IP address from this address pool to the client. If no IP address is available in this address pool, address allocation fails, and the DHCP server will not assign an IP address from other address pools.

Follow these steps to apply an extended address pool on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Apply an extended address pool on the interface	**dhcp server apply ip-pool** *pool-name*	Optional By default, the DHCP server has no extended address pool applied on its interface, and assigns an IP address from a common address pool to a requesting client.

Note

Only an extended address pool can be applied on the interface. The address pool to be referenced must already exist.

Configuring the DHCP Server Security Functions

This configuration is necessary to secure DHCP services on the DHCP server.

Configuration Prerequisites

Before performing this configuration, complete the following configurations on the DHCP server:

- Enable DHCP
- Configure the DHCP address pool

Enabling Unauthorized DHCP Server Detection

Unauthorized DHCP servers may exist on networks, and they reply DHCP clients with wrong IP addresses.

With this feature enabled, upon receiving a DHCP request, the DHCP server will record the IP address of the DHCP server which assigned an IP address to the DHCP client and the receiving interface. The administrator can use this information to check out any unauthorized DHCP servers.

Follow these steps to enable unauthorized DHCP server detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable unauthorized DHCP server detection	**dhcp server detect**	Required Disabled by default.

Note

With the unauthorized DHCP server detection enabled, the device puts a record once for each DHCP server. The administrator needs to find unauthorized DHCP servers from the log information.

Configuring IP Address Conflict Detection

To avoid IP address conflicts, the DHCP server checks whether the address to be assigned is in use by sending ping packets.

The DHCP server pings the IP address to be assigned using ICMP. If the server gets a response within the specified period, the server will select and ping another IP address; otherwise, the server will ping the IP addresses once again until the specified number of ping packets are sent. If still no response is received, the server will assign the IP address to the requesting client (The DHCP client probes the IP address by sending gratuitous ARP packets).

Follow these steps to configure IP address conflict detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify the number of ping packets	**dhcp server ping packets** *number*	Optional One ping packet by default. The value **0** indicates that no ping operation is performed.
Configure a timeout waiting for ping responses	**dhcp server ping timeout** *milliseconds*	Optional 500 ms by default. The value **0** indicates that no ping operation is performed.

Configuring the Handling Mode for Option 82

When the DHCP server receives a message with Option 82, if the server is configured to handle Option 82, it will return a response message carrying Option 82 to assign an IP address to the requesting client.

If the server is configured to ignore Option 82, it will assign an IP address to the client without adding Option 82 in the response message.

Configuration prerequisites

Before performing this configuration, complete the following configuration on the DHCP server:

- Enable DHCP
- Configure the DHCP address pool

Configuring the handling mode for Option 82

Follow these steps to enable the DHCP server to handle Option 82:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the server to handle Option 82	**dhcp server relay information enable**	Optional Enabled by default.

Note

To support Option 82, it is required to perform configuration on both the DHCP server and relay agent (or the device enabled with DHCP snooping). For related configuration details, refer to Configuring the DHCP Relay Agent to Support Option 82 and Configuring DHCP Snooping to Support Option 82.

Displaying and Maintaining the DHCP Server

To do…	Use the command…	Remarks
Display information about IP address conflicts	**display dhcp server conflict** { **all** | **ip** *ip-address* }	Available in any view
Display information about lease expiration	**display dhcp server expired** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }	
Display information about assignable IP addresses	**display dhcp server free-ip**	
Display IP addresses excluded from automatic allocation in the DHCP address pool	**display dhcp server forbidden-ip**	
Display information about bindings	**display dhcp server ip-in-use** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }	
Display information about DHCP server statistics	**display dhcp server statistics**	
Display tree organization information of address pool(s)	**display dhcp server tree** { **all** | **pool** [*pool-name*] }	
Clear information about IP address conflicts	**reset dhcp server conflict** { **all** | **ip** *ip-address* }	Available in user view
Clear information about dynamic bindings	**reset dhcp server ip-in-use** { **all** | **ip** *ip-address* | **pool** [*pool-name*] }	
Clear information about DHCP server statistics	**reset dhcp server statistics**	

Note

Using the **save** command does not save DHCP server lease information. Therefore, when the system boots up or the **reset dhcp server ip-in-use** command is executed, no lease information will be available in the configuration file. In this case, the server will deny the request for lease extension from a client and the client needs to request an IP address again.

DHCP Server Configuration Examples

DHCP networking involves two types:

- The DHCP server and client are on the same subnet and exchange messages directly.
- The DHCP server and client are not on the same subnet and they communicate with each other via a DHCP relay agent.

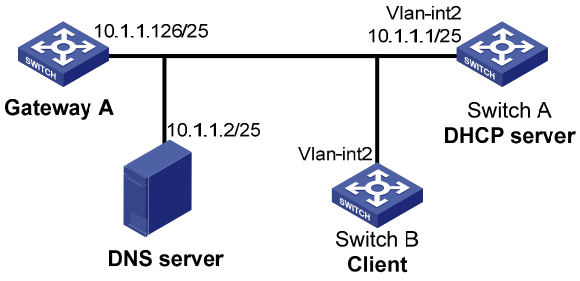
The DHCP server configuration for the two types is the same.

Static IP Address Assignment Configuration Example

Network requirements

As shown in Figure 2-1, Switch B (DHCP client) obtains a static IP address, DNS server address, and gateway address from Switch A (DHCP server).

Figure 2-1 Network diagram for static IP address assignment



Configuration procedure

1) Configure the IP address of VLAN-interface 2 on Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 10.1.1.1 25
[SwitchA-Vlan-interface2] quit
```

2) Configure the DHCP server

Enable DHCP.

```
[SwitchA] dhcp enable
```

Create DHCP address pool 0, configure a static IP-MAC binding, DNS server and gateway in it.

```
[SwitchA] dhcp server ip-pool 0
[SwitchA-dhcp-pool-0] static-bind ip-address 10.1.1.5
[SwitchA-dhcp-pool-0] static-bind mac-address 000f-e200-0002
[SwitchA-dhcp-pool-0] dns-list 10.1.1.2
[SwitchA-dhcp-pool-0] gateway-list 10.1.1.126
[SwitchA-dhcp-pool-0] quit
```

Dynamic IP Address Assignment Configuration Example

Network requirements

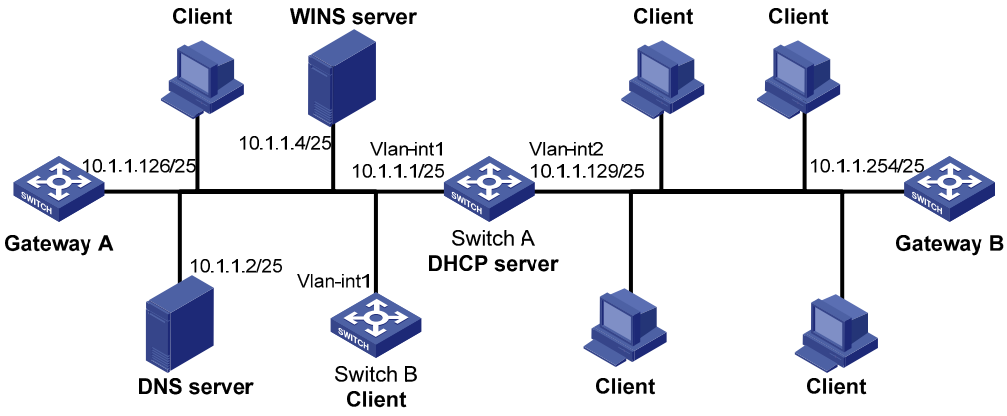
- As shown in Figure 2-2, DHCP server (Switch A) assigns IP address to clients in subnet 10.1.1.0/24, which is subnetted into 10.1.1.0/25 and 10.1.1.128/25.
- The IP addresses of VLAN-interfaces 1 and 2 on Switch A are 10.1.1.1/25 and 10.1.1.129/25 respectively.
- In address pool 10.1.1.0/25, the address lease duration is ten days and twelve hours, domain name suffix aabbcc.com, DNS server address 10.1.1.2/25, gateway 10.1.1.126/25, and WINS server 10.1.1.4/25.
- In address pool 10.1.1.128/25, the address lease duration is five days, domain name suffix aabbcc.com, DNS server address 10.1.1.2/25, and gateway address 10.1.1.254/25, and there is no WINS server address.

- The domain name and DNS server address on subnets 10.1.1.0/25 and 10.1.1.128/25 are the same. Therefore, the domain name suffix and DNS server address can be configured only for subnet 10.1.1.0/24. Subnet 10.1.1.128/25 can inherit the configuration of subnet 10.1.1.0/24.

Note

In this example, the number of requesting clients connected to VLAN-interface 1 should be less than 122, and that of clients connected to VLAN-interface 2 less than 124.

Figure 2-2 DHCP network diagram



Configuration procedure

Specify IP addresses for VLAN interfaces (omitted).

Configure the DHCP server

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Exclude IP addresses (addresses of the DNS server, WINS server and gateways).

```
[SwitchA] dhcp server forbidden-ip 10.1.1.2
[SwitchA] dhcp server forbidden-ip 10.1.1.4
[SwitchA] dhcp server forbidden-ip 10.1.1.126
[SwitchA] dhcp server forbidden-ip 10.1.1.254
```

Configure DHCP address pool 0 (address range, client domain name suffix, and DNS server address).

```
[SwitchA] dhcp server ip-pool 0
[SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[SwitchA-dhcp-pool-0] domain-name aabbcc.com
[SwitchA-dhcp-pool-0] dns-list 10.1.1.2
[SwitchA-dhcp-pool-0] quit
```

Configure DHCP address pool 1 (address range, gateway, lease duration, and WINS server).

```
[SwitchA] dhcp server ip-pool 1
```

```
[SwitchA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.128
[SwitchA-dhcp-pool-1] gateway-list 10.1.1.126
[SwitchA-dhcp-pool-1] expired day 10 hour 12
[SwitchA-dhcp-pool-1] nbns-list 10.1.1.4
[SwitchA-dhcp-pool-1] quit
```

Configure DHCP address pool 2 (address range, gateway, and lease duration).

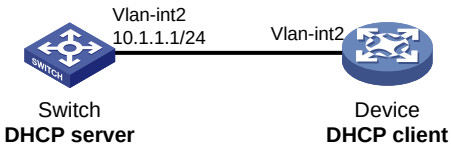
```
[SwitchA] dhcp server ip-pool 2
[SwitchA-dhcp-pool-2] network 10.1.1.128 mask 255.255.255.128
[SwitchA-dhcp-pool-2] expired day 5
[SwitchA-dhcp-pool-2] gateway-list 10.1.1.254
```

Self-Defined Option Configuration Example

Network requirements

- As shown in Figure 2-3, the DHCP client obtains an IP address and PXE server addresses from the DHCP server (Switch).
- The IP address that the DHCP client obtains belongs to network segment 10.1.1.0/24.
- The PXE server addresses that the DHCP client obtains are 1.2.3.4 and 2.2.2.2.
- The DHCP server assigns PXE server addresses to DHCP clients through Option 43, a self-defined option. The format of Option 43 and that of the PXE server address list are shown in Figure 1-5 and Figure 1-7, respectively. The value of Option 43 configured on the DHCP server in this example is 80 0B 00 00 02 01 02 03 04 02 02 02 02. The number 80 is the value of the sub-option type. The number 0B is the value of the sub-option length. The numbers 00 00 are the value of the PXE server type. The number 02 indicates the number of servers. The numbers 01 02 03 04 02 02 02 02 indicate that the PXE server addresses are 1.2.3.4 and 2.2.2.2.

Figure 2-3 Network diagram for self-defined option configuration



Configuration procedure

Specify IP addresses for the interfaces (omitted).

Configure the DHCP server

Enable DHCP.

```
<Switch> system-view
[Switch] dhcp enable
```

Configure DHCP address pool 0.

```
[Switch] dhcp server ip-pool 0
[Switch-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[Switch-dhcp-pool-0] option 43 hex 80 0B 00 00 02 01 02 03 04 02 02 02 02
```

Troubleshooting DHCP Server Configuration

Symptom

A client's IP address obtained from the DHCP server conflicts with another IP address.

Analysis

A host on the subnet may have the same IP address.

Solution

1) Disconnect the client's network cable and ping the client's IP address on another host with a long timeout time to check whether there is a host using the same IP address.
2) If a ping response is received, the IP address has been manually configured on the host. Execute the **dhcp server forbidden-ip** command on the DHCP server to exclude the IP address from dynamic allocation.
3) Connect the client's network cable. Release the IP address and obtain another one on the client. Take WINDOW XP as an example, run **cmd** to enter DOS window. Type **ipconfig/release** to relinquish the IP address and then **ipconfig/renew** to obtain another IP address.

3 DHCP Relay Agent Configuration

When configuring the DHCP relay agent, go to these sections for information you are interested in:

Note

- The DHCP relay agent configuration is supported only on VLAN interfaces.
- DHCP snooping must be disabled on the DHCP relay agent.

Introduction to DHCP Relay Agent

Application Environment

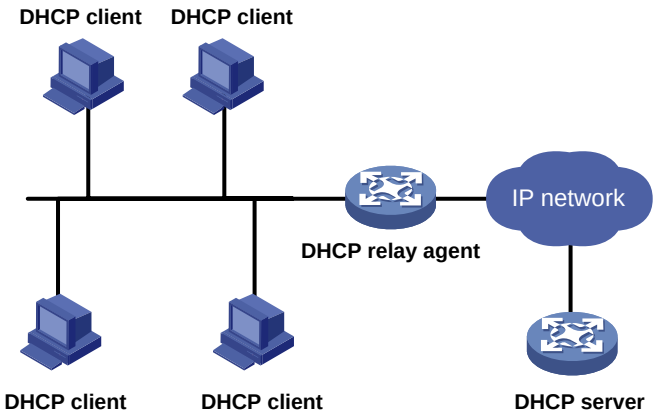
Since DHCP clients request IP addresses via broadcast messages, the DHCP server and clients must be on the same subnet. Therefore, a DHCP server must be available on each subnet, which is not practical.

DHCP relay agent solves the problem. Via a relay agent, DHCP clients communicate with a DHCP server on another subnet to obtain configuration parameters. Thus, DHCP clients on different subnets can contact the same DHCP server for ease of centralized management and cost reduction.

Fundamentals

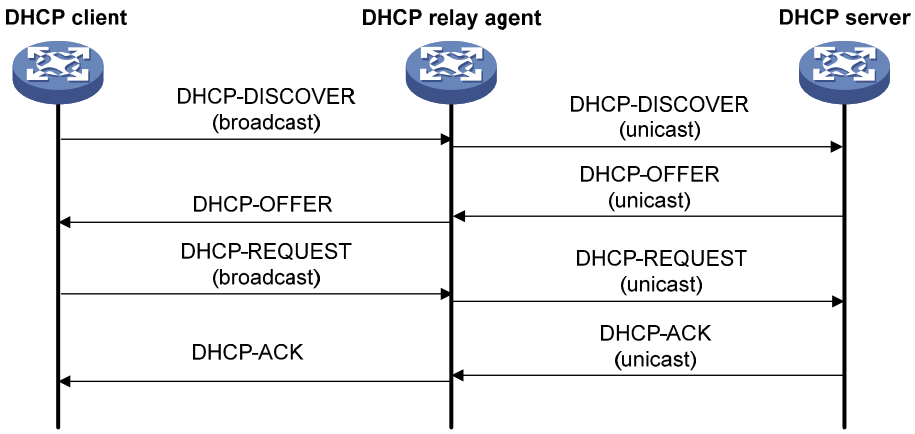
Figure 3-1 shows a typical application of the DHCP relay agent.

Figure 3-1 DHCP relay agent application



No matter whether a relay agent exists or not, the DHCP server and client interact with each other in a similar way (see section Dynamic IP Address Allocation Process). The following describes the forwarding process on the DHCP relay agent.

Figure 3-2 DHCP relay agent work process



As shown in Figure 3-2, the DHCP relay agent works as follows:

1) After receiving a DHCP-DISCOVER or DHCP-REQUEST broadcast message from a DHCP client, the DHCP relay agent fills the giaddr field of the message with its IP address and forwards the message to the designated DHCP server in unicast mode.
2) Based on the giaddr field, the DHCP server returns an IP address and other configuration parameters to the relay agent, which conveys them to the client.

DHCP Relay Agent Support for Option 82

Option 82 records the location information of the DHCP client. The administrator can locate the DHCP client to further implement security control and accounting. For more information, refer toRelay agent option (Option 82).

If the DHCP relay agent supports Option 82, it will handle a client's request according to the contents defined in Option 82, if any. The handling strategies are described in the table below.

If a reply returned by the DHCP server contains Option 82, the DHCP relay agent will remove the Option 82 before forwarding the reply to the client.

If a client's requesting message has…	Handling strategy	Padding format	The DHCP relay agent will…
Option 82	Drop	Random	Drop the message.
	Keep	Random	Forward the message without changing Option 82.
	Replace	normal	Forward the message after replacing the original Option 82 with the Option 82 padded in normal format.
		verbose	Forward the message after replacing the original Option 82 with the Option 82 padded in verbose format.
		user-defined	Forward the message after replacing the original Option 82 with the user-defined Option 82.
no Option 82	—	normal	Forward the message after adding the Option 82 padded in normal format.
	—	verbose	Forward the message after adding the Option 82 padded in verbose format.
	—	user-defined	Forward the message after adding the user-defined Option 82.

DHCP Relay Agent Configuration Task List

Complete the following tasks to configure the DHCP relay agent:

Task	Remarks
Enabling DHCP	Required
Enabling the DHCP Relay Agent on an Interface	Required
Correlating a DHCP Server Group with a Relay Agent Interface	Required
Configuring the DHCP Relay Agent Security Functions	Optional
Configuring the DHCP Relay Agent to Send a DHCP-Release Request	Optional
Configuring the DHCP Relay Agent to Support Option 82	Optional

Configuring the DHCP Relay Agent

Enabling DHCP

Enable DHCP before performing other DHCP-related configurations.

Follow these steps to enable DHCP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable DHCP	**dhcp enable**	Required Disabled by default.

Enabling the DHCP Relay Agent on an Interface

With this task completed, upon receiving a DHCP request from the enabled interface, the relay agent will forward the request to a DHCP server for address allocation.

Follow these steps to enable the DHCP relay agent on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the DHCP relay agent on the current interface	**dhcp select relay**	Required With DHCP enabled, interfaces work in the DHCP server mode.

Note

If the DHCP client obtains an IP address via the DHCP relay agent, the address pool of the subnet to which the IP address of the DHCP relay agent belongs must be configured on the DHCP server. Otherwise, the DHCP client cannot obtain a correct IP address.

Correlating a DHCP Server Group with a Relay Agent Interface

To improve reliability, you can specify several DHCP servers as a group on the DHCP relay agent and correlate a relay agent interface with the server group. When the interface receives requesting messages from clients, the relay agent will forward them to all the DHCP servers of the group.

Follow these steps to correlate a DHCP server group with a relay agent interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a DHCP server group and add a server into the group	**dhcp relay server-group** *group-id* **ip** *ip-address*	Required Not created by default.
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Correlate the DHCP server group with the current interface	**dhcp relay server-select** *group-id*	Required By default, no interface is correlated with any DHCP server group.

Note

- You can specify up to twenty DHCP server groups on the relay agent and eight DHCP server addresses for each DHCP server group.
- The IP addresses of DHCP servers and those of relay agent's interfaces cannot be on the same subnet. Otherwise, the client cannot obtain an IP address.
- A DHCP server group can correlate with one or multiple DHCP relay agent interfaces, while a relay agent interface can only correlate with one DHCP server group. Using the **dhcp relay server-select** command repeatedly overwrites the previous configuration. However, if the specified DHCP server group does not exist, the interface still uses the previous correlation.
- The *group-id* argument in the **dhcp relay server-select** command was specified by the **dhcp relay server-group** command.

Configuring the DHCP Relay Agent Security Functions

Creating static bindings and enable IP address check

The DHCP relay agent can dynamically record clients' IP-to-MAC bindings after clients get IP addresses. It also supports static bindings, which means you can manually configure IP-to-MAC bindings on the DHCP relay agent, so that users can access external network using fixed IP addresses.

For avoidance of invalid IP address configuration, you can configure the DHCP relay agent to check whether a requesting client's IP and MAC addresses match a binding (both dynamic and static bindings) on the DHCP relay agent. If not, the client cannot access outside networks via the DHCP relay agent.

Follow these steps to create a static binding and enable IP address check:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a static binding	**dhcp relay security static** *ip-address mac-address* [**interface** *interface-type interface-number*]	Optional No static binding is created by default.
Enter interface view	**interface** *interface-type interface-number*	—
Enable invalid IP address check	**dhcp relay address-check** { **disable** | **enable** }	Required Disabled by default.

Note

- The **dhcp relay address-check enable** command is independent of other commands of the DHCP relay agent. That is, the invalid address check takes effect when this command is executed, regardless of whether other commands are used.
- The **dhcp relay address-check enable** command only checks IP and MAC addresses of clients.
- You are recommended to configure IP address check on the interface enabled with the DHCP relay agent; otherwise, the valid DHCP clients may not be capable of accessing networks.
- When using the **dhcp relay security static** command to bind an interface to a static binding entry, make sure that the interface is configured as a DHCP relay agent; otherwise, address entry conflicts may occur.

Configuring dynamic binding update interval

Via the DHCP relay agent, a DHCP client sends a DHCP-RELEASE unicast message to the DHCP server to relinquish its IP address. In this case the DHCP relay agent simply conveys the message to the DHCP server, thus it does not remove the IP address from its bindings. To solve this problem, the DHCP relay agent can update dynamic bindings at a specified interval.

The DHCP relay agent uses the IP address of a client and the MAC address of the DHCP relay interface to periodically send a DHCP-REQUEST message to the DHCP server.

- If the server returns a DHCP-ACK message or does not return any message within a specified interval, which means the IP address is assignable now, the DHCP relay agent will update its bindings by aging out the binding entry of the IP address.
- If the server returns a DHCP-NAK message, which means the IP address is still in use, the relay agent will not age it out.

Follow these steps to configure dynamic binding update interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure binding update interval	**dhcp relay security tracker** { *interval* | **auto** }	Optional **auto** by default. (**auto** interval is calculated by the relay agent according to the number of bindings.)

Enabling unauthorized DHCP servers detection

There are unauthorized DHCP servers on networks, which reply DHCP clients with wrong IP addresses.

With this feature enabled, upon receiving a DHCP request, the DHCP relay agent will record the IP address of the DHCP server which assigned an IP address to the DHCP client and the receiving interface. The administrator can use this information to check out any DHCP unauthorized servers.

Follow these steps to enable unauthorized DHCP server detection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable unauthorized DHCP server detection	**dhcp relay server-detect**	Required Disabled by default.

Note

With the unauthorized DHCP server detection enabled, the device puts a record once for each DHCP server. The administrator needs to find unauthorized DHCP servers from the log information. After the information of recorded DHCP servers is cleared, the relay agent will re-record server information following this mechanism.

Configuring the DHCP Relay Agent to Send a DHCP-Release Request

This task allows you to release a client's IP address manually on the DHCP relay agent. After you configure this task, the DHCP relay agent actively sends a DHCP-RELEASE request that contains the client's IP address to be released. Upon receiving the DHCP-RELEASE request, the DHCP server then releases the IP address for the client; meanwhile, the client's IP-to-MAC binding entry is removed from the DHCP relay agent.

Follow these steps to configure the DHCP relay agent in system view to send a DHCP-RELEASE request:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the DHCP relay agent to send a DHCP-RELEASE request	**dhcp relay release ip** *client-ip*	Required

Configuring the DHCP Relay Agent to Support Option 82

Prerequisites

You need to complete the following tasks before configuring the DHCP relay agent to support Option 82.

- Enabling DHCP
- Enabling the DHCP relay agent on the specified interface
- Correlating a DHCP server group with relay agent interfaces

Configuring the DHCP relay agent to support Option 82

Follow these steps to configure the DHCP relay agent to support Option 82:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view		**interface** *interface-type interface-number*	—
Enable the relay agent to support Option 82		**dhcp relay information enable**	Required Disabled by default.
Configure the handling strategy for requesting messages containing Option 82		**dhcp relay information strategy** { **drop** | **keep** | **replace** }	Optional **replace** by default.
Configure non-user-defined Option 82	Configure the padding format for Option 82	**dhcp relay information format** { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] }	Optional **normal** by default.
	Configure the code type for the circuit ID sub-option	**dhcp relay information circuit-id format-type** { **ascii** | **hex** }	Optional By default, the code type depends on the padding format of Option 82. Each field has its own code type. The code type configuration applies to non-user-defined Option 82 only.
	Configure the code type for the remote ID sub-option	**dhcp relay information remote-id format-type** { **ascii** | **hex** }	Optional By default, the code type is **hex**. This code type configuration applies to non-user-defined Option 82 only.
Configure user-defined Option 82	Configure the padding content for the circuit ID sub-option	**dhcp relay information circuit-id string** *circuit-id*	Optional By default, the padding content depends on the padding format of Option 82.
	Configure the padding content for the remote ID sub-option	**dhcp relay information remote-id string** { *remote-id* | **sysname** }	Optional By default, the padding content depends on the padding format of Option 82.

Note

- To support Option 82, it is required to perform related configuration on both the DHCP server and relay agent. Refer to Configuring the Handling Mode for Option 82 for DHCP server configuration of this kind.
- If the handling strategy of the DHCP relay agent is configured as **replace**, you need to configure a padding format for Option 82. If the handling strategy is **keep** or **drop**, you need not configure any padding format.
- If sub-option 1 (node identifier) of Option 82 is padded with the device name (sysname) of a node, the device name must contain no spaces. Otherwise, the DHCP relay agent will drop the message.

Displaying and Maintaining DHCP Relay Agent Configuration

To do…	Use the command…	Remarks
Display information about DHCP server groups correlated to a specified or all interfaces	**display dhcp relay** { **all** | **interface** *interface-type interface-number* }	Available in any view
Display Option 82 configuration information on the DHCP relay agent	**display dhcp relay information** { **all** | **interface** *interface-type interface-number* }	
Display information about bindings of DHCP relay agents	**display dhcp relay security** [*ip-address* | **dynamic** | **static**]	
Display statistics information about bindings of DHCP relay agents	**display dhcp relay security statistics**	
Display information about the refreshing interval for entries of dynamic IP-to-MAC bindings	**display dhcp relay security tracker**	
Display information about the configuration of a specified or all DHCP server groups	**display dhcp relay server-group** { *group-id* | **all** }	
Display packet statistics on relay agent	**display dhcp relay statistics** [**server-group** { *group-id* | **all** }]	
Clear packet statistics from relay agent	**reset dhcp relay statistics** [**server-group** *group-id*]	Available in user view

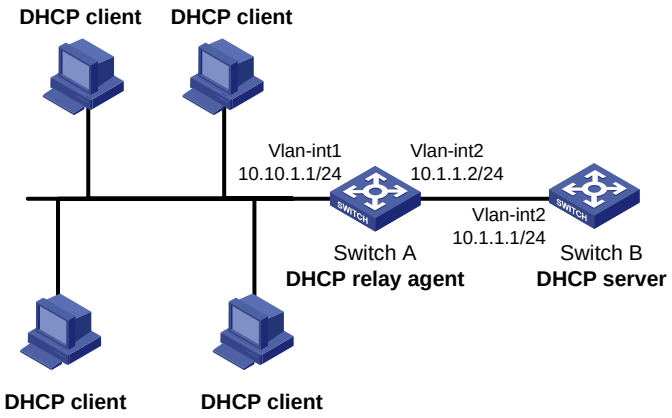
DHCP Relay Agent Configuration Examples

DHCP Relay Agent Configuration Example

Network requirements

VLAN-interface 1 on the DHCP relay agent (Switch A) connects to the network where DHCP clients reside. The IP address of VLAN-interface 1 is 10.10.1.1/24 and IP address of VLAN-interface 2 is 10.1.1.2/24 that communicates with the DHCP server 10.1.1.1/24. As shown in Figure 3-3, Switch A forwards messages between DHCP clients and the DHCP server.

Figure 3-3 Network diagram for DHCP relay agent



Configuration procedure

Specify IP addresses for the interfaces (omitted).

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Add DHCP server 10.1.1.1 into DHCP server group 1.

```
[SwitchA] dhcp relay server-group 1 ip 10.1.1.1
```

Enable the DHCP relay agent on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] dhcp select relay
```

Correlate VLAN-interface 1 to DHCP server group 1.

```
[SwitchA-Vlan-interface1] dhcp relay server-select 1
```

Note

- Performing the configuration on the DHCP server is also required to guarantee the client-server communication via the relay agent. Refer to DHCP Server Configuration Examples for DHCP server configuration information.
- Because the DHCP relay agent and server are on different subnets, you need to configure a static route or dynamic routing protocol to make them reachable to each other.

DHCP Relay Agent Option 82 Support Configuration Example

Network requirements

- As shown in Figure 3-3, Enable Option 82 on the DHCP relay agent (Switch A).
- Configure the handling strategy for DHCP requests containing Option 82 as **replace**.
- Configure the padding content for the circuit ID sub-option as **company001** and for the remote ID sub-option as **device001**.

- Switch A forwards DHCP requests to the DHCP server (Switch B) after replacing Option 82 in the requests, so that the DHCP clients can obtain IP addresses.

Configuration procedure

Specify IP addresses for the interfaces (omitted).

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Add DHCP server 10.1.1.1 into DHCP server group 1.

```
[SwitchA] dhcp relay server-group 1 ip 10.1.1.1
```

Enable the DHCP relay agent on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] dhcp select relay
```

Correlate VLAN-interface 1 to DHCP server group 1.

```
[SwitchA-Vlan-interface1] dhcp relay server-select 1
```

Enable the DHCP relay agent to support Option 82, and perform Option 82-related configurations.

```
[SwitchA-Vlan-interface1] dhcp relay information enable
[SwitchA-Vlan-interface1] dhcp relay information strategy replace
[SwitchA-Vlan-interface1] dhcp relay information circuit-id string company001
[SwitchA-Vlan-interface1] dhcp relay information remote-id string device001
```

Note

You need to perform corresponding configurations on the DHCP server to make the Option 82 configurations function normally.

Troubleshooting DHCP Relay Agent Configuration

Symptom

DHCP clients cannot obtain any configuration parameters via the DHCP relay agent.

Analysis

Some problems may occur with the DHCP relay agent or server configuration. Enable debugging and execute the **display** command on the DHCP relay agent to view the debugging information and interface state information for locating the problem.

Solution

Check that:

- The DHCP is enabled on the DHCP server and relay agent.
- The address pool on the same subnet where DHCP clients reside is available on the DHCP server.
- The routes between the DHCP server and DHCP relay agent are reachable.

- The relay agent interface connected to DHCP clients is correlated with correct DHCP server group and IP addresses for the group members are correct.

4 DHCP Client Configuration

When configuring the DHCP client, go to these sections for information you are interested in:

Note

- The DHCP client configuration is supported only on VLAN interfaces.
- When multiple VLAN interfaces with the same MAC address use DHCP for IP address acquisition via a relay agent, the DHCP server cannot be a Windows 2000 Server or Windows 2003 Server.
- You are not recommended to enable both the DHCP client and the DHCP snooping on the same device. Otherwise, DHCP snooping entries may fail to be generated, or the DHCP client may fail to obtain an IP address.

Introduction to DHCP Client

With the DHCP client enabled on an interface, the interface will use DHCP to obtain configuration parameters such as an IP address from the DHCP server.

Enabling the DHCP Client on an Interface

Follow these steps to enable the DHCP client on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the DHCP client on the interface	**ip address dhcp-alloc** [**client-identifier mac** *interface-type interface-number*]	Required Disabled by default.

Note

- An interface can be configured to acquire an IP address in multiple ways, but these ways are mutually exclusive. The latest configuration will overwrite the previous one.
- After the DHCP client is enabled on an interface, no secondary IP address is configurable for the interface.
- If the IP address assigned by the DHCP server shares a network segment with the IP addresses of other interfaces on the device, the DHCP client enabled interface will not request any IP address of the DHCP server, unless the conflicted IP address is manually deleted and the interface is made UP again by first executing the **shutdown** command and then the **undo shutdown** command or the DHCP client is enabled on the interface by executing the **undo ip address dhcp-alloc** and **ip address dhcp-alloc** commands in sequence.

Displaying and Maintaining the DHCP Client

To do…	Use the command…	Remarks
Display specified configuration information	**display dhcp client** [**verbose**] [**interface** *interface-type interface-number*]	Available in any view

DHCP Client Configuration Example

Network requirements

As shown in Figure 2-2, on a LAN, Switch B contacts the DHCP server via VLAN-interface 1 to obtain an IP address.

Configuration procedure

The following is the configuration on Switch B shown in Figure 2-2.

Enable the DHCP client on VLAN-interface 1.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address dhcp-alloc
```

Note

To implement the DHCP client-server model, you need to perform related configuration on the DHCP server. For details, refer to DHCP Server Configuration Examples.

5 DHCP Snooping Configuration

When configuring DHCP snooping, go to these sections for information you are interested in:

Note

- The DHCP snooping enabled device does not work if it is between the DHCP relay agent and DHCP server, and it can work when it is between the DHCP client and relay agent or between the DHCP client and server.
- The DHCP snooping enabled device cannot be a DHCP server or DHCP relay agent.
- You are not recommended to enable the DHCP client, BOOTP client, and DHCP snooping on the same device. Otherwise, DHCP snooping entries may fail to be generated, or the BOOTP client/DHCP client may fail to obtain an IP address.

DHCP Snooping Overview

Function of DHCP Snooping

As a DHCP security feature, DHCP snooping can implement the following:

1) Ensuring DHCP clients to obtain IP addresses from authorized DHCP servers
2) Recording IP-to-MAC mappings of DHCP clients

Ensuring DHCP clients to obtain IP addresses from authorized DHCP servers

If there is an unauthorized DHCP server on a network, the DHCP clients may obtain invalid IP addresses and network configuration parameters, and cannot normally communicate with other network devices. With DHCP snooping, the ports of a device can be configured as trusted or untrusted, ensuring the clients to obtain IP addresses from authorized DHCP servers.

- Trusted: A trusted port forwards DHCP messages normally.
- Untrusted: An untrusted port discards the DHCP-ACK or DHCP-OFFER messages from any DHCP server.

You should configure ports that connecting to authorized DHCP servers and other DHCP snooping devices as trusted, and other ports as untrusted. With such configurations, DHCP clients obtain IP addresses from authorized DHCP servers only, while unauthorized DHCP servers cannot assign IP addresses to DHCP clients.

Recording IP-to-MAC mappings of DHCP clients

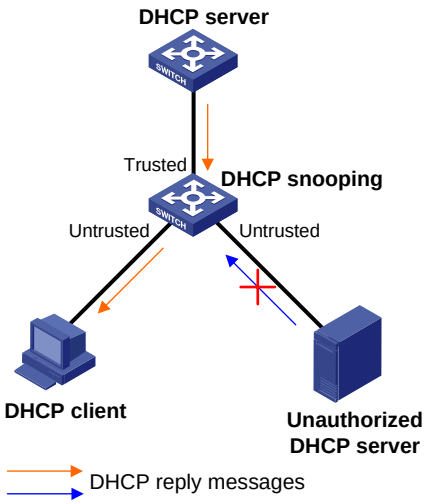
DHCP snooping reads DHCP-REQUEST messages and DHCP-ACK messages from trusted ports to record DHCP snooping entries, including MAC addresses of clients, IP addresses obtained by the clients, ports that connect to DHCP clients, and VLANs to which the ports belong. With DHCP snooping entries, DHCP snooping can implement the following:

- ARP detection: Whether ARP packets are sent from an authorized client is determined based on DHCP snooping entries. This feature prevents ARP attacks from unauthorized clients. For details, refer to *ARP Configuration* in the *IP Services Volume*.
- IP Source Guard: IP Source Guard uses dynamic binding entries generated by DHCP snooping to filter packets on a per-port basis, and thus prevents unauthorized packets from traveling through. For details, refer to *IP Source Guard Configuration* in the *Security Volume*.
- VLAN mapping: The device replaces service provider VLANs (SVLANs) in packets with customer VLANs (CVLANs) by searching corresponding DHCP snooping entries for DHCP client information including IP addresses, MAC addresses, and CVLANs, when sending the packets to clients. For details, refer to *VLAN Mapping Configuration* in the *Access Volume*.

Application Environment of Trusted Ports

Configuring a trusted port connected to a DHCP server

Figure 5-1 Configure trusted and untrusted ports



As shown in Figure 5-1, a DHCP snooping device's port that is connected to an authorized DHCP server should be configured as a trusted port to forward reply messages from the DHCP server, so that the DHCP client is guaranteed to obtain IP addresses from the authorized DHCP server.

Configuring trusted ports in a cascaded network

In a cascaded network involving multiple DHCP snooping devices, the ports connected to other DHCP snooping devices should be configured as trusted ports.

To save system resources, you can disable the trusted ports, which are indirectly connected to DHCP clients, from recording clients' IP-to-MAC bindings upon receiving DHCP requests.

Figure 5-2 Configure trusted ports in a cascaded network

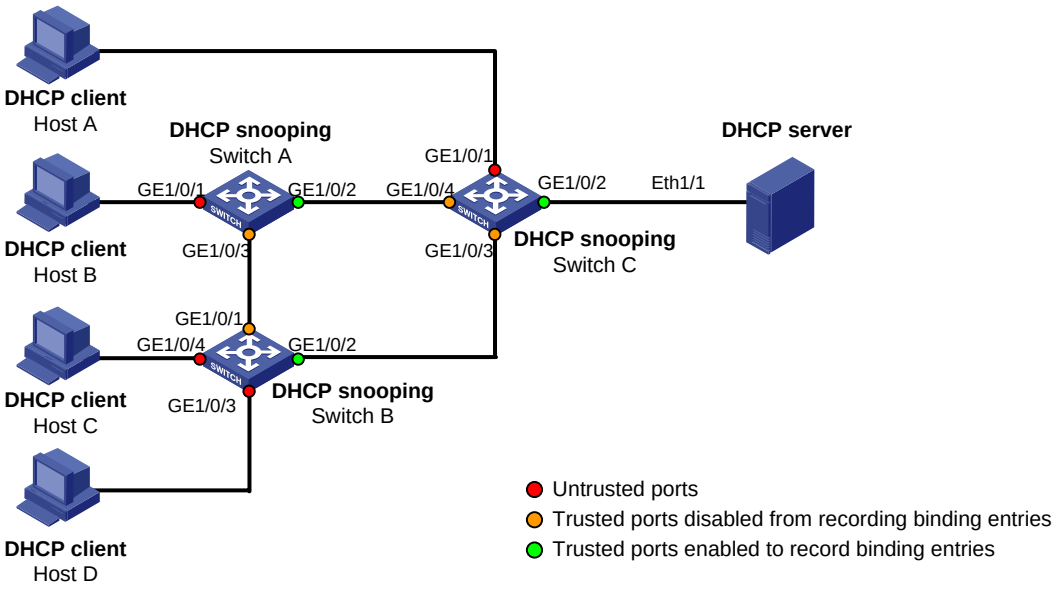


Table 5-1 describes roles of the ports shown in Figure 5-2.

Table 5-1 Roles of ports

Device	Untrusted port	Trusted port disabled from recording binding entries	Trusted port enabled to record binding entries
Switch A	GE1/0/1	GE1/0/3	GE1/0/2
Switch B	GE1/0/3 and GE1/0/4	GE1/0/1	GE1/0/2
Switch C	GE1/0/1	GE1/0/3 and Ethernet 1/4	GE1/0/2

DHCP Snooping Support for Option 82

Option 82 records the location information of the DHCP client. The administrator can locate the DHCP client to further implement security control and accounting. For more information, refer to Relay agent option (Option 82).

If DHCP snooping supports Option 82, it will handle a client's request according to the contents defined in Option 82, if any. The handling strategies are described in the table below.

If a reply returned by the DHCP server contains Option 82, the DHCP snooping device will remove the Option 82 before forwarding the reply to the client. If the reply contains no Option 82, the DHCP snooping device forwards it directly.

If a client's requesting message has…	Handling strategy	Padding format	The DHCP snooping device will…
Option 82	Drop	Random	Drop the message.
	Keep	Random	Forward the message without changing Option 82.
	Replace	normal	Forward the message after replacing the original Option 82 with the Option 82 padded in normal format.
		verbose	Forward the message after replacing the original Option 82 with the Option 82 padded in verbose format.
		user-defined	Forward the message after replacing the original Option 82 with the user-defined Option 82.
no Option 82	—	normal	Forward the message after adding the Option 82 padded in normal format.
	—	verbose	Forward the message after adding the Option 82 padded in verbose format.
	—	user-defined	Forward the message after adding the user-defined Option 82.

Note

The handling strategy and padding format for Option 82 on the DHCP snooping device are the same as those on the relay agent.

Configuring DHCP Snooping Basic Functions

Follow these steps to configure DHCP snooping basic functions:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable DHCP snooping	**dhcp-snooping**	Required Disabled by default.
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Specify the port as trusted	**dhcp-snooping trust** [**no-user-binding**]	Required Untrusted by default.

Note

- You need to specify the ports connected to the valid DHCP servers as trusted to ensure that DHCP clients can obtain valid IP addresses. The trusted port and the port connected to the DHCP client must be in the same VLAN.
- You can specify Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces as trusted ports. For details about aggregate interfaces, refer to *Link Aggregation Configuration* in the *Access Volume*.
- If a Layer 2 Ethernet interface is added to an aggregation group, DHCP snooping configured on the interface will not take effect. After the interface quits from the aggregation group, DHCP snooping will be effective.
- Do not add an untrusted Layer 2 Ethernet interface to an aggregation group.
- Configuring both the DHCP snooping and selective QinQ function on the switch is not recommended because it may result in malfunctioning of DHCP snooping.

Configuring DHCP Snooping to Support Option 82

Prerequisites

You need to enable the DHCP snooping function before configuring DHCP snooping to support Option 82.

Configuring DHCP Snooping to Support Option 82

Follow these steps to configure DHCP snooping to support Option 82:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable DHCP snooping to support Option 82	**dhcp-snooping information enable**	Required Disabled by default.
Configure the handling strategy for requesting messages containing Option 82	**dhcp-snooping information strategy** { **drop** | **keep** | **replace** }	Optional **replace** by default.

To do…		Use the command…	Remarks
Configure non-user-defined Option 82	Configure the padding format for Option 82	**dhcp-snooping information format** { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] }	Optional **normal** by default.
	Configure the code type for the circuit ID sub-option	**dhcp-snooping information circuit-id format-type** { **ascii** | **hex** }	Optional By default, the code type depends on the padding format of Option 82. Each field has its own code type. This code type configuration applies to non-user-defined Option 82 only.
	Configure the code type for the remote ID sub-option	**dhcp-snooping information remote-id format-type** { **ascii** | **hex** }	Optional **hex** by default. The code type configuration applies to non-user-defined Option 82 only.
Configure user-defined Option 82	Configure the padding content for the circuit ID sub-option	**dhcp-snooping information** [**vlan** *vlan-id*] **circuit-id string** *circuit-id*	Optional By default, the padding content depends on the padding format of Option 82.
	Configure the padding content for the remote ID sub-option	**dhcp-snooping information** [**vlan** *vlan-id*] **remote-id string** { *remote-id* | **sysname** }	Optional By default, the padding content depends on the padding format of Option 82.

Note

- You can enable DHCP snooping to support Option 82 on Layer 2 Ethernet interfaces only.
- To support Option 82, it is required to perform related configuration on both the DHCP server and the device enabled with DHCP snooping. Refer to Configuring the Handling Mode for Option 82 for DHCP server configuration of this kind.
- If the handling strategy of the DHCP-snooping-enabled device is configured as **replace**, you need to configure a padding format for Option 82. If the handling strategy is **keep** or **drop**, you need not configure any padding format.
- If the Option 82 is padded with the device name (sysname) of a node, the device name must contain no spaces. Otherwise, the DHCP-snooping-enabled device will drop the message.

Displaying and Maintaining DHCP Snooping

To do…	Use the command…	Remarks
Display DHCP snooping entries	**display dhcp-snooping** [**ip** *ip-address*]	Available in any view
Display Option 82 configuration information on the DHCP snooping device	**display dhcp-snooping information** { **all** | **interface** *interface-type interface-number* }	
Display DHCP packet statistics on the DHCP snooping device	**display dhcp-snooping packet statistics** [**slot** *slot-number*]	
Display information about trusted ports	**display dhcp-snooping trust**	
Clear DHCP snooping entries	**reset dhcp-snooping** { **all** | **ip** *ip-address* }	Available in user view
Clear DHCP packet statistics on the DHCP snooping device	**reset dhcp-snooping packet statistics** [**slot** *slot-number*]	

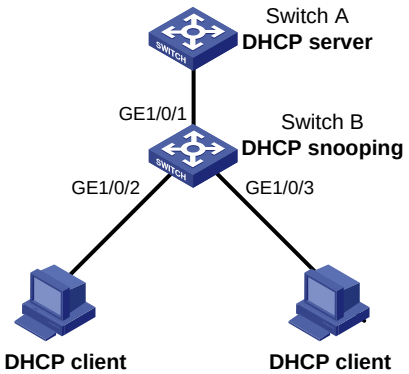
DHCP Snooping Configuration Examples

DHCP Snooping Configuration Example

Network requirements

- As shown in Figure 5-3, Switch B is connected to a DHCP server through GigabitEthernet 1/0/1, and to two DHCP clients through GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.
- GigabitEthernet 1/0/1 forwards DHCP server responses while the other two do not.
- Switch B records clients' IP-to-MAC address bindings in DHCP-REQUEST messages and DHCP-ACK messages received from trusted ports.

Figure 5-3 Network diagram for DHCP snooping configuration



Configuration procedure

Enable DHCP snooping.

```
<SwitchB> system-view
[SwitchB] dhcp-snooping
```

Specify GigabitEthernet 1/0/1 as trusted.

```
[SwitchB] interface gigabitethernet 1/0/1
```

```
[SwitchB-GigabitEthernet1/0/1] dhcp-snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

DHCP Snooping Option 82 Support Configuration Example

Network requirements

- As shown in Figure 5-3, enable DHCP snooping and Option 82 support on Switch B.
- Configure the handling strategy for DHCP requests containing Option 82 as **replace**.
- On GigabitEthernet 1/0/2, configure the padding content for the circuit ID sub-option as **company001** and for the remote ID sub-option as **device001**.
- On GigabitEthernet 1/0/3, configure the padding format as **verbose**, access node identifier as **sysname**, and code type as **ascii** for Option 82.
- Switch B forwards DHCP requests to the DHCP server (Switch A) after replacing Option 82 in the requests, so that the DHCP clients can obtain IP addresses.

Configuration procedure

Enable DHCP snooping.

```
<SwitchB> system-view
[SwitchB] dhcp-snooping
```

Specify GigabitEthernet 1/0/1 as trusted.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dhcp-snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure GigabitEthernet 1/0/2 to support Option 82.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information enable
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information strategy replace
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information circuit-id string company001
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information remote-id string device001
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure GigabitEthernet 1/0/3 to support Option 82.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information enable
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information strategy replace
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information format verbose node-identifier
sysname
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information circuit-id format-type ascii
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information remote-id format-type ascii
```

6 BOOTP Client Configuration

While configuring a BOOTP client, go to these sections for information you are interested in:

- Introduction to BOOTP Client
- Configuring an Interface to Dynamically Obtain an IP Address Through BOOTP
- Displaying and Maintaining BOOTP Client Configuration

Note

- BOOTP client configuration only applies to VLAN interfaces.
- If several VLAN interfaces sharing the same MAC address obtain IP addresses through a BOOTP relay agent, the BOOTP server cannot be a Windows 2000 Server or Windows 2003 Server.
- You are not recommended to enable both the DHCP client and DHCP snooping on the same device. Otherwise, DHCP snooping entries may fail to be generated, or the BOOTP client may fail to obtain an IP address.

Introduction to BOOTP Client

This section covers these topics:

- BOOTP Application
- Obtaining an IP Address Dynamically
- Protocols and Standards

BOOTP Application

After you specify an interface of a device as a BOOTP client, the interface can use BOOTP to get information (such as IP address) from the BOOTP server, which simplifies your configuration.

Before using BOOTP, an administrator needs to configure a BOOTP parameter file for each BOOTP client on the BOOTP server. The parameter file contains information such as MAC address and IP address of a BOOTP client. When a BOOTP client originates a request to the BOOTP server, the BOOTP server will search for the BOOTP parameter file and return the corresponding configuration information.

Because you need to configure a parameter file for each client on the BOOTP server, BOOTP usually runs under a relatively stable environment. If the network changes frequently, DHCP is more suitable.

Note

Because a DHCP server can interact with a BOOTP client, you can use the DHCP server to configure an IP address for the BOOTP client, without any BOOTP server.

Obtaining an IP Address Dynamically

Note

A DHCP server can take the place of the BOOTP server in the following dynamic IP address acquisition.

A BOOTP client dynamically obtains an IP address from a BOOTP server in the following steps:

1) The BOOTP client broadcasts a BOOTP request, which contains its own MAC address.
2) The BOOTP server receives the request and searches the configuration file for the corresponding IP address and other information according to the MAC address of the BOOTP client. The BOOTP server then returns a BOOTP response to the BOOTP client.
3) The BOOTP client obtains the IP address from the received response.

Protocols and Standards

Some protocols and standards related to BOOTP include:

- RFC 951: Bootstrap Protocol (BOOTP)
- RFC 2132: DHCP Options and BOOTP Vendor Extensions
- RFC 1542: Clarifications and Extensions for the Bootstrap Protocol

Configuring an Interface to Dynamically Obtain an IP Address Through BOOTP

Follow these steps to configure an interface to dynamically obtain an IP address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an interface to dynamically obtain IP address through BOOTP	**ip address bootp-alloc**	Required By default, an interface does not use BOOTP to obtain an IP address.

Displaying and Maintaining BOOTP Client Configuration

To do…	Use the command…	Remarks
Display related information on a BOOTP client	**display bootp client** [**interface** *interface-type interface-number*]	Available in any view

BOOTP Client Configuration Example

Network requirement

As shown in Figure 2-2, Switch B's port belonging to VLAN 1 is connected to the LAN. VLAN-interface 1 obtains an IP address from the DHCP server by using BOOTP.

Configuration procedure

The following describes only the configuration on Switch B serving as a client.

Configure VLAN-interface 1 to dynamically obtain an IP address from the DHCP server.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address bootp-alloc
```

Note

To make the BOOTP client obtain an IP address from the DHCP server, you need to perform additional configurations on the DHCP server. For details, refer to DHCP Server Configuration Examples.

Table of Contents

1 DNS Configuration

When configuring DNS, go to these sections for information you are interested in:

Note

This document only covers IPv4 DNS configuration. For information about IPv6 DNS configuration, refer to *IPv6 Basics Configuration* in the *IP Services Volume*.

DNS Overview

Domain Name System (DNS) is a distributed database used by TCP/IP applications to translate domain names into corresponding IP addresses. With DNS, you can use easy-to-remember domain names in some applications and let the DNS server translate them into correct IP addresses.

There are two types of DNS services, static and dynamic. After a user specifies a name, the device checks the local static name resolution table for an IP address. If no IP address is available, it contacts the DNS server for dynamic name resolution, which takes more time than static name resolution. Therefore, some frequently queried name-to-IP address mappings are stored in the local static name resolution table to improve efficiency.

Static Domain Name Resolution

The static domain name resolution means setting up mappings between domain names and IP addresses. IP addresses of the corresponding domain names can be found in the static domain resolution table when you use applications such as Telnet.

Dynamic Domain Name Resolution

Resolving procedure

Dynamic domain name resolution is implemented by querying the DNS server. The resolution procedure is as follows:

1) A user program sends a name query to the resolver of the DNS client.
2) The DNS resolver looks up the local domain name cache for a match. If a match is found, it sends the corresponding IP address back. If not, it sends a query to the DNS server.

3) The DNS server looks up the corresponding IP address of the domain name in its DNS database. If no match is found, it sends a query to a higher level DNS server. This process continues until a result, whether successful or not, is returned.
4) The DNS client returns the resolution result to the application after receiving a response from the DNS server.

Figure 1-1 Dynamic domain name resolution

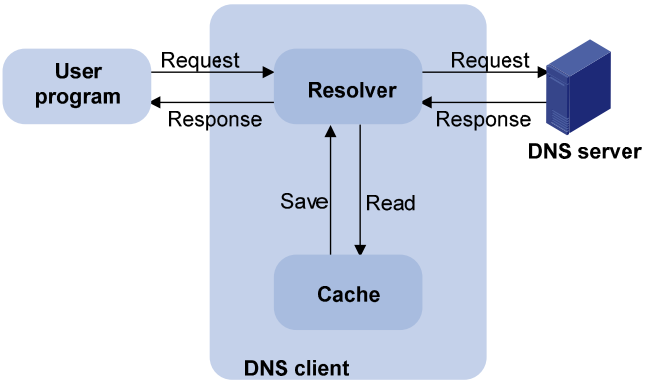


Figure 1-1 shows the relationship between the user program, DNS client, and DNS server.

The resolver and cache comprise the DNS client. The user program and DNS client can run on the same device or different devices, while the DNS server and the DNS client usually run on different devices.

Dynamic domain name resolution allows the DNS client to store latest mappings between domain names and IP addresses in the dynamic domain name cache. There is no need to send a request to the DNS server for a repeated query next time. The aged mappings are removed from the cache after some time, and latest entries are required from the DNS server. The DNS server decides how long a mapping is valid, and the DNS client gets the aging information from DNS messages.

DNS suffixes

The DNS client normally holds a list of suffixes which can be defined by users. It is used when the name to be resolved is incomplete. The resolver can supply the missing part. For example, a user can configure com as the suffix for aabbcc.com. The user only needs to type aabbcc to get the IP address of aabbcc.com. The resolver can add the suffix and delimiter before passing the name to the DNS server.

- If there is no dot in the domain name (for example, aabbcc), the resolver will consider this a host name and add a DNS suffix before query. If no match is found after all the configured suffixes are used respectively, the original domain name (for example, aabbcc) is used for query.
- If there is a dot in the domain name (for example, www.aabbcc), the resolver will directly use this domain name for query. If the query fails, the resolver adds a DNS suffix for another query.
- If the dot is at the end of the domain name (for example, aabbcc.com.), the resolver will consider it a fully qualified domain name (FQDN) and return the query result, successful or failed. Hence, the dot "." at the end of the domain name is called the terminating symbol.

Currently, the device supports static and dynamic DNS services.

Note

If an alias is configured for a domain name on the DNS server, the device can resolve the alias into the IP address of the host.

DNS Proxy

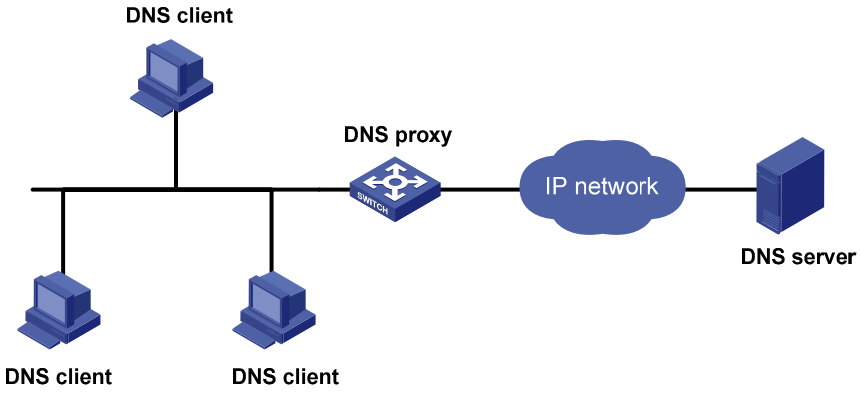
Introduction to DNS proxy

A DNS proxy forwards DNS requests and replies between DNS clients and a DNS server.

As shown in Figure 1-2, a DNS client sends a DNS request to the DNS proxy, which forwards the request to the designated DNS server, and conveys the reply from the DNS server to the client.

The DNS proxy simplifies network management. When the DNS server address is changed, you only need to change the configuration on the DNS proxy instead of on each DNS client.

Figure 1-2 DNS proxy networking application



Operation of a DNS proxy

1) A DNS client considers the DNS proxy as the DNS server, and sends a DNS request to the DNS proxy, that is, the destination address of the request is the IP address of the DNS proxy.
2) The DNS proxy searches the local static domain name resolution table after receiving the request. If the requested information exists in the table, the DNS proxy returns a DNS reply to the client.
3) If the requested information does not exist in the static domain name resolution table, the DNS proxy sends the request to the designated DNS server for domain name resolution.
4) After receiving a reply from the DNS server, the DNS proxy forwards the reply to the DNS client.

Configuring the DNS Client

Configuring Static Domain Name Resolution

Follow these steps to configure static domain name resolution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a mapping between a host name and IP address in the static name resolution table	**ip host** *hostname ip-address*	Required Not configured by default.

Note

The IP address you last assign to the host name will overwrite the previous one if there is any.

You may create up to 50 static mappings between domain names and IP addresses.

Configuring Dynamic Domain Name Resolution

Follow these steps to configure dynamic domain name resolution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable dynamic domain name resolution	**dns resolve**	Required Disabled by default.
Specify a DNS server	**dns server** *ip-address*	Required Not specified by default.
Configure a domain name suffix	**dns domain** *domain-name*	Optional Not configured by default, that is, only the provided domain name is resolved.

Note

You may configure up to six DNS servers and ten DNS suffixes.

Configuring the DNS Proxy

Follow these steps to configure the DNS proxy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable DNS proxy	**dns proxy enable**	Required Disabled by default.

Displaying and Maintaining DNS

To do…	Use the command…	Remarks
Display the static domain name resolution table	**display ip host**	Available in any view
Display DNS server information	**display dns server** [**dynamic**]	
Display domain name suffixes	**display dns domain** [**dynamic**]	Available in any view
Display the information of the dynamic domain name cache	**display dns dynamic-host**	
Clear the information of the dynamic domain name cache	**reset dns dynamic-host**	Available in user view

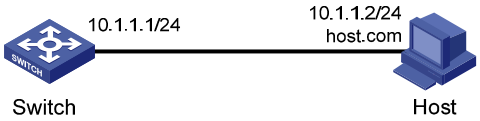
DNS Configuration Examples

Static Domain Name Resolution Configuration Example

Network requirements

Switch uses the static domain name resolution to access Host with IP address 10.1.1.2 through domain name host.com.

Figure 1-3 Network diagram for static domain name resolution



Configuration procedure

Configure a mapping between host name host.com and IP address 10.1.1.2.

```
<Sysname> system-view
[Sysname] ip host host.com 10.1.1.2
```

Execute the **ping host.com** command to verify that the Switch can use the static domain name resolution to get the IP address 10.1.1.2 corresponding to host.com.

```
[Sysname] ping host.com
  PING host.com (10.1.1.2):
```

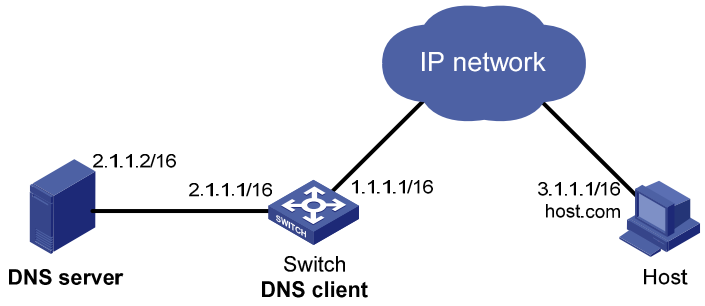
```
56  data bytes, press CTRL_C to break
  Reply from 10.1.1.2: bytes=56 Sequence=1 ttl=128 time=1 ms
  Reply from 10.1.1.2: bytes=56 Sequence=2 ttl=128 time=4 ms
  Reply from 10.1.1.2: bytes=56 Sequence=3 ttl=128 time=3 ms
  Reply from 10.1.1.2: bytes=56 Sequence=4 ttl=128 time=2 ms
  Reply from 10.1.1.2: bytes=56 Sequence=5 ttl=128 time=3 ms

--- host.com ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/2/4 ms
```

Dynamic Domain Name Resolution Configuration Example

Network requirements

- The IP address of the DNS server is 2.1.1.2/16 and the name suffix is com. The mapping between domain name Host and IP address 3.1.1.1/16 is stored in the **com** domain.
- Switch serves as a DNS client, and uses the dynamic domain name resolution and the suffix to access the host with the domain name host.com and the IP address 3.1.1.1/16.

Figure 1-4 Network diagram for dynamic domain name resolution



Configuration procedure

Note

- Before performing the following configuration, make sure that there is a route between the Switch and the host, and the IP addresses of the interfaces are configured as shown Figure 1-4.
- This configuration may vary with different DNS servers. The following configuration is performed on a Windows server 2000.

---

1) Configure the DNS server

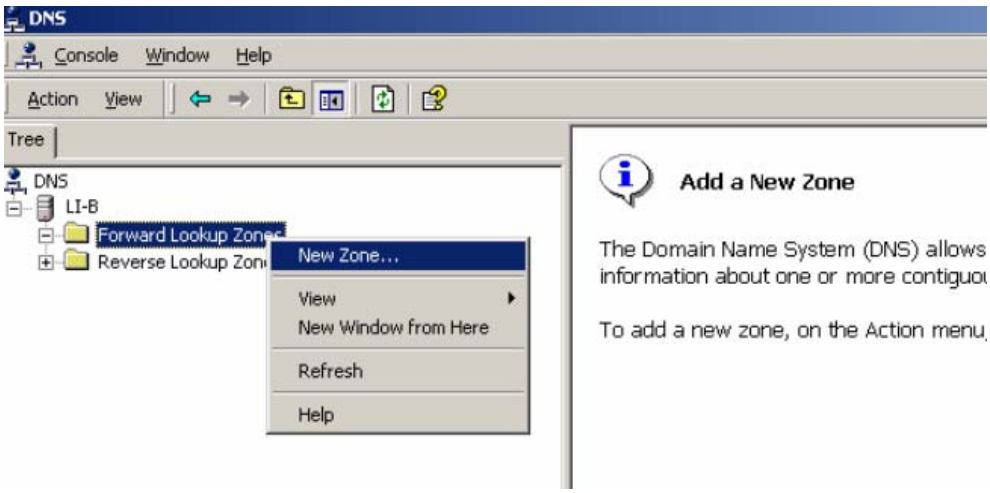
Enter DNS server configuration page.

Select **Start** > **Programs** > **Administrative Tools** > **DNS**.

Create zone com.

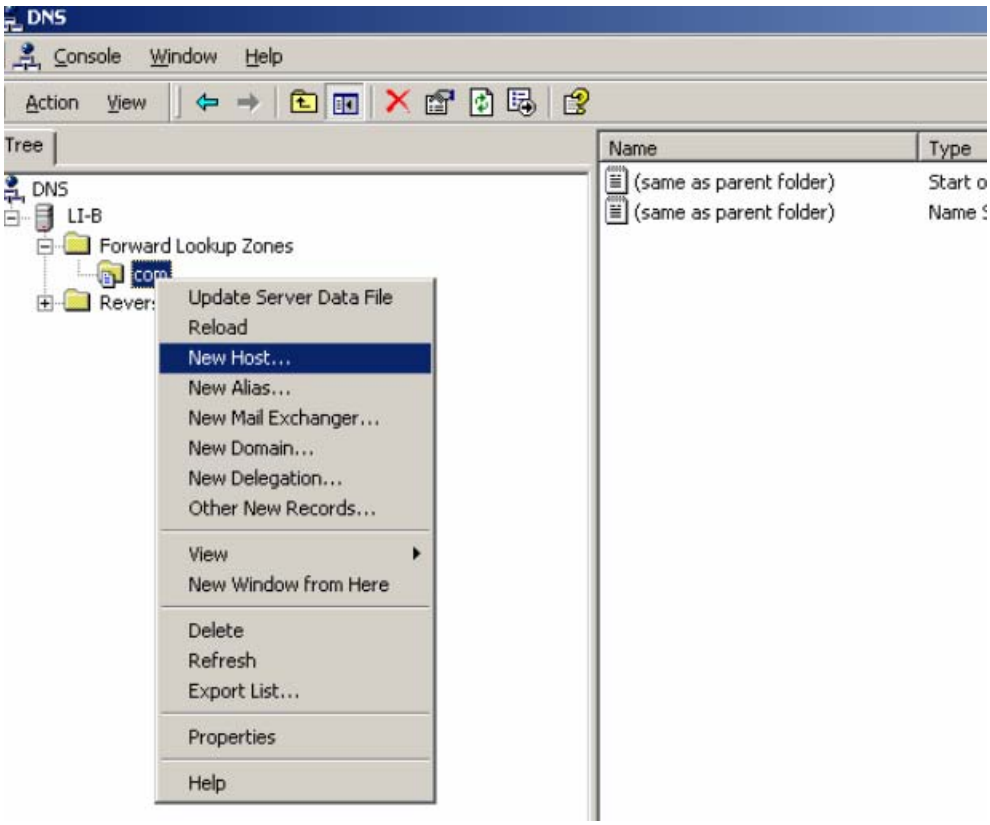
In Figure 1-5, right click **Forward Lookup Zones**, select **New zone**, and then follow the instructions to create a new zone named **com**.

Figure 1-5 Create a zone



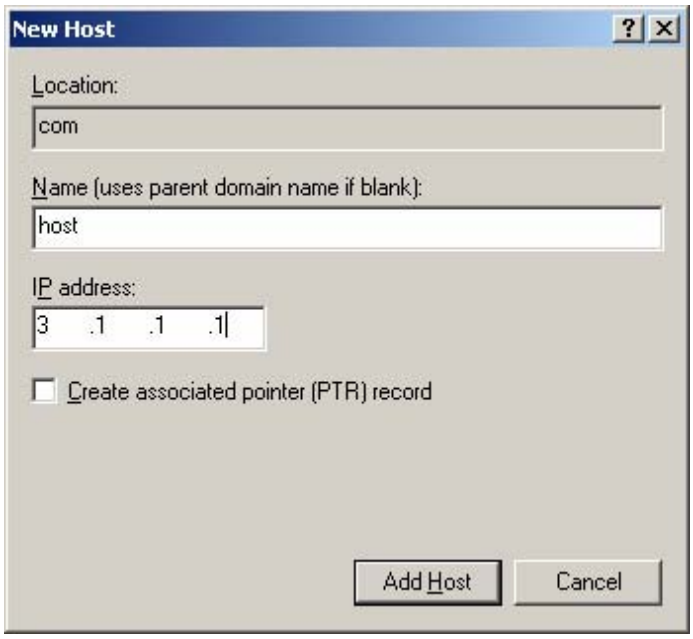
Create a mapping between the host name and IP address.

Figure 1-6 Add a host



In Figure 1-6, right click zone **com**, and then select **New Host** to bring up a dialog box as shown in Figure 1-7. Enter host name host and IP address 3.1.1.1.

Figure 1-7 Add a mapping between domain name and IP address



2) Configure the DNS client

Enable dynamic domain name resolution.

```
<Sysname> system-view
[Sysname] dns resolve
```

Specify the DNS server 2.1.1.2.

```
[Sysname] dns server 2.1.1.2
```

Configure com as the name suffix.

```
[Sysname] dns domain com
```

3) Configuration verification

Execute the **ping host** command on the Switch to verify that the communication between the Switch and the host is normal and that the corresponding destination IP address is 3.1.1.1.

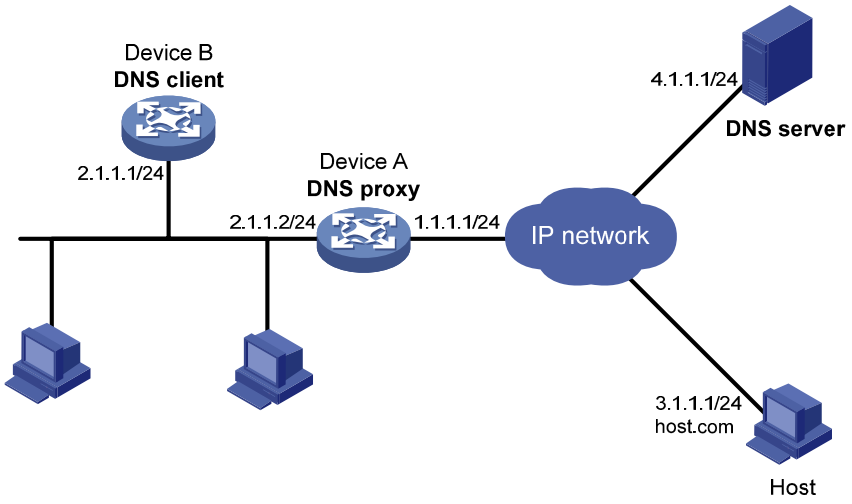
```
[Sysname] ping host
 Trying DNS resolve, press CTRL_C to break
 Trying DNS server (2.1.1.2)
  PING host.com (3.1.1.1):
  56  data bytes, press CTRL_C to break
    Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=126 time=3 ms
    Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=126 time=1 ms

   --- host.com ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/3 ms
```

DNS Proxy Configuration Example

Network requirements

- Specify Switch A as the DNS server of Switch B (the DNS client).
- Switch A acts as a DNS proxy. The IP address of the real DNS server is 4.1.1.1.
- Switch B implements domain name resolution through Switch A.

Figure 1-8 Network diagram for DNS proxy



Configuration procedure

Note

Before performing the following configuration, assume that Switch A, the DNS server, and the host are reachable to each other and the IP addresses of the interfaces are configured as shown in Figure 1-8.

---

1) Configure the DNS server

This configuration may vary with different DNS servers. When a Windows server 2000 acts as the DNS server, refer to Dynamic Domain Name Resolution Configuration Example for related configuration information.

2) Configure the DNS proxy

Specify the DNS server 4.1.1.1.

```
<SwitchA> system-view
[SwitchA] dns server 4.1.1.1
```

Enable DNS proxy.

```
[SwitchA] dns proxy enable
```

3) Configure the DNS client

Enable the domain name resolution function.

```
<SwitchB> system-view
[SwitchB] dns resolve
```

Specify the DNS server 2.1.1.2.

```
[SwitchB] dns server 2.1.1.2
```

4) Configuration verification

Execute the **ping host.com** command on Switch B to verify that the communication between the Switch and the host is normal and that the corresponding destination IP address is 3.1.1.1.

```
[SwitchB] ping host.com
Trying DNS resolve, press CTRL_C to break
 Trying DNS server (2.1.1.2)
  PING host.com (3.1.1.1):
  56  data bytes, press CTRL_C to break
    Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=126 time=3 ms
    Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=126 time=1 ms
    Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=126 time=1 ms

  --- host.com ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/3 ms
```

Troubleshooting DNS Configuration

Symptom

After enabling the dynamic domain name resolution, the user cannot get the correct IP address.

Solution

- Use the **display dns dynamic-host** command to verify that the specified domain name is in the cache.
- If the specified domain name does not exist, check that dynamic domain name resolution is enabled and the DNS client can communicate with the DNS server.
- If the specified domain name is in the cache, but the IP address is incorrect, check that the DNS client has the correct IP address of the DNS server.
- Verify the mapping between the domain name and IP address is correct on the DNS server.

Table of Contents

1 IP Performance Optimization Configuration

When optimizing IP performance, go to these sections for information you are interested in:

IP Performance Overview

In some network environments, you can adjust the IP parameters to achieve best network performance. IP performance optimization configuration includes:

- Enabling the device to receive and forward directed broadcasts
- Configuring TCP timers
- Configuring the TCP buffer size
- Enabling ICMP error packets sending

Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network

Directed broadcast packets are broadcast on a specific network. In the destination IP address of a directed broadcast, the network ID is a network ID identifies the target network, and the host ID is all-one. If a device is allowed to forward directed broadcasts to a directly connected network, hackers may mount attacks to the network. Therefore, the device is disabled from receiving and forwarding directed broadcasts to a directly connected network by default. However, you should enable the feature when using the Wake on LAN function to forward directed broadcasts to a host on the remote network.

Enabling Reception of Directed Broadcasts to a Directly Connected Network

If a device is enabled to receive directed broadcasts, the device will determine whether to forward them according to the configuration on the outgoing interface.

Follow these steps to enable the device to receive directed broadcasts:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the device to receive directed broadcasts	**ip forward-broadcast**	Required By default, the device is disabled from receiving directed broadcasts.

Enabling Forwarding of Directed Broadcasts to a Directly Connected Network

Follow these steps to enable the device to forward directed broadcasts:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the interface to forward directed broadcasts	**ip forward-broadcast** [**acl** *acl-number*]	Required By default, the device is disabled from forwarding directed broadcasts.

Note

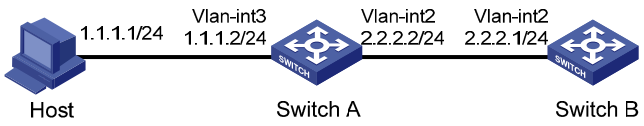
- If an ACL is referenced in the **ip forward-broadcast** [*acl-number*] command, only packets permitted by the ACL can be forwarded.
- If you repeatedly execute the **ip forward-broadcast acl** [*acl-number*] command on an interface, the last executed command takes effect only. If the command executed last time does not include the **acl** *acl-number*, the ACL configured previously will be removed.

Configuration Example

Network requirements

As shown in Figure 1-1, the host's interface and VLAN-interface 3 of Switch A are on the same network segment (1.1.1.0/24). VLAN-interface 2 of Switch A and VLAN-interface 2 of Switch B are on another network segment (2.2.2.0/24). The default gateway of the host is VLAN-interface 3 (IP address 1.1.1.2/24) of Switch A. Configure a static route on Switch B to enable the reachability between host and Switch B.

Figure 1-1 Network diagram for receiving and forwarding directed broadcasts



Configuration procedure

- Configure Switch A

Enable Switch A to receive directed broadcasts.

```
<SwitchA> system-view
[SwitchA] ip forward-broadcast
```

Configure IP addresses for VLAN-interface 3 and VLAN-interface 2.

```
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ip address 1.1.1.2 24
```

```
[SwitchA-Vlan-interface3] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 2.2.2.2 24
```

Enable VLAN-interface 2 to forward directed broadcasts.

```
[SwitchA-Vlan-interface2] ip forward-broadcast
```

- Configure Switch B

Enable Switch B to receive directed broadcasts.

```
<SwitchB> system-view
[SwitchB] ip forward-broadcast
```

Configure a static route to the host.

```
[SwitchB] ip route-static 1.1.1.1 24 2.2.2.2
```

Configure an IP address for VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 2.2.2.1 24
```

After the above configurations, if you ping the subnet broadcast address (2.2.2.255) of VLAN-interface 2 of Switch A on the host, the ping packets can be received by VLAN-interface 2 of Switch B. However, if you disable the **ip forward-broadcast** command, the ping packets cannot be received by the VLAN-interface 2 of Switch B.

Configuring TCP Optional Parameters

TCP optional parameters that can be configured include:

- synwait timer: When sending a SYN packet, TCP starts the synwait timer. If no response packet is received within the synwait timer interval, the TCP connection cannot be created.
- finwait timer: When a TCP connection is changed into FIN_WAIT_2 state, the finwait timer is started. If no FIN packets is received within the timer interval, the TCP connection will be terminated. If a FIN packet is received, the TCP connection state changes to TIME_WAIT. If a non-FIN packet is received, the system restarts the timer upon receiving the last non-FIN packet. The connection is broken after the timer expires.
- Size of TCP receive/send buffer

Follow these steps to configure TCP optional parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the TCP synwait timer	**tcp timer syn-timeout** *time-value*	Optional 75 seconds by default.
Configure the TCP finwait timer	**tcp timer fin-timeout** *time-value*	Optional 675 seconds by default.
Configure the size of TCP receive/send buffer	**tcp window** *window-size*	Optional 8 KB by default.

Caution

The actual length of the finwait timer is determined by the following formula:
Actual length of the finwait timer = (Configured length of the finwait timer – 75) + configured length of the synwait timer

Configuring ICMP to Send Error Packets

Sending error packets is a major function of ICMP. In case of network abnormalities, ICMP packets are usually sent by the network or transport layer protocols to notify corresponding devices so as to facilitate control and management.

Advantages of sending ICMP error packets

There are three kinds of ICMP error packets: redirect packets, timeout packets and destination unreachable packets. Their sending conditions and functions are as follows.

1) Sending ICMP redirect packets

A host may have only a default route to the default gateway in its routing table after startup. The default gateway will send ICMP redirect packets to the source host, telling it to reselect a correct next hop to send the subsequent packets, if the following conditions are satisfied:

- The receiving and forwarding interfaces are the same.
- The selected route has not been created or modified by ICMP redirect packet.
- The selected route is not the default route of the device.
- There is no source route option in the packet.

ICMP redirect packets function simplifies host administration and enables a host to gradually establish a sound routing table to find out the best route.

2) Sending ICMP timeout packets

If the device received an IP packet with a timeout error, it drops the packet and sends an ICMP timeout packet to the source.

The device will send an ICMP timeout packet under the following conditions:

- If the device finds the destination of a packet is not itself and the TTL field of the packet is 1, it will send a “TTL timeout” ICMP error message.
- When the device receives the first fragment of an IP datagram whose destination is the device itself, it starts a timer. If the timer times out before all the fragments of the datagram are received, the device will send a “reassembly timeout” ICMP error packet.

3) Sending ICMP destination unreachable packets

If the device receives an IP packet with the destination unreachable, it will drop the packet and send an ICMP destination unreachable error packet to the source.

Conditions for sending this ICMP packet:

- If neither a route nor the default route for forwarding a packet is available, the device will send a “network unreachable” ICMP error packet.
- If the destination of a packet is local while the transport layer protocol of the packet is not supported by the local device, the device sends a “protocol unreachable” ICMP error packet to the source.

- When receiving a packet with the destination being local and transport layer protocol being UDP, if the packet's port number does not match the running process, the device will send the source a "port unreachable" ICMP error packet.
- If the source uses "strict source routing" to send packets, but the intermediate device finds that the next hop specified by the source is not directly connected, the device will send the source a "source routing failure" ICMP error packet.
- When forwarding a packet, if the MTU of the sending interface is smaller than the packet but the packet has been set "Don't Fragment", the device will send the source a "fragmentation needed and Don't Fragment (DF)-set" ICMP error packet.

Disadvantages of sending ICMP error packets

Although sending ICMP error packets facilitates network control and management, it still has the following disadvantages:

- Sending a lot of ICMP packets will increase network traffic.
- If a device receives a lot of malicious packets that cause it to send ICMP error packets, its performance will be reduced.
- As the redirection function increases the routing table size of a host, the host's performance will be reduced if its routing table becomes very large.
- If a host sends malicious ICMP destination unreachable packets, end users may be affected.

To prevent such problems, you can disable the device from sending ICMP error packets.

Follow these steps to disable sending of ICMP error packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable sending of ICMP redirect packets	**ip redirects enable**	Required Disabled by default.
Disable sending of ICMP timeout packets	**undo ip ttl-expires**	Required Enabled by default.
Enable sending of ICMP destination unreachable packets	**ip unreachables enable**	Required Disabled by default.

Note

The device stops sending "TTL timeout" ICMP error packets after sending ICMP timeout packets is disabled. However, "reassembly timeout" error packets will be sent normally.

Displaying and Maintaining IP Performance Optimization

To do…	Use the command…	Remarks
Display current TCP connection state	**display tcp status**	Available in any view
Display TCP connection statistics	**display tcp statistics**	
Display UDP statistics	**display udp statistics**	
Display statistics of IP packets	**display ip statistics** [**slot** *slot-number*]	
Display statistics of ICMP flows	**display icmp statistics** [**slot** *slot-number*]	
Display socket information	**display ip socket** [**socktype** *sock-type*] [*task-id socket-id*] [**slot** *slot-number*]	
Display FIB information	**display fib** [**vpn-instance** *vpn-instance-name*] [**|** { **begin** | **include** | **exclude** } *regular-expression* | **acl** *acl-number* | **ip-prefix** *ip-prefix-name*]	
Display FIB information matching the specified destination IP address	**display fib** [**vpn-instance** *vpn-instance-name*] *ip-address* [*mask* | *mask-length*]	
Clear statistics of IP packets	**reset ip statistics** [**slot** *slot-number*]	Available in user view
Clear statistics of TCP connections	**reset tcp statistics**	Available in user view
Clear statistics of UDP traffic	**reset udp statistics**	Available in user view

Table of Contents

1 UDP Helper Configuration

When configuring UDP Helper, go to these sections for information you are interested in:

Note

UDP Helper can be currently configured on VLAN interfaces only.

Introduction to UDP Helper

Sometimes, a host needs to forward broadcasts to obtain network configuration information or request the names of other devices on the network. However, if the server or the device to be requested is located in another broadcast domain, the host cannot obtain such information through broadcast.

To solve this problem, the device provides the UDP Helper function to relay specified UDP packets. In other words, UDP Helper functions as a relay agent that converts UDP broadcast packets into unicast packets and forwards them to a specified destination server.

With UDP Helper enabled, the device decides whether to forward a received UDP broadcast packet according to the UDP destination port number of the packet.

- If the destination port number of the packet matches the one pre-configured on the device, the device modifies the destination IP address in the IP header, and then sends the packet to the specified destination server.
- If not, the device sends the packet to the upper layer protocol for processing.

Configuring UDP Helper

Follow these steps to configure UDP Helper:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable UDP Helper	**udp-helper enable**	Required Disabled by default.
Enable the forwarding of packets with the specified UDP destination port number(s)	**udp-helper port** { *port-number* | **dns** | **netbios-ds** | **netbios-ns** | **tacacs** | **tftp** | **time** }	Required No UDP port number is specified by default.

To do…	Use the command…	Remarks
Enter interface view	**interface** *interface-type interface-number*	—
Specify the destination server to which UDP packets are to be forwarded	**udp-helper server** *ip-address*	Required No destination server is specified by default.

Caution

- The UDP Helper enabled device cannot forward DHCP broadcast packets. That is to say, the UDP port number cannot be set to 67 or 68.
- For the **dns**, **netbios-ds**, **netbios-ns**, **tacacs**, **tftp**, and **time** keywords, you can specify port numbers or the corresponding parameters. For example, **udp-helper port** 53 and **udp-helper port dns** specify the same UDP port number.
- The configuration of all UDP ports is removed if you disable UDP Helper.
- You can configure up to 256 UDP port numbers to enable the forwarding of packets with these UDP port numbers.
- You can configure up to 20 destination servers on an interface.

Displaying and Maintaining UDP Helper

To do…	Use the command…	Remarks
Displays the information of forwarded UDP packets	**display udp-helper server** [**interface** *interface-type interface-number*]	Available in any view
Clear statistics about packets forwarded	**reset udp-helper packet**	Available in user view

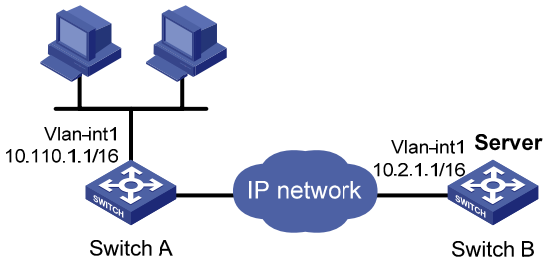
UDP Helper Configuration Examples

UDP Helper Configuration Example

Network requirements

On Switch A, configure UDP helper to forward broadcast packets (with UDP destination port number 55 and destination IP address 255.255.255.255 or 10.110.255.255 to the destination server 10.2.1.1/16.

Figure 1-1 Network diagram for UDP Helper configuration



Configuration procedure

Note

The following configuration assumes that a route from Switch A to the network segment 10.2.0.0/16 is available.

Enable UDP Helper.

```
<SwitchA> system-view
[SwitchA] udp-helper enable
```

Enable the forwarding broadcast packets with the UDP destination port 55.

```
[SwitchA] udp-helper port 55
```

Specify the destination server 10.2.1.1 on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.110.1.1 16
[SwitchA-Vlan-interface1] udp-helper server 10.2.1.1
```

Table of Contents

1 URPF Configuration

When configuring URPF, go to these sections for information you are interested in:

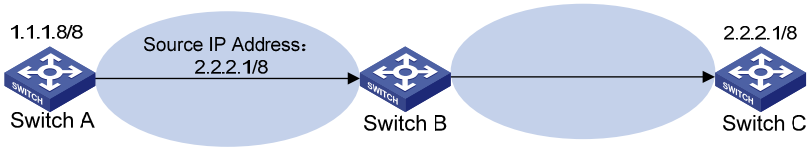
URPF Overview

What is URPF

Unicast Reverse Path Forwarding (URPF) protects a network against source address spoofing attacks.

Attackers launch attacks by creating a series of packets with forged source addresses. For applications using IP-address-based authentication, this type of attacks allows unauthorized users to access the system in the name of authorized users, or even access the system as the administrator. Even if the attackers cannot receive any response packets, the attacks are still disruptive to the attacked target.

Figure 1-1 Attack based on source address spoofing



As shown in Figure 1-1, Switch A originates a request to the server (Switch B) by sending a packet with a forged source IP address of 2.2.2.1/8, and Switch B sends a packet to Switch C at 2.2.2.1/8 in response to the request. Consequently, both Switch B and Switch C are attacked.

URPF can prevent source address spoofing attacks.

How URPF Works

URPF works as follows:

1) First, URPF checks the source address validity, and then:
- Discards packets with broadcast source addresses.
- Discards packets with all-zero source addresses but non-broadcast destination addresses. (A packet with source address 0.0.0.0 and destination address 255.255.255.255 might be a DHCP or BOOT packet, and thus is not discarded.)

2) If the source address of an incoming packet is found in the FIB table, URPF does a reverse route lookup for routes to the source address of the packet. If at least one outgoing interface of such a route matches the receiving interface, the packet passes the check. Otherwise, the packet is rejected.

3) If the source address of an incoming packet is not found in the FIB table, the packet is rejected.

Configuring URPF

Follow these steps to configure URPF:

To do...	Use the command…	Remarks
Enter system view	**system-view**	—
Enable URPF check	**ip urpf strict**	Required Disabled by default.

Table of Contents

1 IPv6 Basics Configuration

When configuring IPv6 basics, go to these sections for information you are interested in:

Note

The term “router” or the router icon in this document refers to a router in a generic sense or a Layer 3 Ethernet switch running a routing protocol.

IPv6 Overview

Internet Protocol Version 6 (IPv6), also called IP next generation (IPng), was designed by the Internet Engineering Task Force (IETF) as the successor to Internet Protocol Version 4 (IPv4). The significant difference between IPv6 and IPv4 is that IPv6 increases the IP address size from 32 bits to 128 bits. This section covers the following:

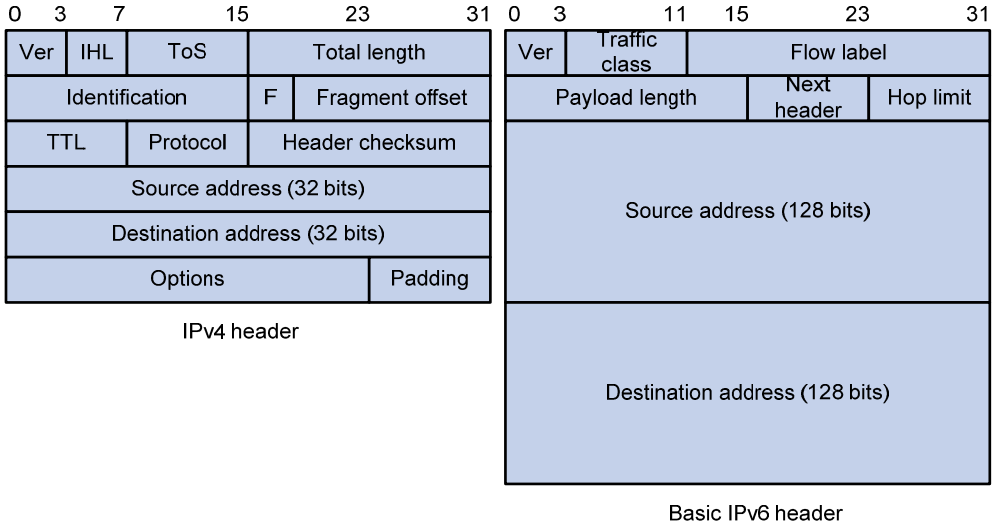
IPv6 Features

Header format simplification

IPv6 cuts down some IPv4 header fields or move them to the IPv6 extension headers to reduce the length of the basic IPv6 header. IPv6 uses the basic header with a fixed length, thus making IPv6 packet handling simple and improving the forwarding efficiency. Although the IPv6 address size is four times

the IPv4 address size, the basic IPv6 header size is 40 bytes and is only twice the IPv4 header size (excluding the Options field).

Figure 1-1 Comparison between IPv4 packet header format and basic IPv6 packet header format



Adequate address space

The source and destination IPv6 addresses are both 128 bits (16 bytes) long. IPv6 can provide 3.4 x 10^{38} addresses to fully meet the requirements of hierarchical address division as well as allocation of public and private addresses.

Hierarchical address structure

IPv6 adopts the hierarchical address structure to quicken route search and reduce the system sources occupied by the IPv6 routing table by route aggregation.

Automatic address configuration

To simplify host configuration, IPv6 supports stateful and stateless address configuration.

- Stateful address configuration means that a host acquires an IPv6 address and related information from a server (for example, a DHCP server).
- Stateless address configuration means that a host automatically generates an IPv6 address and related information on the basis of its own link-layer address and the prefix information advertised by a router.

In addition, a host can generate a link-local address on the basis of its own link-layer address and the default prefix (FE80::/10) to communicate with other hosts on the same link.

Built-in security

IPv6 uses IPSec as its standard extension header to provide end-to-end security. This feature provides a standard for network security solutions and enhances the interoperability between different IPv6 applications.

QoS support

The Flow Label field in the IPv6 header allows the device to label packets of a flow and provide special handling for these packets.

Enhanced neighbor discovery mechanism

The IPv6 neighbor discovery protocol is implemented through a group of Internet Control Message Protocol Version 6 (ICMPv6) messages that manage the information exchange between neighbor nodes on the same link. The group of ICMPv6 messages takes the place of Address Resolution Protocol (ARP) messages, Internet Control Message Protocol version 4 (ICMPv4) router discovery messages, and ICMPv4 redirection messages and provides a series of other functions.

Flexible extension headers

IPv6 cancels the Options field in the IPv4 header but introduces multiple extension headers to provide scalability while improving efficiency. The Options field contains 40 bytes at most, while the size of IPv6 extension headers is restricted to the maximum size of IPv6 packets.

Introduction to IPv6 Address

IPv6 address format

An IPv6 address is represented as a set of 16-bit hexadecimals, separated by colons. An IPv6 address is divided into eight groups, and the 16 bits of each group are represented by four hexadecimal numbers, for example, 2001:0000:130F:0000:0000:09C0:876A:130B.

To simplify the representation of IPv6 addresses, zeros in IPv6 addresses can be handled as follows:

- Leading zeros in each group can be removed. For example, the above-mentioned address can be represented in a shorter format as 2001:0:130F:0:0:9C0:876A:130B.
- If an IPv6 address contains two or more consecutive groups of zeros, they can be replaced by a double-colon ::. For example, the above-mentioned address can be represented in the shortest format as 2001:0:130F::9C0:876A:130B.

Caution

A double-colon can be used only once in an IPv6 address. Otherwise, the device is unable to determine how many zeros that double-colons represent when converting them to zeros to restore a 128-bit IPv6 address.

An IPv6 address consists of two parts: address prefix and interface ID. The address prefix and the interface ID are respectively equivalent to the network ID and the host ID in an IPv4 address.

An IPv6 address prefix is written in IPv6-address/prefix-length notation, where the IPv6-address is in any of the notations above mentioned, and prefix-length is a decimal number indicating how many bits from the left-most of an IPv6 address is the address prefix.

IPv6 address classification

IPv6 addresses fall into three types: unicast address, multicast address, and anycast address.

- Unicast address: An identifier for a single interface, similar to an IPv4 unicast address. A packet sent to a unicast address is delivered to the interface identified by that address.
- Multicast address: An identifier for a set of interfaces (typically belonging to different nodes), similar to an IPv4 multicast address. A packet sent to a multicast address is delivered to all interfaces identified by that address.

- Anycast address: An identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the target interface is nearest to the source, according to a routing protocol's measure of distance).

Note

There are no broadcast addresses in IPv6. Their function is replaced by multicast addresses.

The type of an IPv6 address is designated by the first several bits called format prefix. Table 1-1 lists the mappings between address types and format prefixes.

Table 1-1 Mappings between address types and format prefixes

Type		Format prefix (binary)	IPv6 prefix ID
Unicast address	Unassigned address	00...0 (128 bits)	::/128
	Loopback address	00...1 (128 bits)	::1/128
	Link-local address	1111111010	FE80::/10
	Site-local address	1111111011	FEC0::/10
	Global unicast address	other forms	—
Multicast address		11111111	FF00::/8
Anycast address		Anycast addresses are taken from unicast address space and are not syntactically distinguishable from unicast addresses.	

Unicast address

There are several types of unicast addresses, including aggregatable global unicast address, link-local address, and site-local address.

- The aggregatable global unicast addresses, equivalent to public IPv4 addresses, are provided for network service providers. This type of address allows efficient prefix aggregation to restrict the number of global routing entries.
- The link-local addresses are used for communication between link-local nodes in neighbor discovery and stateless autoconfiguration. Packets with link-local source or destination addresses are not forwarded to other links.
- IPv6 unicast site-local addresses are similar to private IPv4 addresses. Packets with site-local source or destination addresses are not forwarded out of the local site (a private network).
- Loopback address: The unicast address 0:0:0:0:0:0:0:1 (represented in the shortest format as ::1) is called the loopback address and may never be assigned to any physical interface. Like the loopback address in IPv4, it may be used by a node to send an IPv6 packet to itself.
- Unassigned address: The unicast address "::" is called the unassigned address and may not be assigned to any node. Before acquiring a valid IPv6 address, a node may fill this address in the source address field of an IPv6 packet. It cannot be used as a destination IPv6 address.

Multicast address

IPv6 multicast addresses listed in Table 1-2 are reserved for special purpose.

Table 1-2 Reserved IPv6 multicast addresses

Address	Application
FF01::1	Node-local scope all nodes multicast address
FF02::1	Link-local scope all nodes multicast address
FF01::2	Node-local scope all routers multicast address
FF02::2	Link-local scope all routers multicast address
FF05::2	Site-local scope all routers multicast address

Besides, there is another type of multicast address: solicited-node address. A solicited-node multicast address is used to acquire the link-layer address of a neighbor node on the same link, and is also used for duplicate address detection (DAD). Each IPv6 unicast or anycast address has a corresponding solicited-node address. The format of a solicited-node multicast address is as follows:

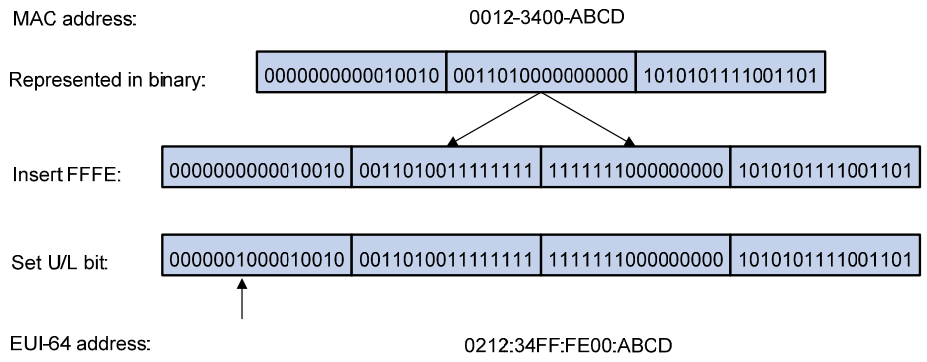
FF02:0:0:0:0:1:FFXX:XXXX

Where, FF02:0:0:0:0:1:FF is permanent and consists of 104 bits, and XX:XXXX is the last 24 bits of an IPv6 unicast or anycast address.

Interface identifier in IEEE EUI-64 format

An interface identifier is used to identify a unique interface on a link and is 64 bits long. An interface identifier in IEEE EUI-64 format is derived from the link-layer address (MAC) of an interface. A MAC address is 48 bits long and therefore, to get the interface identifier, the hexadecimal number FFFE needs to be inserted in the middle of the MAC address (behind the 24 high-order bits). To ensure the interface identifier obtained from a MAC address is unique, it is necessary to set the universal/local (U/L) bit (the seventh high-order bit) to “1”. Thus, an interface identifier in IEEE EUI-64 format is obtained.

Figure 1-2 Convert a MAC address into an EUI-64 interface identifier



Introduction to IPv6 Neighbor Discovery Protocol

The IPv6 Neighbor Discovery Protocol (NDP) uses five types of ICMPv6 messages to implement the following functions:

- Address resolution
- Neighbor reachability detection

- Duplicate address detection
- Router/prefix discovery and address autoconfiguration
- Redirection

Table 1-3 lists the types and functions of ICMPv6 messages used by the NDP.

Table 1-3 Types and functions of ICMPv6 messages

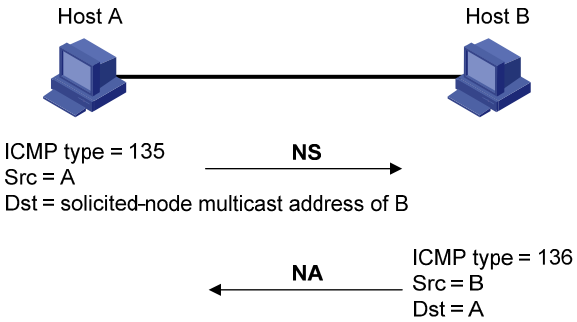
ICMPv6 message	Number	Function
Neighbor solicitation (NS) message	135	Used to acquire the link-layer address of a neighbor
		Used to verify whether the neighbor is reachable
		Used to perform a duplicate address detection
Neighbor advertisement (NA) message	136	Used to respond to an NS message
		When the link layer changes, the local node initiates an NA message to notify neighbor nodes of the node information change.
Router solicitation (RS) message	133	After started, a node sends an RS message to request the router for an address prefix and other configuration information for the purpose of autoconfiguration.
Router advertisement (RA) message	134	Used to respond to an RS message
		With the RA message suppression disabled, the router regularly sends an RA message containing information such as prefix information options and flag bits.
Redirect message	137	When a certain condition is satisfied, the default gateway sends a redirect message to the source host so that the host can reselect a correct next hop router to forward packets.

The NDP mainly provides the following functions:

Address resolution

Similar to the ARP function in IPv4, a node acquires the link-layer addresses of neighbor nodes on the same link through NS and NA messages. Figure 1-3 shows how node A acquires the link-layer address of node B.

Figure 1-3 Address resolution



The address resolution procedure is as follows:

1) Node A multicasts an NS message. The source address of the NS message is the IPv6 address of the sending interface of node A and the destination address is the solicited-node multicast address of node B. The NS message contains the link-layer address of node A.

2) After receiving the NS message, node B judges whether the destination address of the packet is its solicited-node multicast address. If yes, node B learns the link-layer address of node A, and then unicasts an NA message containing its link-layer address.
3) Node A acquires the link-layer address of node B from the NA message.

Neighbor reachability detection

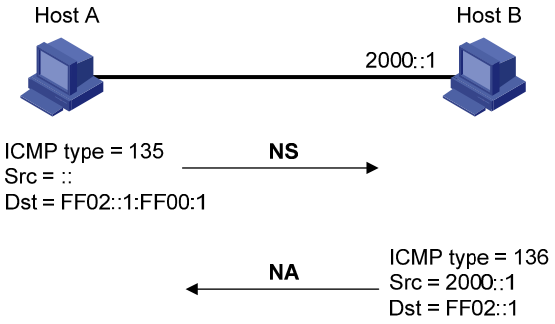
After node A acquires the link-layer address of its neighbor node B, node A can verify whether node B is reachable according to NS and NA messages.

1) Node A sends an NS message whose destination address is the IPv6 address of node B.
2) If node A receives an NA message from node B, node A considers that node B is reachable. Otherwise, node B is unreachable.

Duplicate address detection

After node A acquires an IPv6 address, it will perform duplicate address detection (DAD) to determine whether the address is being used by any other node (similar to the gratuitous ARP function of IPv4). DAD is accomplished through NS and NA messages. Figure 1-4 shows the DAD procedure.

Figure 1-4 Duplicate address detection



The DAD procedure is as follows:

1) Node A sends an NS message whose source address is the unassigned address :: and destination address is the corresponding solicited-node multicast address of the IPv6 address to be detected. The NS message contains the IPv6 address.
2) If node B uses this IPv6 address, node B returns an NA message. The NA message contains the IPv6 address of node B.
3) Node A learns that the IPv6 address is being used by node B after receiving the NA message from node B. Otherwise, node B is not using the IPv6 address and node A can use it.

Router/prefix discovery and address autoconfiguration

Router/prefix discovery means that a node locates the neighboring routers, and learns the prefix of the network where the host is located, and other configuration parameters from the received RA message.

Stateless address autoconfiguration means that a node automatically generates an IPv6 address according to the information obtained through router/prefix discovery.

The router/prefix discovery is implemented through RS and RA messages. The router/prefix discovery procedure is as follows:

1) After started, a node sends an RS message to request the router for the address prefix and other configuration information for the purpose of autoconfiguration.

2) The router returns an RA message containing information such as prefix information option. (The router also regularly sends an RA message.)
3) The node automatically generates an IPv6 address and other information for its interface according to the address prefix and other configuration parameters in the RA message.

Note

- In addition to an address prefix, the prefix information option also contains the preferred lifetime and valid lifetime of the address prefix. After receiving a periodic RA message, the node updates the preferred lifetime and valid lifetime of the address prefix accordingly.
- An automatically generated address is applicable within the valid lifetime and is removed when the valid lifetime times out.

Redirection

When a host is started, its routing table may contain only the default route to the gateway. When certain conditions are satisfied, the gateway sends an ICMPv6 redirect message to the source host so that the host can select a better next hop to forward packets (similar to the ICMP redirection function in IPv4).

The gateway sends an IPv6 ICMP redirect message when the following conditions are satisfied:

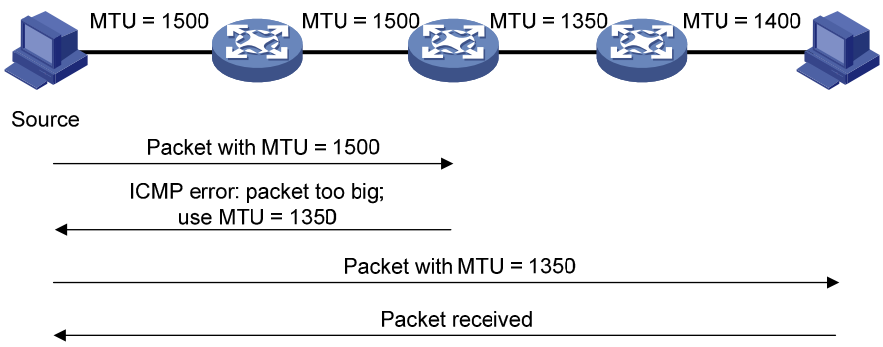
- The receiving interface is the forwarding interface.
- The selected route itself is not created or modified by an IPv6 ICMP redirect message.
- The selected route is not the default route.
- The forwarded IPv6 packet does not contain any routing extension header.

IPv6 PMTU Discovery

The links that a packet passes from the source to the destination may have different MTUs. In IPv6, when the packet size exceeds the path MTU (the minimum MTU of all links), the packet will be fragmented at the source end so as to reduce the processing pressure of forwarding devices and utilize network resources properly.

The path MTU (PMTU) discovery mechanism is to find the minimum MTU of all links in the path from the source to the destination. Figure 1-5 shows the working procedure of PMTU discovery.

Figure 1-5 Working procedure of PMTU discovery



The working procedure of the PMTU discovery is as follows:

1) The source host uses its MTU to send packets to the destination host.
2) If the MTU supported by a forwarding interface is smaller than the packet size, the forwarding device will discard the packet and return an ICMPv6 error packet containing the interface MTU to the source host.
3) After receiving the ICMPv6 error packet, the source host uses the returned MTU to send packets to the destination.
4) Step 2 to step 3 are repeated until the destination host receives the packet. In this way, the minimum MTU of all links in the path from the source host to the destination host is determined.

Introduction to IPv6 DNS

IPv6 Domain Name System (DNS) is responsible for translating domain names into IPv6 addresses, instead of IPv4 addresses.

Like IPv4 DNS, IPv6 DNS also involves static domain name resolution and dynamic domain name resolution. The function and implementation of these two types of domain name resolution are the same as those of IPv4 DNS. For details, refer to *DNS Configuration* in the *IP Services Volume*.

Usually, the DNS server connecting IPv4 and IPv6 networks not only contains A records (IPv4 addresses), but also AAAA records (IPv6 addresses). The DNS server can convert domain names into IPv4 addresses or IPv6 addresses. In this way, the DNS server implements the functions of both IPv6 DNS and IPv4 DNS.

Protocols and Standards

Protocols and standards related to IPv6 include:

- RFC 1881: IPv6 Address Allocation Management
- RFC 1887: An Architecture for IPv6 Unicast Address Allocation
- RFC 1981: Path MTU Discovery for IP version 6
- RFC 2375: IPv6 Multicast Address Assignments
- RFC 2460: Internet Protocol, Version 6 (IPv6) Specification.
- RFC 2461: Neighbor Discovery for IP Version 6 (IPv6)
- RFC 2462: IPv6 Stateless Address Autoconfiguration
- RFC 2463: Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
- RFC 2464: Transmission of IPv6 Packets over Ethernet Networks
- RFC 2526: Reserved IPv6 Subnet Anycast Addresses
- RFC 3307: Allocation Guidelines for IPv6 Multicast Addresses
- RFC 3513: Internet Protocol Version 6 (IPv6) Addressing Architecture
- RFC 3596: DNS Extensions to Support IP Version 6

IPv6 Basics Configuration Task List

Complete the following tasks to perform IPv6 basics configuration:

Task	Remarks
Configuring Basic IPv6 Functions	Required
Configuring IPv6 NDP	Optional
Configuring PMTU Discovery	Optional
Configuring IPv6 TCP Properties	Optional

Task	Remarks
Configuring ICMPv6 Packet Sending	Optional
Configuring IPv6 DNS Client	Optional

Configuring Basic IPv6 Functions

Enabling IPv6

Before performing IPv6-related configurations, you need to Enable IPv6. Otherwise, an interface cannot forward IPv6 packets even if it has an IPv6 address configured.

Follow these steps to Enable IPv6:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IPv6	**ipv6**	Required Disabled by default.

Configuring an IPv6 Unicast Address

IPv6 site-local addresses and aggregatable global unicast addresses can be configured in the following ways:

- EUI-64 format: When the EUI-64 format is adopted, the IPv6 address prefix of an interface is the configured prefix, and the interface identifier is derived from the link-layer address of the interface.
- Manual configuration: IPv6 site-local addresses or aggregatable global unicast addresses are configured manually.

IPv6 link-local addresses can be configured in either of the following ways:

- Automatic generation: The device automatically generates a link-local address for an interface according to the link-local address prefix (FE80::/10) and the link-layer address of the interface.
- Manual assignment: IPv6 link-local addresses can be assigned manually.

Follow these steps to configure an IPv6 unicast address:

To do...		Use the command...	Remarks
Enter system view		**system-view**	—
Enter interface view		**interface** *interface-type interface-number*	—
Configure an IPv6 aggregatable global unicast address or site-local address	Manually assign an IPv6 address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }	One of the two commands is required. By default, no site-local address or aggregatable global unicast address is configured for an interface.
	Adopt the EUI-64 format to form an IPv6 address	**ipv6 address** *ipv6-address*/*prefix-length* **eui-64**	

To do...		Use the command...	Remarks
Configure an IPv6 link-local address	Automatically generate a link-local address for the interface	**ipv6 address auto link-local**	Optional By default, after an IPv6 site-local address or aggregatable global unicast address is configured for an interface, a link-local address will be generated automatically.
	Manually assign a link-local address for the interface	**ipv6 address** *ipv6-address* **link-local**	

Note

- After an IPv6 site-local address or aggregatable global unicast address is configured for an interface, a link-local address is generated automatically. The automatically generated link-local address is the same as the one generated by using the **ipv6 address auto link-local** command. If a link-local address is manually assigned to an interface, this manual link-local address takes effect. If the manually assigned link-local address is removed, the automatically generated link-local address takes effect.
- Manual assignment takes precedence over automatic generation. That is, if you first adopt automatic generation and then manual assignment, the manually assigned link-local address will overwrite the automatically generated one. If you first adopt manual assignment and then automatic generation, the automatically generated link-local address will not take effect and the link-local address of an interface is still the manually assigned one. If you delete the manually assigned address, the automatically generated link-local address is validated.
- The **undo ipv6 address auto link-local** command can only remove the link-local addresses generated through the **ipv6 address auto link-local** command. However, if an IPv6 site-local address or aggregatable global unicast address is already configured for an interface, the interface still has a link-local address because the system automatically generates one for the interface. If no IPv6 site-local address or aggregatable global unicast address is configured, the interface has no link-local address.

Configuring IPv6 NDP

Configuring a Static Neighbor Entry

The IPv6 address of a neighbor node can be resolved into a link-layer address dynamically through NS and NA messages or through a manually configured static neighbor entry.

The device uniquely identifies a static neighbor entry according to the neighbor IPv6 address and the local Layer 3 interface ID. Currently, there are two configuration methods:

- Associate a neighbor IPv6 address and link-layer address with a Layer 3 interface.
- Associate a neighbor IPv6 address and link-layer address with a port in a VLAN.

Follow these steps to configure a static neighbor entry:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure a static neighbor entry	**ipv6 neighbor** *ipv6-address mac-address* { *vlan-id port-type port-number* | **interface** *interface-type interface-number* }	Required

Caution

You can adopt either of the two methods above to configure a static neighbor entry.

- After a static neighbor entry is configured by using the first method, the device needs to resolve the corresponding Layer 2 port information of the VLAN interface.
- If you adopt the second method, you should ensure that the corresponding VLAN interface exists and that the Layer 2 port specified by *port-type port-number* belongs to the VLAN specified by *vlan-id*. After a static neighbor entry is configured, the device relates the VLAN interface to the IPv6 address to uniquely identify a static neighbor entry.

Configuring the Maximum Number of Neighbors Dynamically Learned

The device can dynamically acquire the link-layer address of a neighbor node through NS and NA messages and add it into the neighbor table. Too large a neighbor table may reduce the forwarding performance of the device. You can restrict the size of the neighbor table by setting the maximum number of neighbors that an interface can dynamically learn. When the number of dynamically learned neighbors reaches the threshold, the interface will stop learning neighbor information.

Follow these steps to configure the maximum number of neighbors dynamically learned:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the maximum number of neighbors dynamically learned by an interface	**ipv6 neighbors max-learning-num** *number*	Optional The default value is 4096.

Configuring Parameters Related to RA Messages

You can enable an interface to send RA messages, and configure the interval for sending RA messages and parameters in RA messages. After receiving an RA message, a host can use these parameters to perform corresponding operations. Table 1-4 lists the configurable parameters in an RA message and their descriptions.

Table 1-4 Parameters in an RA message and their descriptions

Parameters	Description
Cur hop limit	When sending an IPv6 packet, a host uses the value to fill the Cur Hop Limit field in IPv6 headers. The value is also filled into the Cur Hop Limit field in response messages of a device.
Prefix information options	After receiving the prefix information advertised by the device, the hosts on the same link can perform stateless autoconfiguration.
M flag	This field determines whether hosts use the stateful autoconfiguration to acquire IPv6 addresses. If the M flag is set to 1, hosts use the stateful autoconfiguration to acquire IPv6 addresses (for example, through a DHCP server). Otherwise, hosts use the stateless autoconfiguration to acquire IPv6 addresses, that is, hosts generate IPv6 addresses according to their own link-layer addresses and the prefix information issued by the router.
O flag	This field determines whether hosts use the stateful autoconfiguration to acquire information other than IPv6 addresses. If the O flag is set to 1, hosts use the stateful autoconfiguration to acquire information other than IPv6 addresses (for example, through a DHCP server). Otherwise, hosts use the stateless autoconfiguration to acquire information other than IPv6 addresses.
Router lifetime	This field is used to set the lifetime of the router that sends RA messages to serve as the default router of hosts. According to the router lifetime in the received RA messages, hosts determine whether the router sending RA messages can serve as the default router.
Retrans timer	If the device fails to receive a response message within the specified time after sending an NS message, the device will retransmit the NS message.
Reachable time	If the neighbor reachability detection shows that a neighbor is reachable, the device considers the neighbor reachable within the specified reachable time. If the device needs to send a packet to a neighbor after the specified reachable time expires, the device will reconfirm whether the neighbor is reachable.

Note

The values of the Retrans Timer and the Reachable Time configured for an interface are sent to hosts via RA messages. Furthermore, this interface sends NS messages at the interval of Retrans Timer and considers a neighbor reachable within the Reachable Time.

Follow these steps to configure parameters related to an RA message:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the hop limit	**ipv6 nd hop-limit** *value*	Optional 64 by default.
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Disable the RA message suppression	**undo ipv6 nd ra halt**	Required By default, RA messages are suppressed.
Configure the maximum and minimum intervals for sending RA messages	**ipv6 nd ra interval** *max-interval-value min-interval-value*	Optional By default, the maximum interval for sending RA messages is 600 seconds, and the minimum interval is 200 seconds. The device sends RA messages at random intervals between the maximum interval and the minimum interval. The minimum interval should be less than or equal to 0.75 times the maximum interval.
Configure the prefix information in RA messages	**ipv6 nd ra prefix** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* } *valid-lifetime preferred-lifetime* [**no-autoconfig** | **off-link**] *	Optional By default, no prefix information is configured for RA messages, and the IPv6 address of the interface sending RA messages is used as the prefix information.
Set the M flag bit to 1	**ipv6 nd autoconfig managed-address-flag**	Optional By default, the M flag bit is set to 0, that is, hosts acquire IPv6 addresses through stateless autoconfiguration.
Set the O flag bit to 1	**ipv6 nd autoconfig other-flag**	Optional By default, the O flag bit is set to 0, that is, hosts acquire other information through stateless autoconfiguration.
Configure the router lifetime in RA messages	**ipv6 nd ra router-lifetime** *value*	Optional 1800 seconds by default.
Set the NS retransmission timer	**ipv6 nd ns retrans-timer** *value*	Optional By default, the local interface sends NS messages at an interval of 1000 milliseconds, and the value of the Retrans Timer field in RA messages sent by the local interface is 0.
Set the reachable time	**ipv6 nd nud reachable-time** *value*	Optional By default, the neighbor reachable time on the local interface is 30000 milliseconds, and the value of the Reachable Timer field in RA messages is 0.

Caution

The maximum interval for sending RA messages should be less than or equal to the router lifetime in RA messages.

Configuring the Maximum Number of Attempts to Send an NS Message for DAD

An interface sends a neighbor solicitation (NS) message for duplicate address detection after acquiring an IPv6 address. If the interface does not receive a response within a specified time (determined by the **ipv6 nd ns retrans-timer** command), it continues to send an NS message. If it still does not receive a response after the number of sent attempts reaches a configurable threshold, the acquired address is considered usable.

Follow these steps to configure the attempts to send an NS message for DAD:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the number of attempts to send an NS message for DAD	**ipv6 nd dad attempts** *value*	Optional 1 by default. When the *value* argument is set to 0, DAD is disabled.

Configuring PMTU Discovery

Configuring a Static PMTU for a Specified IPv6 Address

You can configure a static PMTU for a specified destination IPv6 address. When a source host sends a packet through an interface, it compares the interface MTU with the static PMTU of the specified destination IPv6 address. If the packet size is larger than the smaller one between the two values, the host fragments the packet according to the smaller value.

Follow these steps to configure a static PMTU for a specified address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a static PMTU for a specified IPv6 address	**ipv6 pathmtu** *ipv6-address* [*value*]	Required By default, no static PMTU is configured.

Configuring the Aging Time for Dynamic PMTUs

After the path MTU from a source host to a destination host is dynamically determined (refer to IPv6 PMTU Discovery), the source host sends subsequent packets to the destination host on basis of this

MTU. After the aging time expires, the dynamic PMTU is removed and the source host re-determines a dynamic path MTU through the PMTU mechanism.

The aging time is invalid for a static PMTU.

Follow these steps to configure the aging time for dynamic PMTUs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the aging time for dynamic PMTUs	**ipv6 pathmtu age** *age-time*	Optional 10 minutes by default.

Configuring IPv6 TCP Properties

The IPv6 TCP properties you can configure include:

- synwait timer: When a SYN packet is sent, the synwait timer is triggered. If no response packet is received before the synwait timer expires, the IPv6 TCP connection establishment fails.
- finwait timer: When the IPv6 TCP connection status is FIN_WAIT_2, the finwait timer is triggered. If no packet is received before the finwait timer expires, the IPv6 TCP connection is terminated. If a FIN packet is received, the IPv6 TCP connection status becomes TIME_WAIT. If non-FIN packets are received, the finwait timer is reset upon receipt of the last non-FIN packet and the connection is terminated after the finwait timer expires.
- Size of the IPv6 TCP sending/receiving buffer.

Follow these steps to configure IPv6 TCP properties:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the finwait timer	**tcp ipv6 timer fin-timeout** *wait-time*	Optional 675 seconds by default.
Set the synwait timer	**tcp ipv6 timer syn-timeout** *wait-time*	Optional 75 seconds by default.
Set the size of the IPv6 TCP sending/receiving buffer	**tcp ipv6 window** *size*	Optional 8 KB by default.

Configuring ICMPv6 Packet Sending

Configuring the Maximum ICMPv6 Error Packets Sent in an Interval

If too many ICMPv6 error packets are sent within a short time in a network, network congestion may occur. To avoid network congestion, you can control the maximum number of ICMPv6 error packets sent within a specified time, currently by adopting the token bucket algorithm.

You can set the capacity of a token bucket, namely, the number of tokens in the bucket. In addition, you can set the update interval of the token bucket, namely, the interval for restoring the configured capacity. One token allows one ICMPv6 error packet to be sent. Each time an ICMPv6 error packet is sent, the number of tokens in a token bucket decreases by one. If the number of ICMPv6 error packets

successively sent exceeds the capacity of the token bucket, the additional ICMPv6 error packets cannot be sent out until the capacity of the token bucket is restored.

Follow these steps to configure the capacity and update interval of the token bucket:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the capacity and update interval of the token bucket	**Ipv6 icmp-error** { **bucket** *bucket-size* | **ratelimit** *interval* } *	Optional By default, the capacity of a token bucket is 10 and the update interval is 100 milliseconds. That is, at most 10 IPv6 ICMP error packets can be sent within 100 milliseconds. The update interval “0” indicates that the number of ICMPv6 error packets sent is not restricted.

Enable Sending of Multicast Echo Replies

If hosts are capable of answering multicast echo requests, Host A can attack Host B by sending an echo request with the source being Host B to a multicast address, then all the hosts in the multicast group will send echo replies to Host B. Therefore, to prevent such an attack, a device is disabled from replying multicast echo requests by default.

Follow these steps to enable sending of multicast echo replies:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable sending of multicast echo replies	**ipv6 icmpv6 multicast-echo-reply enable**	Not enabled by default.

Enabling Sending of ICMPv6 Time Exceeded Packets

A device sends an ICMPv6 time exceeded packet in the following cases.

- If a received IPv6 packet’s destination IP address is not the local address and its hop count is 1, the device sends an ICMPv6 time-to-live count exceeded packet to the source.
- Upon receiving the first fragment of an IPv6 datagram with the destination IP address being the local address, the device starts a timer. If the timer expires before all the fragments arrive, an ICMPv6 fragment reassembly time exceeded packet is sent to the source.

If large amounts of malicious packets are received, the performance of a device degrades greatly because it has to send back ICMP time exceeded packets. You can disable sending of ICMPv6 time-to-live count exceeded packets.

Follow these steps to enable sending of ICMPv6 time exceeded packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable sending of ICMPv6 time exceeded packets	**ipv6 hoplimit-expires enable**	Optional Enabled by default.

Configuring IPv6 DNS Client

Configuring Static IPv6 Domain Name Resolution

Configuring static IPv6 domain name resolution is to establish the mapping between a host name and an IPv6 address. When using such applications as Telnet, you can directly input a host name and the system will resolve the host name into an IPv6 address. Each host name can correspond to only one IPv6 address.

Follow these steps to configure static IPv6 domain name resolution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a host name to IPv6 address mapping	**ipv6 host** *hostname ipv6-address*	Required

Configuring Dynamic IPv6 Domain Name Resolution

You can use the following command to enable the dynamic domain name resolution function. In addition, you need to configure a DNS server so that a query request message can be sent to the correct server for resolution. The system can support at most six DNS servers.

You can configure a DNS suffix so that you only need to enter part of a domain name, and the system can automatically add the preset suffix for address resolution. The system can support at most 10 DNS suffixes.

Follow these steps to configure dynamic IPv6 domain name resolution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the dynamic domain name resolution function	**dns resolve**	Required Disabled by default.
Configure an IPv6 DNS server	**dns server ipv6** *ipv6-address* [*interface-type interface-number*]	Required If the IPv6 address of the DNS server is a link-local address, you need to specify the *interface-type* and *interface-number* argument.
Configure a DNS suffix	**dns domain** *domain-name*	Required By default, no domain name suffix is configured, that is, the domain name is resolved according to the input information.

Note

The **dns resolve** and **dns domain** commands are the same as those of IPv4 DNS. For details about the commands, refer to *DNS Commands* in the *IP Services Volume*.

Displaying and Maintaining IPv6 Basics Configuration

To do…	Use the command…	Remarks
Display DNS suffix information	**display dns domain** [**dynamic**]	Available in any view
Display IPv6 dynamic domain name cache information	**display dns ipv6 dynamic-host**	
Display IPv6 DNS server information	**display dns ipv6 server** [**dynamic**]	
Display the IPv6 FIB entries	**display ipv6 fib** [*slot-number*] [*ipv6-address*]	
Display the host name to IPv6 address mappings in the static DNS database	**display ipv6 host**	
Display the IPv6 interface settings	**display ipv6 interface** [*interface-type* [*interface-number*]] [**verbose**]	
Display neighbor information	**display ipv6 neighbors** { { *ipv6-address* | **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* | **vlan** *vlan-id* } [**|** { **begin** | **exclude** | **include** } *regular-expression*]	
Display the total number of neighbor entries satisfying the specified conditions	**display ipv6 neighbors** { { **all** | **dynamic** | **static** } [**slot** *slot-number*] | **interface** *interface-type interface-number* | **vlan** *vlan-id* } **count**	
Display the PMTU information of an IPv6 address	**display ipv6 pathmtu** { *ipv6-address* | **all** | **dynamic** | **static** }	
Display socket information	**display ipv6 socket** [**socktype** *socket-type*] [*task-id socket-id*] [**slot** *slot-number*]	
Display the statistics of IPv6 packets and ICMPv6 packets	**display ipv6 statistics** [**slot** *slot-number*]	
Display the IPv6 TCP connection statistics	**display tcp ipv6 statistics**	
Display the IPv6 TCP connection status information	**display tcp ipv6 status**	
Display the IPv6 UDP connection statistics	**display udp ipv6 statistics**	
Clear IPv6 dynamic domain name cache information	**reset dns ipv6 dynamic-host**	Available in user view
Clear IPv6 neighbor information	**reset ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **slot** *slot-number* | **static** }	
Clear the specified PMTU values	**reset ipv6 pathmtu** { **all** | **static** | **dynamic**}	
Clear the statistics of IPv6 and ICMPv6 packets	**reset ipv6 statistics** [**slot** *slot-number*]	
Clear all IPv6 TCP connection statistics	**reset tcp ipv6 statistics**	
Clear the statistics of all IPv6 UDP packets	**reset udp ipv6 statistics**	

Note

The **display dns domain** command is the same as the one of IPv4 DNS. For details about the commands, refer to *DNS Commands* in the *IP Services Volume*.

IPv6 Configuration Example

Network requirements

- Host, Switch A and Switch B are directly connected through Ethernet ports. Add the Ethernet ports into corresponding VLANs, configure IPv6 addresses for the VLAN interfaces and verify the connectivity between them.
- The aggregatable global unicast addresses of VLAN-interface 2 and VLAN-interface 1 on Switch A are 3001::1/64 and 2001::1/64 respectively.
- The aggregatable global unicast address of VLAN-interface 2 on Switch B is 3001::2/64, and a route to Host is available.
- IPv6 is enabled for Host to automatically get an IPv6 address through IPv6 NDP, and a route to Switch B is available.

Figure 1-6 Network diagram for IPv6 address configuration

Note

The VLAN interfaces have been created on the switch.

Configuration procedure

- Configure Switch A

Enable IPv6.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Specify an aggregatable global unicast address for VLAN-interface 2.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address 3001::1/64
[SwitchA-Vlan-interface2] quit
```

Specify an aggregatable global unicast address for VLAN-interface 1, and allow it to advertise RA messages (no interface advertises RA messages by default).

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ipv6 address 2001::1/64
[SwitchA-Vlan-interface1] undo ipv6 nd ra halt
```

- Configure Switch B

Enable IPv6.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure an aggregatable global unicast address for VLAN-interface 2.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address 3001::2/64
```

Configure an IPv6 static route with destination IP address 2001::/64 and next hop address 3001::1.

```
[SwitchB-Vlan-interface2] ipv6 route-static 2001:: 64 3001::1
```

- Configure Host

Enable IPv6 for Host to automatically get an IPv6 address through IPv6 NDP.

```
[SwitchA-Vlan-interface1] display ipv6 neighbors interface gigabitethernet 1/0/2
                Type: S-Static    D-Dynamic
IPv6 Address                Link-layer      VID  Interface    State T Age
FE80::215:E9FF:FEA6:7D14    0015-e9a6-7d14   1   GE1/0/2      STALE D 1238
2001::15B:E0EA:3524:E791    0015-e9a6-7d14   1   GE1/0/2      STALE D 1248
```

The above information shows that the IPv6 aggregatable global unicast address that Host obtained is 2001::15B:E0EA:3524:E791.

Verification

Display the IPv6 interface settings on Switch A.

```
[SwitchA-Vlan-interface1] display ipv6 interface vlan-interface 2 verbose
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:2
  Global unicast address(es):
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FF00:2
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   25829
  InTooShorts:                  0
  InTruncatedPkts:              0
  InHopLimitExceeds:            0
  InBadHeaders:                 0
  InBadOptions:                 0
```

```
  ReasmReqds:                   0
  ReasmOKs:                     0
  InFragDrops:                  0
  InFragTimeouts:               0
  OutFragFails:                 0
  InUnknownProtos:              0
  InDelivers:                   47
  OutRequests:                  89
  OutForwDatagrams:             48
  InNoRoutes:                   0
  InTooBigErrors:               0
  OutFragOKs:                   0
  OutFragCreates:               0
  InMcastPkts:                  6
  InMcastNotMembers:            25747
  OutMcastPkts:                 48
  InAddrErrors:                 0
  InDiscards:                   0
  OutDiscards:                  0
[SwitchA-Vlan-interface1] display ipv6 interface vlan-interface 1 verbose
Vlan-interface1 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1C0
  Global unicast address(es):
    2001::1, subnet is 2001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FF00:1C0
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   272
  InTooShorts:                  0
  InTruncatedPkts:              0
  InHopLimitExceeds:            0
  InBadHeaders:                 0
  InBadOptions:                 0
  ReasmReqds:                   0
```

```
  ReasmOKs:                       0
  InFragDrops:                    0
  InFragTimeouts:                 0
  OutFragFails:                   0
  InUnknownProtos:                0
  InDelivers:                     159
  OutRequests:                    1012
  OutForwDatagrams:               35
  InNoRoutes:                     0
  InTooBigErrors:                 0
  OutFragOKs:                     0
  OutFragCreates:                 0
  InMcastPkts:                    79
  InMcastNotMembers:              65
  OutMcastPkts:                   938
  InAddrErrors:                   0
  InDiscards:                     0
  OutDiscards:                    0
```

Display the IPv6 interface settings on Switch B.

```
[SwitchB-Vlan-interface2] display ipv6 interface vlan-interface 2 verbose
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1234
  Global unicast address(es):
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:0
    FF02::1:FF00:2
    FF02::1:FF00:1234
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                     117
  InTooShorts:                    0
  InTruncatedPkts:                0
  InHopLimitExceeds:              0
  InBadHeaders:                   0
  InBadOptions:                   0
  ReasmReqds:                     0
  ReasmOKs:                       0
  InFragDrops:                    0
  InFragTimeouts:                 0
```

```
  OutFragFails:                 0
  InUnknownProtos:              0
  InDelivers:                   117
  OutRequests:                  83
  OutForwDatagrams:             0
  InNoRoutes:                   0
  InTooBigErrors:               0
  OutFragOKs:                   0
  OutFragCreates:               0
  InMcastPkts:                  28
  InMcastNotMembers:            0
  OutMcastPkts:                 7
  InAddrErrors:                 0
  InDiscards:                   0
  OutDiscards:                  0
```

Ping Switch A and Switch B on Host, and ping Switch A and Host on Switch B to verify the connectivity between them.

Caution

When you ping a link-local address, you should use the “–i” parameter to specify an interface for the link-local address.

```
[SwitchB-Vlan-interface2] ping ipv6 -c 1 3001::1
  PING 3001::1 : 56  data bytes, press CTRL_C to break
    Reply from 3001::1
    bytes=56 Sequence=1 hop limit=64  time = 2 ms

  --- 3001::1 ping statistics ---
    1 packet(s) transmitted
    1 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 2/2/2 ms
[SwitchB-Vlan-interface2] ping ipv6 -c 1 2001::15B:E0EA:3524:E791
  PING 2001::15B:E0EA:3524:E791 : 56  data bytes, press CTRL_C to break
    Reply from 2001::15B:E0EA:3524:E791
    bytes=56 Sequence=1 hop limit=63  time = 3 ms

  --- 2001::15B:E0EA:3524:E791 ping statistics ---
    1 packet(s) transmitted
    1 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 3/3/3 ms
```

As shown in the output information, Host can ping Switch B and Switch A.

Troubleshooting IPv6 Basics Configuration

Symptom

The peer IPv6 address cannot be pinged.

Solution

- Use the **display current-configuration** command in any view or the **display this** command in system view to verify that IPv6 is enabled.
- Use the **display ipv6 interface** command in any view to verify that the IPv6 address of the interface is correct and the interface is up.
- Use the **debugging ipv6 packet** command in user view to enable the debugging for IPv6 packets to help locate the cause.

Table of Contents

1 Dual Stack Configuration

When configuring dual stack, go to these sections for information you are interested in:

- Dual Stack Overview
- Configuring Dual Stack

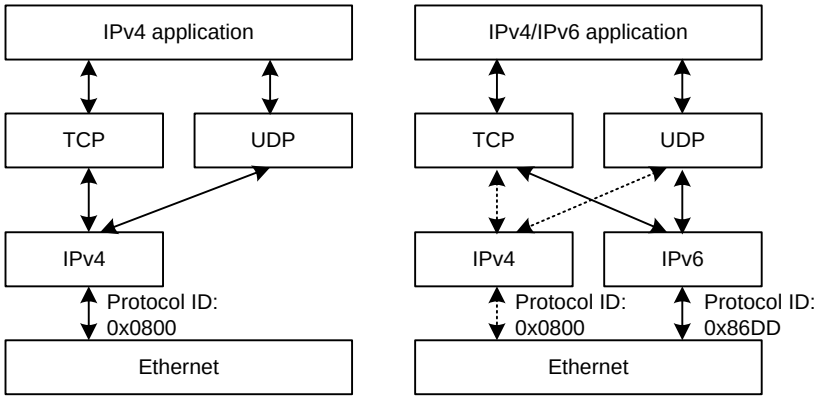
Dual Stack Overview

Dual stack is the most direct approach to making IPv6 nodes compatible with IPv4 nodes. The best way for an IPv6 node to be compatible with an IPv4 node is to maintain a complete IPv4 stack. A network node that supports both IPv4 and IPv6 is called a dual stack node. A dual stack node configured with an IPv4 address and an IPv6 address can have both IPv4 and IPv6 packets transmitted.

For an upper layer application supporting both IPv4 and IPv6, either TCP or UDP can be selected at the transport layer, while IPv6 stack is preferred at the network layer.

Figure 1-1 illustrates the IPv4/IPv6 dual stack in relation to the IPv4 stack.

Figure 1-1 IPv4/IPv6 dual stack in relation to IPv4 stack (on Ethernet)



Configuring Dual Stack

You must enable the IPv6 packet forwarding function before dual stack. Otherwise, the device cannot forward IPv6 packets even if IPv6 addresses are configured for interfaces.

Follow these steps to configure dual stack on a gateway:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the IPv6 packet forwarding function	**ipv6**	Required Disabled by default.
Enter interface view	**interface** *interface-type interface-number*	—

To do…			Use the command…	Remarks
Configure an IPv4 address for the interface			**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Required By default, no IP address is configured.
Configure an IPv6 address on the interface	Configure an IPv6 global unicast address or site-local address	Manually specify an IPv6 address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* }	Use either command. By default, no site-local address or global unicast address is configured on an interface.
		Configure an IPv6 address in the EUI-64 format	**ipv6 address** *ipv6-address/prefix-length* **eui-64**	
	Configure an IPv6 link-local address	Automatically create an IPv6 link-local address	**ipv6 address auto link-local**	Optional By default, after you configured an IPv6 site-local address or global unicast address, a link local address is automatically created.
		Manually specify an IPv6 link-local address	**ipv6 address** *ipv6-address* **link-local**	

Note

- For information about IPv4 addressing, refer to *IP Addressing Configuration* in the *IP Services Volume*.
- For more information about IPv6 address, refer to *IPv6 Basics Configuration* in the *IP Services Volume*.
- For how to enable IPv6 and configure an IPv6 address on an interface, refer to *IPv6 Basics Commands* in the *IP Services Volume*.

Table of Contents

1 Tunneling Configuration

When configuring tunneling, go to these sections for information you are interested in:

Note

The tunnel interface number is in the A/B/C format, where A, B, and C represent the stack member device ID, the sub-slot number, and the tunnel interface number respectively. The value ranges of A and B vary with devices. C is in the range of 0 to 126.

Introduction to Tunneling

The expansion of the Internet results in scarce IPv4 addresses. The technologies such as temporary IPv4 address allocation and Network Address Translation (NAT) relieve the problem of IPv4 address shortage to some extent. However, these technologies not only increase the overhead in address resolution and processing, but also lead to upper-layer application failures. Furthermore, they will still face the problem that IPv4 addresses will eventually be used up. Internet Protocol Version 6 (IPv6) adopting the 128-bit addressing scheme completely solves the above problem. Since significant improvements have been made in address space, security, network management, mobility, and QoS, IPv6 becomes one of the core standards for the next generation Internet protocol. IPv6 is compatible with all protocols except IPv4 in the TCP/IP suite. Therefore, IPv6 can completely take the place of IPv4.

Before IPv6 becomes the dominant protocol, networks using the IPv6 protocol stack are expected to communicate with the Internet using IPv4. Therefore, an IPv6-IPv4 interworking technology must be developed to ensure the smooth transition from IPv4 to IPv6. In addition, the interworking technology should provide efficient, seamless information transfer. The Internet Engineering Task Force (IETF) sets up the next generation transition (NGTRANS) working group to study problems about IPv4-to-IPv6 transition and efficient, seamless IPv4-IPv6 interworking. Currently, multiple transition technologies and interworking solutions are available. With their own characteristics, they are used to solve communication problems in different transition stages under different environments.

Currently, there are three major transition technologies: dual stack (RFC 2893), tunneling (RFC 2893), and NAT-PT (RFC 2766).

Tunneling is an encapsulation technology, which utilizes one network protocol to encapsulate packets of another network protocol and transfer them over the network. A tunnel is a virtual point-to-point

connection. In practice, the virtual interface that supports only point-to-point connections is called tunnel interface. One tunnel provides one channel to transfer encapsulated packets. Packets can be encapsulated and decapsulated at both ends of a tunnel. Tunneling refers to the whole process from data encapsulation to data transfer to data decapsulation.

Note

- For related configuration about the dual protocol stack, refer to *Dual Stack Configuration* in the *IP Services Volume*.
- The 3Com Switches 4800G do not support NAT-PT.

IPv6 over IPv4 Tunnel

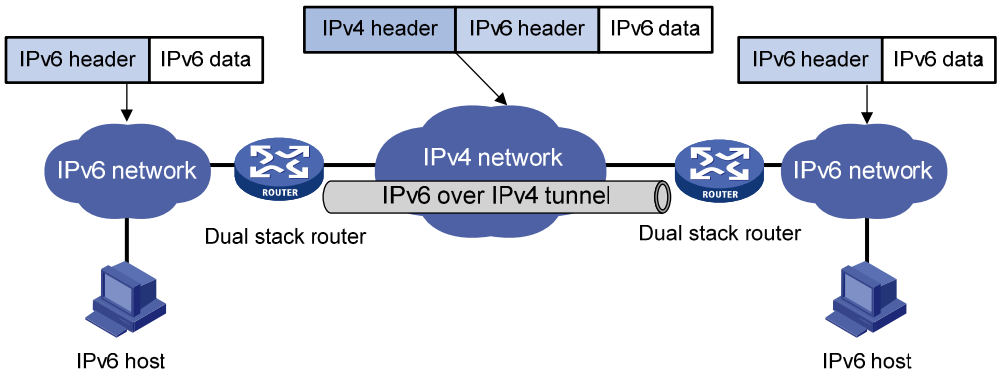
Implementation

The IPv6 over IPv4 tunneling mechanism encapsulates an IPv4 header in IPv6 data packets so that IPv6 packets can pass an IPv4 network through a tunnel to realize interworking between isolated IPv6 networks, as shown in Figure 1-1.

Caution

The devices at both ends of an IPv6 over IPv4 tunnel must support IPv4/IPv6 dual stack.

Figure 1-1 IPv6 over IPv4 tunnel



The IPv6 over IPv4 tunnel processes packets in the following way:

1) A host in the IPv6 network sends an IPv6 packet to the device at the source end of the tunnel.
2) After determining according to the routing table that the packet needs to be forwarded through the tunnel, the device at the source end of the tunnel encapsulates the IPv6 packet with an IPv4 header and forwards it through the physical interface of the tunnel.

3) The encapsulated packet goes through the tunnel to reach the device at the destination end of the tunnel. The device at the destination end decapsulates the packet if the destination address of the encapsulated packet is the device itself.
4) The destination device forwards the packet according to the destination address in the decapsulated IPv6 packet. If the destination address is the device itself, the device forwards the IPv6 packet to the upper-layer protocol for processing.

Configured tunnel and automatic tunnel

An IPv6 over IPv4 tunnel can be established between hosts, between hosts and devices, and between devices. The tunnel destination needs to forward packets if the tunnel destination is not the final destination of the IPv6 packet.

Tunnels are divided into configured tunnels and automatic tunnels depending on how the IPv4 address of the tunnel destination is acquired.

- If the destination address of an IPv6 over IPv4 tunnel cannot be acquired from the destination address of IPv6 packets, it needs to be configured manually. Such a tunnel is called a configured tunnel.
- If the interface address of an IPv6 over IPv4 tunnel has an IPv4 address embedded into an IPv6 address, the IPv4 address of the tunnel destination can be acquired automatically. Such a tunnel is called an automatic tunnel.

Type

According to the way an IPv6 packet is encapsulated, IPv6 over IPv4 tunnels are divided into the following types:

Tunnel type	Tunnel mode
Manually configured tunnel	IPv6 manual tunnel
Automatic tunnel	6to4 tunnel
	Intra-site automatic tunnel addressing protocol (ISATAP) tunnel

The configuration parameters for each tunnel mode are listed in the following table:

Tunnel mode	Source/destination IP address of the tunnel	IP address of the tunnel interface
IPv6 manual tunnel	The source/destination IP address is a manually configured IPv4 address.	IPv6 address
6to4 tunnel	The source IP address is a manually configured IPv4 address, while the destination IP address does not need to be configured.	6to4 address, in the format of 2002:IPv4-source-address::/48
ISATAP tunnel	The source IP address is a manually configured IPv4 address, while the destination IP address does not need to be configured.	ISATAP address, in the format of Prefix:0:5EFE:IPv4-source-address/64

1) IPv6 manually configured tunnel

A manually configured tunnel is a point-to-point link. Each link is a separate tunnel. IPv6 manually configured tunnels are mainly used to provide stable connections for regular secure communication between border routers or between border routers and hosts for access to remote IPv6 networks.

2) 6to4 tunnel

An automatic 6to4 tunnel is a point-to-multipoint tunnel and is used to connect multiple isolated IPv6 networks over an IPv4 network to remote IPv6 networks. The embedded IPv4 address in an IPv6 address is used to automatically acquire the destination IPv4 address of the tunnel.

The automatic 6to4 tunnel adopts 6to4 addresses. The address format is 2002:abcd:efgh:subnet number::interface ID/64, where 2002 represents the fixed IPv6 address prefix, and abcd:efgh represents the 32-bit globally unique source IPv4 address of the 6to4 tunnel, in hexadecimal notation. For example, 1.1.1.1 can be represented by 0101:0101. The part that follows 2002:abcd:efgh uniquely identifies a host in a 6to4 network. The tunnel destination is automatically determined by the embedded IPv4 address, which makes it easy to create a 6to4 tunnel.

Because the 16-bit subnet number of the 64-bit address prefix in 6to4 addresses can be customized and the first 48 bits in the address prefix are fixed to a permanent value and the IPv4 address of the tunnel source or destination, it is possible that IPv6 packets can be forwarded by the tunnel.

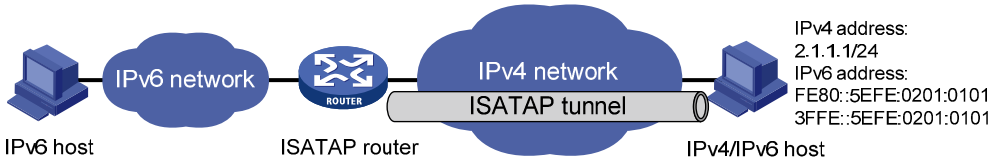
3) ISATAP tunnel

With the application of the IPv6 technology, there will be more and more IPv6 hosts in the existing IPv4 network. The ISATAP tunneling technology provides a satisfactory solution for IPv6 application. An ISATAP tunnel is a point-to-point automatic tunnel. The destination of a tunnel can automatically be acquired from the embedded IPv4 address in the destination address of an IPv6 packet.

When an ISATAP tunnel is used, the destination address of an IPv6 packet and the IPv6 address of a tunnel interface both adopt special ISATAP addresses. The ISATAP address format is prefix(64bit):0:5EFE:ip-address. The 64-bit prefix is the prefix of a valid IPv6 unicast address, while ip-address is a 32-bit source IPv4 address in the form of a.b.c.d or abcd:efgh, which need not be globally unique. Through the embedded IPv4 address, an ISATAP tunnel can automatically be created to transfer IPv6 packets.

The ISATAP tunnel is mainly used for connection between IPv6 routers or between a host and an IPv6 router over an IPv4 network.

Figure 1-2 Principle of ISATAP tunnel



Protocols and Standards

RFC 1853: IP in IP Tunneling

RFC 2473: Generic Packet Tunneling in IPv6 Specification

RFC 2893: Transition Mechanisms for IPv6 Hosts and Routers

RFC 3056: Connection of IPv6 Domains via IPv4 Clouds

RFC 4214: Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)

Tunneling Configuration Task List

Complete the following tasks to configure the tunneling feature:

Task		Remarks
Configuring IPv6 over IPv4 tunnel	Configuring IPv6 Manual Tunnel	Optional
	Configuring 6to4 Tunnel	Optional
	Configuring ISATAP Tunnel	Optional

Configuring IPv6 Manual Tunnel

Configuration Prerequisites

- Configure IP addresses for interfaces (such as the VLAN interface and loopback interface) on the device to ensure normal communication.
- Specify one of the above interfaces as the source interface of the tunnel.
- Ensure reachability between the tunnel source and destination addresses.

Configuration Procedure

Follow these steps to configure an IPv6 manual tunnel:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable IPv6		**ipv6**	Required By default, the IPv6 packet forwarding function is disabled.
Create a tunnel interface and enter tunnel interface view		**interface tunnel** *number*	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	Configure a global unicast IPv6 address or a site-local address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-addressIprefix-length* }	Required Use either command. By default, no IPv6 global unicast address or site-local address is configured for the tunnel interface.
		ipv6 address *ipv6-addressIprefix-length* **eui-64**	
	Configure a link-local IPv6 address	**ipv6 address auto link-local**	Optional By default, a link-local address will automatically be created when an IPv6 global unicast address or site-local address is configured.
		ipv6 address *ipv6-address* **link-local**	
Specify the IPv6 manual tunnel mode		**tunnel-protocol ipv6-ipv4**	Required By default, the tunnel is an IPv6 manual tunnel. The same tunnel mode should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.

To do…	Use the command…	Remarks
Configure a source address or interface for the tunnel	**source** { *ip-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Configure a destination address for the tunnel	**destination** *ip-address*	Required By default, no destination address is configured for the tunnel.
Reference a service loopback group	**service-loopback-group** *number*	Required By default, the tunnel does not reference any service loopback group.

Caution

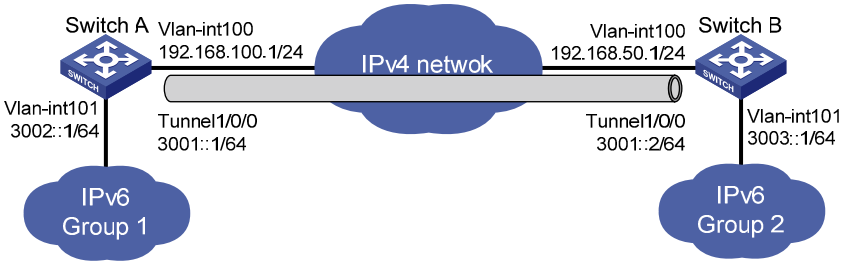
- After a tunnel interface is deleted, all the above features configured on the tunnel interface will be deleted.
- If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a forwarding route through the tunnel to the peer must be configured so that the encapsulated packet can be forwarded normally. You need to configure static or dynamic routing at both ends of the tunnel. For detailed configuration, refer to *Static Routing Configuration* or other routing protocol configuration in the *IP Routing Volume*.
- When you configure a static route at one tunnel end, you need to configure a route to the destination IPv6 address of the packet, instead of the IPv4 address of the tunnel destination, and set the outbound interface to the tunnel interface at the local end or set the next-hop to the tunnel interface at the peer end. The similar configuration needs to be performed at the other tunnel end.
- When you configure dynamic routing at both tunnel ends, you need to enable the dynamic routing protocol on the tunnel interfaces. For related configurations, refer to related contents in the *IP Routing Volume*.
- To reference a service loopback group ID on the tunnel interface to receive and send packets, you must have configured the service loopback group. Otherwise, the tunnel interface will not be up to communicate. For creation and configuration of a service loopback group, refer to *Service Loopback Group Configuration* in the *Access Volume.*

Configuration Example

Network requirements

As shown in Figure 1-3, two IPv6 networks are connected to an IPv4 network through Switch A and Switch B respectively. Configure an IPv6 manual tunnel between Switch A and Switch B to make the two IPv6 networks reachable to each other.

Figure 1-3 Network diagram for an IPv6 manual tunnel



Configuration procedure

Note

Make sure that Switch A and Switch B have the corresponding VLAN interfaces created and are reachable to each other.

- Configuration on Switch A

Enable IPv6.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.100.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 3002::1 64
[SwitchA-Vlan-interface101] quit
```

Create a service loopback group. Note that you need to disable STP on a port before adding it to a service loopback group.

```
[SwitchA] service-loopback group 1 type tunnel
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] stp disable
[SwitchA-GigabitEthernet1/0/1] port service-loopback group 1
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure a manual IPv6 tunnel.

```
[SwitchA] interface tunnel 1/0/0
[SwitchA-Tunnel1/0/0] ipv6 address 3001::1/64
[SwitchA-Tunnel1/0/0] source vlan-interface 100
[SwitchA-Tunnel1/0/0] destination 192.168.50.1
[SwitchA-Tunnel1/0/0] tunnel-protocol ipv6-ipv4
```

Reference service loopback group 1 in tunnel interface view.

```
[SwitchA-Tunnel1/0/0] service-loopback-group 1
[SwitchA-Tunnel1/0/0] quit
```

Configure a static route to IPv6 Group 2 through tunnel 1/0/0 on Switch A.

```
[SwitchA] ipv6 route-static 3003:: 64 tunnel 1/0/0
```

- Configuration on Switch B

Enable IPv6.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.168.50.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 3003::1 64
[SwitchA-Vlan-interface101] quit
```

Create a service loopback group. Note that you need to disable STP on a port before adding it to a service loopback group.

```
[SwitchB] service-loopback group 1 type tunnel
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] stp disable
[SwitchB-GigabitEthernet1/0/1] port service-loopback group 1
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure an IPv6 manual tunnel.

```
[SwitchB] interface tunnel 1/0/0
[SwitchB-Tunnel1/0/0] ipv6 address 3001::2/64
[SwitchB-Tunnel1/0/0] source vlan-interface 100
[SwitchB-Tunnel1/0/0] destination 192.168.100.1
[SwitchB-Tunnel1/0/0] tunnel-protocol ipv6-ipv4
```

Reference service loopback group 1 in tunnel interface view.

```
[SwitchB-Tunnel1/0/0] service-loopback-group 1
[SwitchB-Tunnel1/0/0] quit
```

Configure a static route to IPv6 Group 1 through tunnel 1/0/0 on Switch B.

```
[SwitchB] ipv6 route-static 3002:: 64 tunnel 1/0/0
```

Configuration verification

After the above configurations, display the status of the tunnel interfaces on Switch A and Switch B, respectively.

```
[SwitchA] display ipv6 interface tunnel 1/0/0 verbose
Tunnel1/0/0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:6401
```

```
  Global unicast address(es):
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FFA8:6401
    FF02::1:FF00:1
    FF02::1:FF00:0
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   55
...
[SwitchB] display ipv6 interface tunnel 1/0/0 verbose
Tunnel1/0/0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:3201
  Global unicast address(es):
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FFA8:3201
    FF02::1:FF00:1
    FF02::1:FF00:0
    FF02::2
    FF02::1
  MTU is 1480 bytes
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
IPv6 Packet statistics:
  InReceives:                   55
...
```

Ping the IPv6 address of VLAN-interface 101 at the peer end from Switch A.

```
[SwitchA] ping ipv6 3003::1
  PING 3003::1 : 56  data bytes, press CTRL_C to break
    Reply from 3003::1
    bytes=56 Sequence=1 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=2 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=3 hop limit=64  time = 1 ms
    Reply from 3003::1
    bytes=56 Sequence=4 hop limit=64  time = 1 ms
    Reply from 3003::1
```

```
  bytes=56 Sequence=5 hop limit=64  time = 1 ms

--- 3003::1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/1/1 ms
```

Configuring 6to4 Tunnel

Configuration Prerequisites

- Configure IP addresses for interfaces (such as the VLAN interface and loopback interface) on the device to ensure normal communication.
- Specify one of the above interfaces as the source interface of the tunnel.
- Ensure reachability between the tunnel source and destination addresses.

Configuration Procedure

Follow these steps to configure a 6to4 tunnel:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable IPv6		**ipv6**	Required By default, the IPv6 packet forwarding function is disabled.
Create a tunnel interface and enter tunnel interface view		**interface tunnel** *number*	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	Configure an IPv6 global unicast address or a site-local address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }	Required. Use either command. By default, no IPv6 global unicast address or site-local address is configured for the tunnel interface.
		ipv6 address *ipv6-address*/*prefix-length* **eui-64**	
	Configure an IPv6 link-local address	**ipv6 address auto link-local**	Optional By default, a link-local address will automatically be generated when an IPv6 global unicast address or site-local address is configured.
		ipv6 address *ipv6-address* **link-local**	
Set a 6to4 tunnel		**tunnel-protocol ipv6-ipv4 6to4**	Required By default, the tunnel is an IPv6 manual tunnel. The same tunnel mode should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.

To do…	Use the command…	Remarks
Configure a source address or interface for the tunnel	**source** { *ip-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Reference a service loopback group	**service-loopback-group** *number*	Required By default, no service loopback group is referenced.

Caution

- No destination address needs to be configured for a 6to4 tunnel because the destination address can automatically be obtained from the IPv4 address embedded in the 6to4 IPv6 address.
- If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a route to the peer must be configured so that the encapsulated packet can be forwarded normally. You can configure static or dynamic routing. Automatic tunnels do not support dynamic routing. You need to configure a route to the peer at both end of the tunnel. For the detailed configuration, refer to *Static Routing Configuration* or other routing protocol configuration in the *IP Routing Volume*.
- The automatic tunnel interfaces using the same encapsulation protocol cannot share the same source IP address.
- When you configure a static route at one tunnel end, you need to configure a route to the destination IPv6 address of the packet, instead of the IPv4 address of the tunnel destination, and set the outbound interface to the tunnel interface at the local end or set the next-hop to the tunnel interface at the peer end. The similar configuration needs to be performed at the other tunnel end.
- To reference a service loopback group on the tunnel interface to receive and send packets, you must have configured the service loopback group. Otherwise, the tunnel interface will not be up to communicate. For creation and configuration of a service loopback group, refer to *Service Loopback Group Configuration* in the *Access Volume.*

6to4 Tunnel Configuration Example

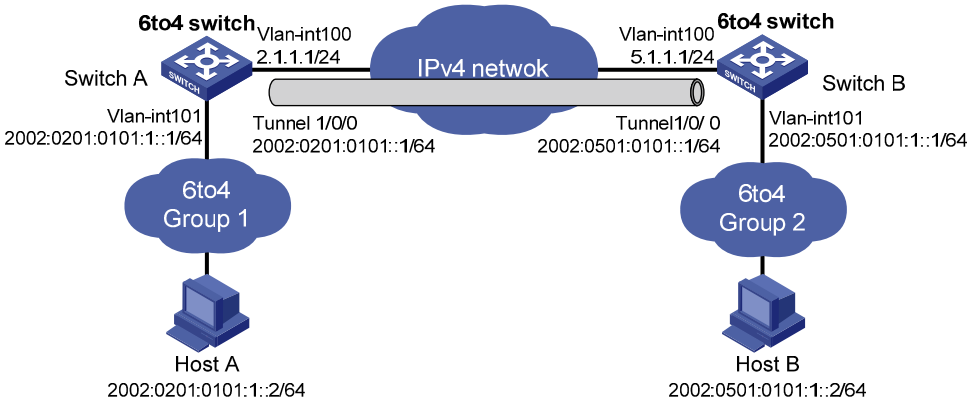
Network requirements

As shown in Figure 1-4, two 6to4 networks are connected to an IPv4 network through two 6to4 switches (Switch A and Switch B) respectively. Configure a 6to4 tunnel to make Host A and Host B reachable to each other.

To enable communication between 6to4 networks, you need to configure 6to4 addresses for 6to4 switches and hosts in the 6to4 networks.

- The IPv4 address of VLAN-interface 100 on Switch A is 2.1.1.1/24, and the corresponding 6to4 prefix is 2002:0201:0101::/48 after it is translated to an IPv6 address. Assign interface tunnel 1/0/0 to subnet 2002:0201:0101::/64 and VLAN-interface 101 to subnet 2002:0201:0101:1::/64.
- The IPv4 address of VLAN-interface 100 on Switch B is 5.1.1.1/24, and the corresponding 6to4 prefix is 2002:0501:0101::/48 after it is translated to an IPv6 address. Assign interface tunnel 1/0/0 to subnet 2002:0501:0101::/64 and VLAN-interface 101 to subnet 2002:0501:0101:1::/64.

Figure 1-4 Network diagram for a 6to4 tunnel



Configuration procedure

Note

Make sure that Switch A and Switch B have the corresponding VLAN interfaces created and are reachable to each other.

- Configuration on Switch A

Enable IPv6.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 2.1.1.1 24
[SwitchA-Vlan-interface100] quit
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 2002:0201:0101:1::1/64
[SwitchA-Vlan-interface101] quit
```

Create a service loopback group. Note that you need to disable STP on a port before adding it to a service loopback group.

```
[SwitchA] service-loopback group 1 type tunnel
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] stp disable
[SwitchA-GigabitEthernet1/0/1] port service-loopback group 1
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure a 6to4 tunnel.

```
[SwitchA] interface tunnel 1/0/0
[SwitchA-Tunnel1/0/0] ipv6 address 2002:201:101::1/64
[SwitchA-Tunnel1/0/0] source vlan-interface 100
```

```
[SwitchA-Tunnel1/0/0] tunnel-protocol ipv6-ipv4 6to4
```

Reference service loopback group 1 in tunnel interface view.

```
[SwitchA-Tunnel1/0/0] service-loopback-group 1
[SwitchA-Tunnel1/0/0] quit
```

Configure a static route whose destination address is 2002::/16 and next-hop is the tunnel interface.

```
[SwitchA] ipv6 route-static 2002:: 16 tunnel 1/0/0
```

- Configuration on Switch B

Enable IPv6.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 5.1.1.1 24
[SwitchB-Vlan-interface100] quit
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2002:0501:0101:1::1/64
[SwitchB-Vlan-interface101] quit
```

Create a service loopback group. Note that you need to disable STP on a port before adding it to a service loopback group.

```
[SwitchB] service-loopback group 1 type tunnel
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] stp disable
[SwitchB-GigabitEthernet1/0/1] port service-loopback group 1
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure the 6to4 tunnel.

```
[SwitchB] interface tunnel 1/0/0
[SwitchB-Tunnel1/0/0] ipv6 address 2002:0501:0101::1/64
[SwitchB-Tunnel1/0/0] source vlan-interface 100
[SwitchB-Tunnel1/0/0] tunnel-protocol ipv6-ipv4 6to4
```

Reference service loopback group 1 in tunnel interface view.

```
[SwitchB-Tunnel1/0/0] service-loopback-group 1
[SwitchB-Tunnel1/0/0] quit
```

Configure a static route whose destination address is 2002::/16 and the next hop is the tunnel interface.

```
[SwitchB] ipv6 route-static 2002:: 16 tunnel 1/0/0
```

Configuration verification

After the above configuration, ping Host B from Host A or ping Host A from Host B.

```
D:\>ping6 -s 2002:201:101:1::2 2002:501:101:1::2

Pinging 2002:501:101:1::2
```

```
from 2002:201:101:1::2 with 32 bytes of data:

Reply from 2002:501:101:1::2: bytes=32 time=13ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time<1ms

Ping statistics for 2002:501:101:1::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

Configuring ISATAP Tunnel

Configuration Prerequisites

- Configure IP addresses for interfaces (such as the VLAN interface and loopback interface) on the device to ensure normal communication.
- Specify one of the above interfaces as the source interface of the tunnel.
- Ensure reachability between the tunnel source and destination addresses.

Configuration Procedure

Follow these steps to configure an ISATAP tunnel:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable IPv6		**ipv6**	Required By default, the IPv6 forwarding function is disabled.
Create a tunnel interface and enter tunnel interface view		**interface tunnel** *number*	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	Configure an IPv6 global unicast address or site-local address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }	Required. Use either command. By default, no IPv6 global unicast address is configured for the tunnel interface.
		ipv6 address *ipv6-address*/*prefix-length* **eui-64**	
	Configure an IPv6 link-local address	**ipv6 address auto link-local**	Optional By default, a link-local address will automatically be generated when an IPv6 global unicast address or link-local address is configured.
		ipv6 address *ipv6-address* **link-local**	

To do…	Use the command…	Remarks
Set an ISATAP tunnel	**tunnel-protocol ipv6-ipv4 isatap**	Required By default, the tunnel is an IPv6 manual tunnel. The same tunnel mode should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or interface for the tunnel	**source** { *ip-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Reference a service loopback group	**service-loopback-group** *number*	Required By default, no service loopback group ID is referenced.

Caution

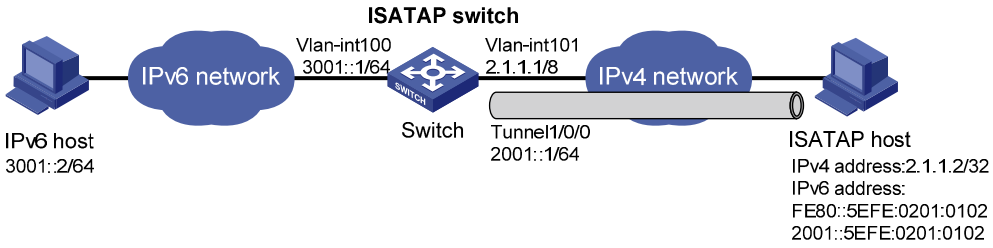
- If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a route to the peer must be configured at both ends so that the encapsulated packet can be forwarded normally. You can configure static or dynamic routing. Automatic tunnels do not support dynamic routing. For the detailed configuration, refer to *Static Routing Configuration* or other routing protocol configuration in the *IP Routing Volume*.
- The automatic tunnel interfaces using the same encapsulation protocol cannot share the same source IP address.
- When you configure a static route at one tunnel end, you need to configure a route to the destination IPv6 address of the packet, instead of the IPv4 address of the tunnel destination, and set the outbound interface to the tunnel interface at the local end or set the next-hop to the tunnel interface at the peer end. The similar configuration needs to be performed at the other tunnel end.
- To reference a service loopback group on the tunnel interface to receive and send packets, you must have configured the service loopback group. Otherwise, the tunnel interface will not be up to communicate. For creation and configuration of a service loopback group, refer to *Service Loopback Group Configuration* in the *Access Volume.*

Configuration Example

Network requirements

As shown in Figure 1-5, an IPv6 network is connected to an IPv4 network through an ISATAP switch. The destination address of the tunnel is an ISATAP address. It is required that IPv6 hosts in the IPv4 network can access the IPv6 network through the ISATAP tunnel.

Figure 1-5 Network diagram for an ISATAP tunnel



Configuration procedure

Note

- Make sure that the corresponding VLAN interfaces have been created on the switch.
- Make sure that VLAN-interface 101 on the ISATAP switch and the ISATAP host are reachable to each other.

- Configuration on the switch

Enable IPv6.

```
<Switch> system-view
[Switch] ipv6
```

Configure addresses for interfaces.

```
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] ipv6 address 3001::1/64
[Switch-Vlan-interface100] quit
[Switch] interface vlan-interface 101
[Switch-Vlan-interface101] ip address 2.1.1.1 255.0.0.0
[Switch-Vlan-interface101] quit
```

Create a service loopback group. Note that you need to disable STP on a port before adding it to a service loopback group.

```
[Switch] service-loopback group 1 type tunnel
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] stp disable
[Switch-GigabitEthernet1/0/1] port service-loopback group 1
[Switch-GigabitEthernet1/0/1] quit
```

Configure an ISATAP tunnel.

```
[Switch] interface tunnel 1/0/0
[Switch-Tunnel1/0/0] ipv6 address 2001::1/64 eui-64
[Switch-Tunnel1/0/0] source vlan-interface 101
[Switch-Tunnel1/0/0] tunnel-protocol ipv6-ipv4 isatap
```

Reference service loopback group 1 in tunnel interface view.

```
[Switch-Tunnel1/0/0] service-loopback-group 1
```

Disable the RA suppression so that hosts can acquire information such as the address prefix from the RA message released by the ISATAP switch.

```
[Switch-Tunnel1/0/0] undo ipv6 nd ra halt
[Switch-Tunnel1/0/0] quit
```

Configure a static route to the ISATAP host.

```
[Switch] ipv6 route-static 2001:: 16 tunnel 1/0/0
```

- Configuration on the ISATAP host

The specific configuration on the ISATAP host is related to its operating system. The following example shows the configuration of the host running the Windows XP.

On a Windows XP-based host, the ISATAP interface is usually interface 2. Configure the IPv4 address of the ISATAP router on the interface to complete the configuration on the host. Before doing that, display the ISATAP interface information:

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  does not use Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 0.0.0.0
  router link-layer address: 0.0.0.0
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1280 (true link MTU 65515)
  current hop limit 128
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
  DAD transmits 0
  default site prefix length 48
```

A link-local address (fe80::5efe:2.1.1.2) in the ISATAP format was automatically generated for the ISATAP interface. Configure the IPv4 address of the ISATAP switch on the ISATAP interface.

```
C:\>ipv6 rlu 2 2.1.1.1
```

After carrying out the above command, look at the information on the ISATAP interface.

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  uses Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 2.1.1.2
  router link-layer address: 2.1.1.1
    preferred global 2001::5efe:2.1.1.2, life 29d23h59m46s/6d23h59m46s (public)
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1500 (true link MTU 65515)
  current hop limit 255
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
```

```
  DAD transmits 0
  default site prefix length 48
```

By comparison, it is found that the host acquires the address prefix 2001::/64 and automatically generates the address 2001::5efe:2.1.1.2. Meanwhile, “uses Router Discovery” is displayed, indicating that the router discovery function is enabled on the host. At this time, ping the IPv6 address of the tunnel interface of the switch. If the address is successfully pinged, an ISATAP tunnel is established.

```
C:\>ping 2001::5efe:2.1.1.1

Pinging 2001::5efe:2.1.1.1 with 32 bytes of data:

Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms
Reply from 2001::5efe:2.1.1.1: time=1ms

Ping statistics for 2001::5efe:2.1.1.1:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

Configuration verification

After the above configurations, the ISATAP host can access the host in the IPV6 network.

Displaying and Maintaining Tunneling Configuration

To do…	Use the command…	Remarks
Display information about a specified tunnel interface	**display interface tunnel** [*number*]	Available in any view
Display IPv6 information related to a specified tunnel interface	**display ipv6 interface tunnel** [*number*] [**verbose**]	Available in any view

Troubleshooting Tunneling Configuration

Symptom: After the configuration of related parameters such as tunnel source address, tunnel destination address, and tunnel mode, the tunnel interface is still not up.

Solution: Follow the steps below:

1) The common cause is that the physical interface of the tunnel source is not up. Use the **display interface tunnel** or **display ipv6 interface tunnel** commands to view whether the physical interface of the tunnel source is up. If the physical interface is down, use the **debugging tunnel event** command in user view to view the cause.
2) Another possible cause is that the tunnel destination is unreachable. Use the **display ipv6 routing-table** or **display ip routing-table** command to view whether the tunnel destination is reachable. If no routing entry is available for tunnel communication in the routing table, configure related routes.

Table of Contents

1 sFlow Configuration

When configuring sFlow, go to these sections for information you are interested in:

sFlow Overview

Introduction to sFlow

Sampled Flow (sFlow) is a traffic monitoring technology mainly used to collect and analyze traffic statistics.

The sFlow system involves an sFlow agent embedded in a device and a remote sFlow collector. The sFlow agent collects traffic statistics and packets from the sFlow enabled ports on the device, encapsulates the information into sFlow packets, and sends the packets to the sFlow collector. The sFlow collector analyzes the sFlow packets and displays the results.

sFlow has the following two sampling mechanisms:

- Packet-based sampling: An sFlow enabled port samples one packet out of a configurable number of packets passing through it.
- Time-based sampling: The sFlow agent samples the statistics of all sFlow enabled ports at a configurable interval.

As a traffic monitoring technology, sFlow has the following advantages:

- Supporting traffic monitoring on Gigabit and higher-speed networks.
- Providing scalability to allow one sFlow collector to monitor multiple or more sFlow agents.
- Implementing the low-cost sFlow agent.

Note

Currently, only the sFlow agent function is supported on the device.

Operation of sFlow

sFlow operates as follows:

1) With sFlow enabled, a physical port encapsulates sampled data into packets and sends them to the sFlow agent.
2) The sFlow agent periodically collects the statistics of all sFlow enabled ports.

3) When the sFlow packet buffer overflows or the one-second timer expires, the sFlow agent sends sFlow packets to the specified sFlow collector.

Configuring sFlow

The sFlow feature enables the remote sFlow collector to monitor the network and analyze sFlow packet statistics.

Follow these steps to configure sFlow:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify the IP address of the sFlow agent	**sflow agent ip** *ip-address*	Required Not configured by default.
Specify the IP address and port number of the sFlow collector	**sflow collector ip** *ip-address* [**port** *port-num*]	Required Not specified by default.
Set the counter sampling interval at which the sFlow agent collects the statistics of sFlow enabled ports	**sflow interval** *interval-time*	Optional 20 seconds by default.
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Enable sFlow in the inbound or outbound direction	**sflow enable** { **inbound** | **outbound** }	Required Not enabled by default.
Specify the sFlow sampling mode	**sflow sampling-mode** { **determine** | **random** }	Optional **random** by default. Currently, the **determine** mode is not supported on 3Com Switch 4800G.
Specify the number of packets out of which the interface will sample a packet	**sflow sampling-rate** *rate*	Optional 200000 by default.

Caution

- The sFlow agent and sFlow collector must not have the same IP address.
- Currently, you can specify at most two sFlow collectors on the device.

Displaying and Maintaining sFlow

To do…	Use the command…	Remarks
Display sFlow configuration information	**display sflow** [**slot** *slot-id*]	Available in any view

sFlow Configuration Example

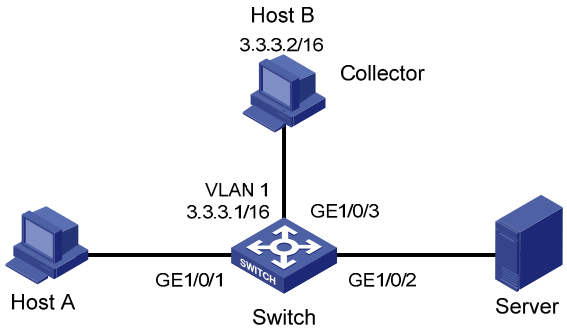
Network requirements

- Host A and Server are connected to Switch through GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 respectively.
- Host B works as an sFlow collector with IP address 3.3.3.2 and port number 6343, and is connected to Switch through GigabitEthernet 1/0/3.
- GigabitEthernet 1/0/3 belongs to VLAN 1, having an IP address of 3.3.3.1.

Run sFlow agent on Switch, and enable sFlow on GigabitEthernet 1/0/1 to monitor traffic on this interface. Switch sends sFlow packets through GigabitEthernet 1/0/3 to Host B, which then analyzes the sFlow packets and displays the results.

Network diagram

Figure 1-1 Network diagram for sFlow configuration



Configuration procedure

Configure an IP address for the sFlow agent.

```
<Switch> system-view
[Switch] sflow agent ip 3.3.3.1
```

Specify the IP address and port number of the sFlow collector.

```
[Switch] sflow collector ip 3.3.3.2
```

Set the sFlow interval to 30 seconds.

```
[Switch] sflow interval 30
```

Enable sFlow in both the inbound and outbound directions on GigabitEthernet 1/0/1.

```
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] sflow enable inbound
[Switch-GigabitEthernet1/0/1] sflow enable outbound
```

Specify the traffic sampling rate.

```
[Switch-GigabitEthernet1/0/1] sflow sampling-rate 100000
```

Display the sFlow configuration information.

```
[Switch-GigabitEthernet1/0/1] display sflow
sFlow Version: 5
sFlow Global Information:
Agent                   IP:3.3.3.1
```

```
Collector               IP:3.3.3.2  Port:6343
Interval(s): 30
sFlow Port Information:
Interface    Direction       Rate         Mode      Status
Eth1/1       In/Out          100000       Random    Active
```

Troubleshooting sFlow Configuration

The Remote sFlow Collector Cannot Receive sFlow Packets

Symptom

The remote sFlow collector cannot receive sFlow packets.

Analysis

- sFlow is not enabled globally because the sFlow agent or/and the sFlow collector is/are not specified.
- No port is enabled with sFlow to sample data.
- The IP address of the sFlow collector specified on the sFlow agent is different from that of the remote sFlow collector.
- No IP address is configured for the Layer 3 interface on the device, or the IP address is configured, but the UDP packets with the IP address being the source cannot reach the sFlow collector.
- The physical link between the device and the sFlow collector fails.

Solution

1) Check whether sFlow is correctly configured by displaying sFlow configuration with the **display sflow** command.
2) Check whether the correct IP address is configured for the device to communicate with the sFlow collector.
3) Check whether the physical link between the device and the sFlow collector is normal.

IP Routing Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The IP Routing Volume is organized as follows:

Features	Description
IP Routing Overview	This document introduces the Display commands for IP Routing Table.
Static Routing	This document introduces the commands for Static Routing.
RIP	Routing Information Protocol (RIP) is a simple Interior Gateway Protocol (IGP), mainly used in small-sized networks. This document introduces the commands for RIP configuration.
OSPF	Open Shortest Path First (OSPF) is an Interior Gateway Protocol based on the link state developed by IETF. This document introduces the commands for OSPF configuration.
IS-IS	Intermediate System-to-Intermediate System (IS-IS) is a link state protocol, which uses the shortest path first (SPF) algorithm. This document introduces the commands for IS-IS configuration.
BGP	Border gateway protocol (BGP) is an inter-autonomous system (inter-AS) dynamic route discovery protocol. This document introduces the commands for BGP configuration.
IPv6 Static Routing	Similar to IPv4 static routes, IPv6 static routes work well in simple IPv6 network environments. This document introduces the commands for IPv6 Static Routing configuration.
IPv6 RIPng	RIP next generation (RIPng) is an extension of RIP-2 for IPv4. RIPng for IPv6 is IPv6 RIPng. This document introduces the commands for RIPng configuration.
IPv6 OSPFv3	OSPFv3 is OSPF version 3 for short, supporting IPv6 and compliant with RFC2740 (OSPF for IPv6). This document introduces the commands for OSPFv3 configuration.
IPv6 IS-IS	IS-IS with IPv6 support is called IPv6 IS-IS dynamic routing protocol. This document introduces the commands for IPv6 IS-IS configuration.
IPv6 BGP	To support multiple network layer protocols, IETF extended BGP-4 by introducing IPv6 BGP. This document introduces the commands for IPv6 BGP configuration.

Features	Description
Routing Policy	Routing policy is used on the router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed. This document introduces the commands for Routing Policy configuration.
BFD	Bidirectional forwarding detection (BFD) provides a single mechanism to quickly detect and monitor the connectivity of links in networks. This document introduces the commands for BFD configuration.
MCE	Multi-CE (MCE) enables a switch to function as the CEs of multiple VPN instances in a BGP/MPLS VPN network, thus reducing the investment on network equipment. This document introduces the commands for MCE configuration.

Table of Contents

1 IP Routing Overview

Go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch.

IP Routing and Routing Table

Routing

Routing in the Internet is achieved through routers. Upon receiving a packet, a router finds an optimal route based on the destination address and forwards the packet to the next router in the path until the packet reaches the last router, which forwards the packet to the intended destination host.

Routing Table

Routing table

Routing tables play a key role in routing. Each router maintains a routing table, and each entry in the table specifies which physical interface a packet destined for a certain destination should go out to reach the next hop (the next router) or the directly connected destination.

Routes in a routing table can be divided into three categories by origin:

- Direct routes: Routes discovered by data link protocols, also known as interface routes.
- Static routes: Routes that are manually configured.
- Dynamic routes: Routes that are discovered dynamically by routing protocols.

Contents of a routing table

A routing table includes the following key items:

- Destination address: Destination IP address or destination network.
- Network mask: Specifies, in company with the destination address, the address of the destination network. A logical AND operation between the destination address and the network mask yields the address of the destination network. For example, if the destination address is 129.102.8.10 and the mask 255.255.0.0, the address of the destination network is 129.102.0.0. A network mask is made of a certain number of consecutive 1s. It can be expressed in dotted decimal format or by the number of the 1s.

- Outbound interface: Specifies the interface through which the IP packets are to be forwarded.
- IP address of the next hop: Specifies the address of the next router on the path. If only the outbound interface is configured, its address will be the IP address of the next hop.
- Priority for the route. Routes to the same destination but having different nexthops may have different priorities and be found by various routing protocols or manually configured. The optimal route is the one with the highest priority (with the smallest metric).

Routes can be divided into two categories by destination:

- Subnet routes: The destination is a subnet.
- Host routes: The destination is a host.

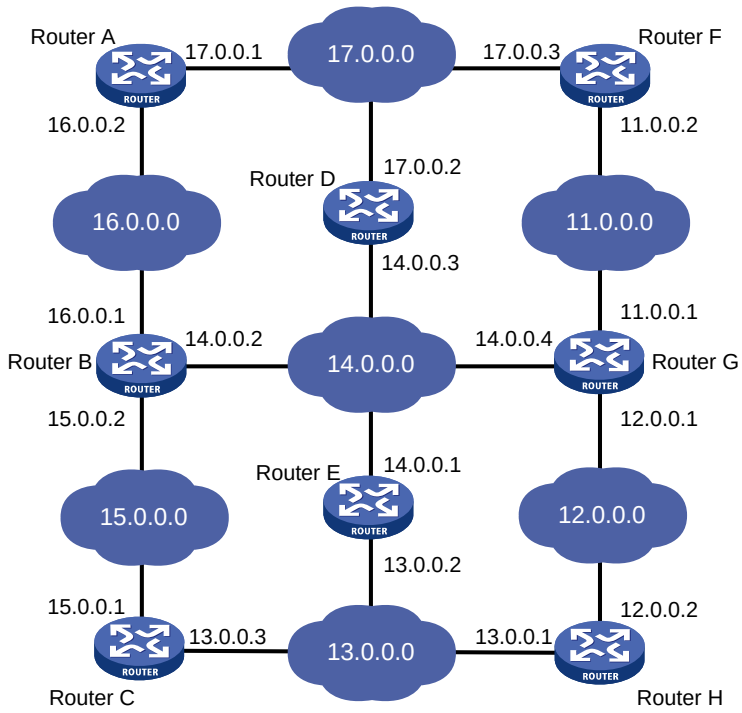
Based on whether the destination is directly connected to a given router, routes can be divided into:

- Direct routes: The destination is directly connected to the router.
- Indirect routes: The destination is not directly connected to the router.

To prevent the routing table from getting too large, you can configure a default route. All packets without matching any entry in the routing table will be forwarded through the default route.

In Figure 1-1, the IP address on each cloud represents the address of the network. Router G is connected to three networks and therefore has three IP addresses for its three physical interfaces. Its routing table is shown under the network topology.

Figure 1-1 A sample routing table



Destination Network	Nexthop	Interface
11.0.0.0	11.0.0.1	2
12.0.0.0	12.0.0.1	1
13.0.0.0	12.0.0.2	1
14.0.0.0	14.0.0.4	3
15.0.0.0	14.0.0.2	3
16.0.0.0	14.0.0.2	3
17.0.0.0	11.0.0.2	2

Routing Protocol Overview

Static Routing and Dynamic Routing

Static routing is easy to configure and requires less system resources. It works well in small, stable networks with simple topologies. Its major drawback is that you must perform routing configuration again whenever the network topology changes; it cannot adjust to network changes by itself.

Dynamic routing is based on dynamic routing protocols, which can detect network topology changes and recalculate the routes accordingly. Therefore, dynamic routing is suitable for large networks. Its disadvantages are that it is difficult to configure, and that it not only imposes higher requirements on the system, but also consumes a certain amount of network resources.

Classification of Dynamic Routing Protocols

Dynamic routing protocols can be classified based on the following standards:

Operational scope

- Interior gateway protocols (IGPs): Work within an autonomous system, including RIP, OSPF, and IS-IS.
- Exterior gateway protocols (EGPs): Work between autonomous systems. The most popular one is BGP.

Note

An autonomous system refers to a group of routers that share the same routing policy and work under the same administration.

Routing algorithm

- Distance-vector protocols: RIP and BGP. BGP is also considered a path-vector protocol.
- Link-state protocols: OSPF and IS-IS.

The main differences between the above two types of routing algorithms lie in the way routes are discovered and calculated.

Type of the destination address

- Unicast routing protocols: RIP, OSPF, BGP, and IS-IS.
- Multicast routing protocols: PIM-SM and PIM-DM.

This chapter focuses on unicast routing protocols. For information on multicast routing protocols, refer to the *IP Multicast Volume*.

Version of IP protocol

IPv4 routing protocols: RIP, OSPFv2, BGP4, and IS-IS.

IPv6 routing protocols: RIPng, OSPFv3, BGP4+, and IPv6 IS-IS.

Routing Protocols and Routing Priority

Different routing protocols may find different routes to the same destination. However, not all of those routes are optimal. In fact, at a particular moment, only one protocol can uniquely determine the current optimal route to the destination. For the purpose of route selection, each routing protocol (including static routes) is assigned a priority. The route found by the routing protocol with the highest priority is preferred.

The following table lists some routing protocols and the default priorities for routes found by them:

Routing approach	Priority
DIRECT	0
OSPF	10
IS-IS	15
STATIC	60
RIP	100
OSPF ASE	150
OSPF NSSA	150
IBGP	255
EBGP	255
UNKNOWN	256

Note

- The smaller the priority value, the higher the priority.
- The priority for a direct route is always 0, which you cannot change. Any other type of routes can have their priorities manually configured.
- Each static route can be configured with a different priority.
- IPv4 and IPv6 routes have their own respective routing tables.

Load Balancing and Route Backup

Load Balancing

In multi-route mode, a routing protocol can be configured with multiple equal-cost routes to the same destination. These routes have the same priority and will all be used to accomplish load balancing if there is no route with a higher priority available. A given routing protocol may find several routes with the same metric to the same destination, and if this protocol has the highest priority among all the active protocols, these routes will be considered valid routes for load balancing.

Note

- The number of routes for load balancing varies by device.
- In current implementations, routing protocols supporting load balancing are static routing, RIP, OSPF, BGP, and IS-IS.

Route backup

Route backup can help improve network reliability. With route backup, you can configure multiple routes to the same destination, expecting the one with the highest priority to be the main route and all the rest backup routes.

Under normal circumstances, packets are forwarded through the main route. When the main route goes down, the route with the highest priority among the backup routes is selected to forward packets. When the main route recovers, the route selection process is performed again and the main route is selected again to forward packets.

Route Recursion

The nexthops of some BGP routes (except eBGP routes) and static routes configured with nexthops may not be directly connected. To forward the packets, the outgoing interface to reach the nexthop must be available. Route recursion is used to find the outgoing interface based on the nexthop information of the route. Link-state routing protocols, such as OSPF and IS-IS, do not need route recursion because they obtain nexthop information through route calculation.

Sharing of Routing Information

As different routing protocols use different routing algorithms to calculate routes, they may find different routes. In a large network with multiple routing protocols, it is required for routing protocols to share their routing information. Each routing protocol has its own route redistribution mechanism. For detailed information, refer to the *IP Routing Volume*.

Configuring a Router ID

Some routing protocols use a router ID to identify a device. You can configure a global router ID for a device. If no router ID is configured for a protocol, the global router ID is used.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a router ID	**router id** *router-id*	Optional Not configured by default.

Displaying and Maintaining a Routing Table

To do…	Use the command…	Remarks
Display brief information about the active routes in the routing table	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] [**verbose** | **|** { **begin** | **exclude** | **include** } *regular-expression*]	Available in any view
Display information about routes to the specified destination	**display ip routing-table** *ip-address* [*mask-length* | *mask*] [**longer-match**] [**verbose**]	Available in any view
Display information about routes with destination addresses in the specified range	**display ip routing-table** *ip-address1* { *mask-length* | *mask* } *ip-address2* { *mask-length* | *mask* } [**verbose**]	Available in any view
Display information about routes permitted by an IPv4 basic ACL	**display ip routing-table acl** *acl-number* [**verbose**]	Available in any view
Display routing information permitted by an IPv4 prefix list	**display ip routing-table ip-prefix** *ip-prefix-name* [**verbose**]	Available in any view
Display routes of a routing protocol	**display ip routing-table protocol** *protocol* [**inactive** | **verbose**]	Available in any view
Display statistics about the network routing table or a VPN routing table	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **statistics**	Available in any view
Display the router ID	**display router id**	Available in any view
Clear statistics for the routing table or a VPN routing table	**reset ip routing-table statistics protocol** [**vpn-instance** *vpn-instance-name*] { **all** | *protocol* }	Available in user view
Display brief IPv6 routing table information	**display ipv6 routing-table**	Available in any view
Display verbose IPv6 routing table information	**display ipv6 routing-table verbose**	Available in any view
Display routing information for a specified destination IPv6 address	**display ipv6 routing-table** *ipv6-address prefix-length* [**longer-match**] [**verbose**]	Available in any view
Display routing information permitted by an IPv6 ACL	**display ipv6 routing-table acl** *acl6-number* [**verbose**]	Available in any view
Display routing information permitted by an IPv6 prefix list	**display ipv6 routing-table ipv6-prefix** *ipv6-prefix-name* [**verbose**]	Available in any view
Display IPv6 routing information of a routing protocol	**display ipv6 routing-table protocol** *protocol* [**inactive** | **verbose**]	Available in any view
Display IPv6 routing statistics	**display ipv6 routing-table statistics**	Available in any view
Display IPv6 routing information for an IPv6 address range	**display ipv6 routing-table** *ipv6-address1 prefix-length1 ipv6-address2 prefix-length2* [**verbose**]	Available in any view
Clear specified IPv6 routing table statistics	**reset ipv6 routing-table statistics protocol** { **all** | *protocol* }	Available in user view

Table of Contents

1 Static Routing Configuration

When configuring a static route, go to these sections for information you are interested in:

Note

The term "router" in this document refers to a router in a generic sense or a Layer 3 switch.

Introduction

Static Route

A static route is a manually configured. If a network's topology is simple, you only need to configure static routes for the network to work normally. The proper configuration and usage of static routes can improve network performance and ensure bandwidth for important network applications.

The disadvantage of using static routes is that they cannot adapt to network topology changes. If a fault or a topological change occurs in the network, the routes will be unreachable and the network breaks. In this case, the network administrator has to modify the static routes manually.

Default Route

If the destination address of a packet fails to match any entry in the routing table, the packet will be discarded.

After a default route is configured on a router, any packet whose destination IP address matches no entry in the routing table can be forwarded to a designated upstream router.

A router selects the default route only when it cannot find any matching entry in the routing table.

- If the destination address of a packet fails to match any entry in the routing table, the router selects the default route to forward the packet.
- If there is no default route and the destination address of the packet fails to match any entry in the routing table, the packet will be discarded and an ICMP packet will be sent to the source to report that the destination or the network is unreachable.

Default routes can be configured in two ways:

- The network administrator can configure a default route with both destination and mask being 0.0.0.0. The router forwards any packet whose destination address fails to match any entry in the routing table to the next hop of the default static route.
- Some dynamic routing protocols, such as OSPF, RIP and IS-IS, can also generate a default route. For example, an upstream router running OSPF can generate a default route and advertise it to other routers, which install the default route with the next hop being the upstream router.

Application Environment of Static Routing

Before configuring a static route, you need to know the following concepts:

1) Destination address and mask

In the **ip route-static** command, an IPv4 address is in dotted decimal format and a mask can be either in dotted decimal format or in the form of mask length (the digits of consecutive 1s in the mask).

2) Output interface and next hop address

While configuring a static route, you can specify either the output interface or the next hop address depending on the specific occasion. For a NULL0 or loopback interface, if the output interface has already been configured, there is no need to configure the next hop address

In fact, all the route entries must have a next hop address. When forwarding a packet, a router first searches the routing table for the route to the destination address of the packet. The system can find the corresponding link layer address and forward the packet only after the next hop address is specified. The next hop address can not be a local interface IP address; otherwise, the route configuration will not take effect.

3) Other attributes

You can configure different preferences for different static routes so that route management policies can be applied more flexibly. For example, specifying the same preference for different routes to the same destination enables load sharing, while specifying different preferences for these routes enables route backup.

Configuring a Static Route

Configuration Prerequisites

Before configuring a static route, you need to finish the following tasks:

- Configure the physical parameters for related interfaces
- Configure the link-layer attributes for related interfaces
- Configure the IP addresses for related interfaces

Configuration Procedure

Follow these steps to configure a static route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a static route	**ip route-static** *dest-address* { *mask* | *mask-length* } { *next-hop-address* **|** *interface-type interface-number next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } [**preference** *preference-value*] [**tag** *tag-value*] [*description description-text*]	Required By default, **preference** for static routes is 60, **tag** is 0, and no description information is configured.
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* **|** *mask-length* } { *next-hop-address* [**public**] | *interface-type interface-number next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	
Configure the default preference for static routes	**ip route-static default-preference** *default-preference-value*	Optional 60 by default

Note

- When configuring a static route, the static route does not take effect if you specify the next hop address first and then configure it as the IP address of a local interface.
- If you do not specify the preference when configuring a static route, the default preference will be used. Reconfiguring the default preference applies only to newly created static routes.
- You can flexibly control static routes by configuring tag values and using the tag values in the routing policy.
- If the destination IP address and mask are both configured as 0.0.0.0 with the **ip route-static** command, the route is the default route.

Detecting Reachability of the Static Route's Nexthop

If a static route fails due to a topology change or a fault, the connection will be interrupted. To improve network stability, the system needs to detect reachability of the static route's next hop and switch to a backup route once the next hop is unreachable.

There are two methods of detecting reachability of the static route's next hop. Note that only one of the two methods can be used at a time.

Detecting Nexthop Reachability Through BFD

Bidirectional forwarding detection (BFD) provides a general-purpose, standard, medium- and protocol-independent fast failure detection mechanism. It can uniformly and quickly detect the failures of the bidirectional forwarding paths between two routers for upper-layer protocols, such as routing

protocols and Multiprotocol Label Switching (MPLS). For details about BFD, refer to *BFD Configuration* in the *IP Routing Volume*.

After a static route is configured, you can enable BFD to detect the reachability of the static route's nexthop.

Network requirements

To detect the reachability of the static route's nexthop through BFD, you need to enable BFD first. For BFD configuration, refer to *BFD Configuration* in the *IP Routing Volume*.

Configuration procedure

Follow these steps to detect reachability of the static route's nexthop through BFD:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Detect reachability of the static route's nexthop through BFD	**ip route-static** *dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* [**bfd** { **control-packe**t | **echo-packet** }] [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Required Not configured by default
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* [**bfd** { **control-packet** | **echo-packet** }] [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

Note

- To implement BFD in the **control-packet** mode, the remote end must create a BFD session; otherwise the BFD function cannot work. To implement BFD in the **echo-packet** mode, the BFD function can work without the remote end needing to create any BFD session.
- If a route flap occurs, enabling BFD may worsen the flapping. Be cautious for use of this feature.

Detecting Nexthop Reachability Through Track

If you specify the nexthop but not outgoing interface when configuring a static route, you can associate the static route with a track entry to check the static route validity:

- When the track entry is positive, the static route's nexthop is reachable and the static route takes effect.
- When the track entry is negative, the static route's nexthop is unreachable and the static route is invalid. For details about track, refer to *Track Configuration* in the *System Volume*.

Network requirements

To detect the reachability of a static route's nexthop through a Track entry, you need to create a Track first. For detailed Track configuration procedure, refer to *Track Configuration* in the *System Volume*.

Configuration procedure

Follow these steps to detect the reachability of a static route's nexthop through Track:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Associate the static route with a track entry	**ip route-static** *dest-address* { *mask* **|** *mask-length* } { *next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } **track** *track-entry-number* [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Required
	ip route-static vpn-instance *s-vpn-instance-name*&<1-6> *dest-address* { *mask* **|** *mask-length* } { *next-hop-address* **track** *track-entry-number* [**public**] **| vpn-instance** *d-vpn-instance-name next-hop-address* **track** *track-entry-number* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Not configured by default

Note

- To configure this feature for an existing static route, simply associate the static route with a track entry. For a non-existent static route, configure it and associate it with a Track entry.
- If the track module uses NQA to detect the reachability of the private network static route's nexthop, the VPN instance number of the static route's nexthop must be identical to that configured in the NQA test group.
- If a static route needs route recursion, the associated track entry must monitor the nexthop of the recursive route instead of that of the static route; otherwise, a valid route may be mistakenly considered invalid.

Displaying and Maintaining Static Routes

To do…	Use the command…	Remarks
Display the current configuration information	**display current-configuration**	Available in any view
Display the brief information of the IP routing table	**display ip routing-table**	
Display the detailed information of the IP routing table	**display ip routing-table verbose**	
View information of static routes	**display ip routing-table protocol static** [**inactive** | **verbose**]	
Delete all the static routes	**delete** [**vpn-instance** *vpn-instance-name*] **static-routes all**	Available In system view

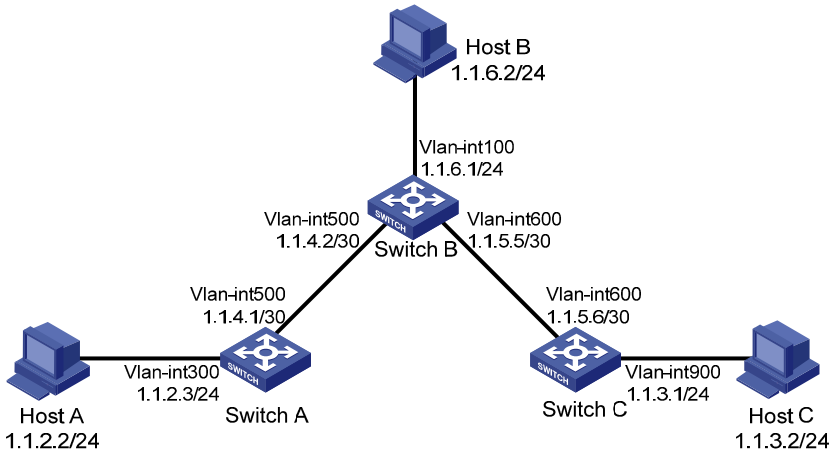
Static Route Configuration Example

Basic Static Route Configuration Example

Network requirements

The IP addresses and masks of the switches and hosts are shown in the following figure. Static routes are required for interconnection between any two hosts.

Figure 1-1 Network diagram for static route configuration



Configuration procedure

1) Configuring IP addresses for interfaces (omitted)
2) Configuring static routes

Configure a default route on Switch A.

```
<SwitchA> system-view
[SwitchA] ip route-static 0.0.0.0 0.0.0.0 1.1.4.2
```

Configure two static routes on Switch B.

```
<SwitchB> system-view
[SwitchB] ip route-static 1.1.2.0 255.255.255.0 1.1.4.1
[SwitchB] ip route-static 1.1.3.0 255.255.255.0 1.1.5.6
```

Configure a default route on Switch C

```
<SwitchC> system-view
[SwitchC] ip route-static 0.0.0.0 0.0.0.0 1.1.5.5
```

3) Configure the hosts.

The default gateways for the three hosts A, B and C are 1.1.2.3, 1.1.6.1 and 1.1.3.1 respectively. The configuration procedure is omitted.

4) Display the configuration.

Display the IP routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 7       Routes : 7
```

```
Destination/Mask    Proto  Pre  Cost         NextHop         Interface

0.0.0.0/0           Static 60   0            1.1.4.2         Vlan500
1.1.2.0/24          Direct 0    0            1.1.2.3         Vlan300
1.1.2.3/32          Direct 0    0            127.0.0.1       InLoop0
1.1.4.0/30          Direct 0    0            1.1.4.1         Vlan500
1.1.4.1/32          Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

Display the IP routing table of Switch B.

```
[SwitchB] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

1.1.2.0/24          Static 60   0            1.1.4.1         Vlan500
1.1.3.0/24          Static 60   0            1.1.5.6         Vlan600
1.1.4.0/30          Direct 0    0            1.1.4.2         Vlan500
1.1.4.2/32          Direct 0    0            127.0.0.1       InLoop0
1.1.5.4/30          Direct 0    0            1.1.5.5         Vlan600
1.1.5.5/32          Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
1.1.6.0/24          Direct 0    0            192.168.1.47    Vlan100
1.1.6.1/32          Direct 0    0            127.0.0.1       InLoop0
```

Use the **ping** command on Host B to check reachability to Host A, assuming Windows XP runs on the two hosts.

```
C:\Documents and Settings\Administrator>ping 1.1.2.2

Pinging 1.1.2.2 with 32 bytes of data:

Reply from 1.1.2.2: bytes=32 time=1ms TTL=255
Reply from 1.1.2.2: bytes=32 time=1ms TTL=255
Reply from 1.1.2.2: bytes=32 time=1ms TTL=255
Reply from 1.1.2.2: bytes=32 time=1ms TTL=255

Ping statistics for 1.1.2.2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

Use the **tracert** command on Host B to check reachability to Host A.

```
[HostB] tracert 1.1.2.2

Tracing route to 1.1.2.2 over a maximum of 30 hops
```

```
  1    <1 ms    <1 ms    <1 ms  1.1.6.1
  2    <1 ms    <1 ms    <1 ms  1.1.4.1
  3     1 ms    <1 ms    <1 ms  1.1.2.2

Trace complete.
```

Table of Contents

1 RIP Configuration

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch.

When configuring RIP, go to these sections for information you are interested in:

RIP Overview

RIP is a simple Interior Gateway Protocol (IGP), mainly used in small-sized networks, such as academic networks and simple LANs. RIP is not applicable to complex networks.

RIP is still widely used in practical networking due to easier implementation, configuration and maintenance than OSPF and IS-IS.

Operation of RIP

Introduction

RIP is a distance vector routing protocol, using UDP packets for exchanging information through port 520.

RIP uses a hop count to measure the distance to a destination. The hop count from a router to a directly connected network is 0. The hop count from a router to a directly connected router is 1. To limit convergence time, the range of RIP metric value is from 0 to 15. A metric value of 16 (or greater) is considered infinite, which means the destination network is unreachable. That is why RIP is not suitable for large-scaled networks.

RIP prevents routing loops by implementing the split horizon and poison reverse functions.

RIP routing table

A RIP router has a routing table containing routing entries of all reachable destinations, and each routing entry contains:

- Destination address: IP address of a host or a network.
- Next hop: IP address of the adjacent router’s interface to reach the destination.

- Egress interface: Packet outgoing interface.
- Metric: Cost from the local router to the destination.
- Route time: Time elapsed since the routing entry was last updated. The time is reset to 0 every time the routing entry is updated.
- Route tag: Identifies a route, used in a routing policy to flexibly control routes. For information about routing policy, refer to *Routing Policy Configuration* in the *IP Routing Volume*.

RIP timers

RIP employs four timers, update, timeout, suppress, and garbage-collect.

- The update timer defines the interval between routing updates.
- The timeout timer defines the route aging time. If no update for a route is received within the aging time, the metric of the route is set to 16 in the routing table.
- The suppress timer defines how long a RIP route stays in the suppressed state. When the metric of a route is 16, the route enters the suppressed state. In the suppressed state, only routes which come from the same neighbor and whose metric is less than 16 will be received by the router to replace unreachable routes.
- The garbage-collect timer defines the interval from when the metric of a route becomes 16 to when it is deleted from the routing table. During the garbage-collect timer length, RIP advertises the route with the routing metric set to 16. If no update is announced for that route after the garbage-collect timer expires, the route will be deleted from the routing table.

Routing loops prevention

RIP is a distance vector (D-V) routing protocol. Since a RIP router advertises its own routing table to neighbors, routing loops may occur.

RIP uses the following mechanisms to prevent routing loops.

- Counting to infinity. The metric value of 16 is defined as unreachable. When a routing loop occurs, the metric value of the route will increment to 16.
- Split horizon. A router does not send the routing information learned from a neighbor to the neighbor to prevent routing loops and save bandwidth.
- Poison reverse. A router sets the metric of routes received from a neighbor to 16 and sends back these routes to the neighbor to help delete such information from the neighbor's routing table.
- Triggered updates. A router advertises updates once the metric of a route is changed rather than after the update period expires to speed up network convergence.

Operation of RIP

The following procedure describes how RIP works.

1) After RIP is enabled, the router sends request messages to neighboring routers. Neighboring routers return Response messages including information about their routing tables.
2) After receiving such information, the router updates its local routing table, and sends triggered update messages to its neighbors. All routers on the network do the same to keep the latest routing information.
3) By default, a RIP router sends its routing table to neighbors every 30 seconds.
4) RIP ages out routes by adopting an aging mechanism to keep only valid routes.

RIP Version

RIP has two versions, RIPv1 and RIPv2.

RIPv1, a classful routing protocol, supports message advertisement via broadcast only. RIPv1 protocol messages do not carry mask information, which means it can only recognize routing information of natural networks such as Class A, B, C. That is why RIPv1 does not support discontiguous subnets.

RIPv2 is a classless routing protocol. Compared with RIPv1, RIPv2 has the following advantages.

- Supporting route tags. Route tags are used in routing policies to flexibly control routes.
- Supporting masks, route summarization and Classless Inter-Domain Routing (CIDR).
- Supporting designated next hops to select the best next hops on broadcast networks.
- Supporting multicast routing update to reduce resource consumption.
- Supporting plain text authentication and MD5 authentication to enhance security.

Note

RIPv2 has two types of message transmission: broadcast and multicast. Multicast is the default type using 224.0.0.9 as the multicast address. The interface working in the RIPv2 broadcast mode can also receive RIPv1 messages.

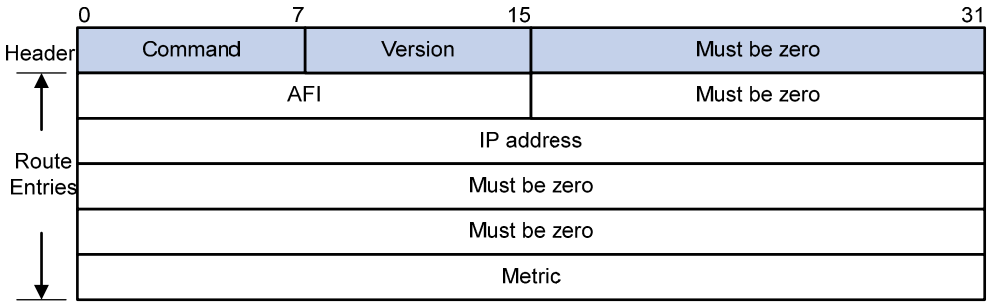
RIP Message Format

A RIPv1 message consists of a header and up to 25 route entries. (A RIPv2 authentication message uses the first route entry as the authentication entry, so it has up to 24 route entries.)

RIPv1 message format

Figure 1-1 shows the format of RIPv1 message.

Figure 1-1 RIPv1 Message Format

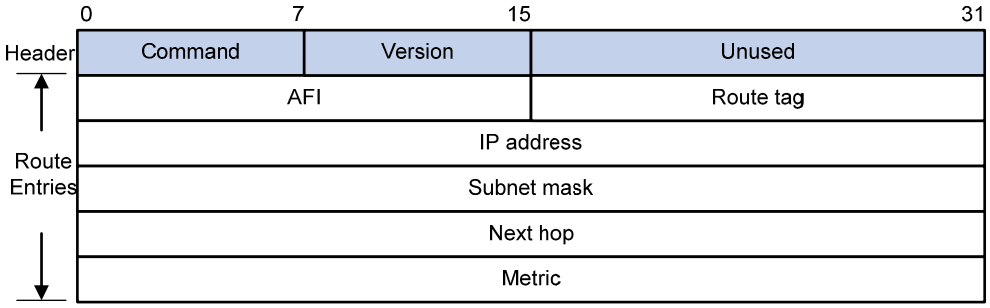


- Command: Type of message. 1 indicates request, which is used to request all or part of the routing information from the neighbor, and 2 indicates response, which contains all or part of the routing information. A response message consists of up to 25 route entries.
- Version: Version of RIP, 0x01 for RIPv1.
- Must be zero: This field must be zero.
- AFI: Address Family Identifier, 2 for IP, and 0 for request messages.
- IP Address: Destination IP address of the route. It can be a natural network, subnet or a host address.
- Metric: Cost of the route, 16 for request messages.

RIPv2 message format

The format of RIPv2 message is similar to RIPv1. Figure 1-2 shows it.

Figure 1-2 RIPv2 Message Format



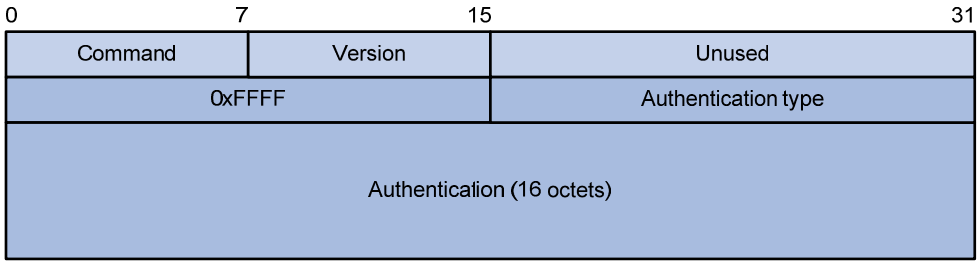
The differences from RIPv1 are stated as following.

- Version: Version of RIP. For RIPv2 the value is 0x02.
- Route Tag: Route Tag.
- IP Address: Destination IP address. It can be a natural network address, subnet address or host address.
- Subnet Mask: Mask of the destination address.
- Next Hop: If set to 0.0.0.0, it indicates that the originator of the route is the best next hop; otherwise it indicates a next hop better than the originator of the route.

RIPv2 authentication

RIPv2 sets the AFI field of the first route entry to 0xFFFF to identify authentication information. See Figure 1-3.

Figure 1-3 RIPv2 Authentication Message



- Authentication Type: A value of 2 represents plain text authentication, while a value of 3 represents MD5.
- Authentication: Authentication data, including password information when plain text authentication is adopted or including key ID, MD5 authentication data length and sequence number when MD5 authentication is adopted.

Note

- RFC 1723 only defines plain text authentication. For information about MD5 authentication, refer to RFC 2453 “RIP Version 2”.
- With RIPv1, you can configure the authentication mode in interface view. However, the configuration will not take effect because RIPv1 does not support authentication.

Supported RIP Features

The current implementation supports the following RIP features.

- RIPv1 and RIPv2
- RIP multi-instance.

Protocols and Standards

- RFC 1058: Routing Information Protocol
- RFC 1723: RIP Version 2 - Carrying Additional Information
- RFC 1721: RIP Version 2 Protocol Analysis
- RFC 1722: RIP Version 2 Protocol Applicability Statement
- RFC 1724: RIP Version 2 MIB Extension
- RFC 2082: RIPv2 MD5 Authentication
- RFC2453: RIP Version 2

Configuring RIP Basic Functions

Configuration Prerequisites

Before configuring RIP basic functions, complete the following tasks.

- Configure the link layer protocol.
- Configure an IP address on each interface, and make sure all adjacent routers are reachable to each other.

Configuration Procedure

Enabling RIP and a RIP interface

Follow these steps to enable RIP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable a RIP process and enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	Required Not enabled by default
Enable RIP on the interface attached to the specified network	**network** *network-address*	Required Disabled by default

Note

- If you make some RIP configurations in interface view before enabling RIP, those configurations will take effect after RIP is enabled.
- RIP runs only on the interfaces residing on the specified networks. Therefore, you need to specify the network after enabling RIP to validate RIP on a specific interface.
- You can enable RIP on all interfaces using the command **network** 0.0.0.0.

Configuring the interface behavior

Follow these steps to configure the interface behavior:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Disable an or all interfaces from sending routing updates (the interfaces can still receive updates)	**silent-interface** { *interface-type interface-number* | **all** }	Optional All interfaces can send routing updates by default.
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the interface to receive RIP messages	**rip input**	Optional Enabled by default
Enable the interface to send RIP messages	**rip output**	Optional Enabled by default

Configuring a RIP version

You can configure a RIP version in RIP or interface view.

- If neither global nor interface RIP version is configured, the interface sends RIPv1 broadcasts and can receive RIPv1 broadcast and unicast packets, and RIPv2 broadcast, multicast, and unicast packets.
- If an interface has no RIP version configured, it uses the global RIP version; otherwise it uses the RIP version configured on it.
- With RIPv1 configured, an interface sends RIPv1 broadcasts, and can receive RIPv1 broadcasts and RIPv1 unicasts.
- With RIPv2 configured, a multicast interface sends RIPv2 multicasts and can receive RIPv2 unicasts, broadcasts and multicasts.
- With RIPv2 configured, a broadcast interface sends RIPv2 broadcasts and can receive RIPv1 unicasts, and broadcasts, and RIPv2 broadcasts, multicasts and unicasts.

Follow these steps to configure a RIP version:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify a global RIP version	**version** { **1** | **2** }	Optional By default, if an interface has a RIP version specified, the version takes precedence over the global one. If no RIP version is specified for an interface, the interface can send RIPv1 broadcasts, and receive RIPv1 broadcasts, unicasts, RIPv2 broadcasts, multicasts and unicasts.
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify a RIP version for the interface	**rip version** { **1** | **2** [**broadcast** | **multicast**] }	Optional

Configuring RIP Route Control

In complex networks, you need to configure advanced RIP features.

This section covers the following topics:

- Configuring an Additional Routing Metric
- Configuring RIPv2 Route Summarization
- Disabling Host Route Reception
- Advertising a Default Route
- Configuring Inbound/Outbound Route Filtering
- Configuring a Priority for RIP
- Configuring RIP Route Redistribution

Before configuring RIP routing feature, complete the following tasks:

- Configure an IP address for each interface, and make sure all neighboring routers are reachable to each other.
- Configure RIP basic functions

Configuring an Additional Routing Metric

An additional routing metric can be added to the metric of an inbound or outbound RIP route.

The outbound additional metric is added to the metric of a sent route, and the route's metric in the routing table is not changed.

The inbound additional metric is added to the metric of a received route before the route is added into the routing table, and the route's metric is changed. If the sum of the additional metric and the original metric is greater than 16, the metric of the route will be 16.

Follow these steps to configure additional routing metrics:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Define an inbound additional routing metric	**rip metricin** [**route-policy** *route-policy-name*] *value*	Optional 0 by default
Define an outbound additional routing metric	**rip metricout** [**route-policy** *route-policy-name*] *value*	Optional 1 by default

Configuring RIPv2 Route Summarization

Route summarization means that subnets in a natural network are summarized into a natural network that is sent to other networks. This feature can reduce the size of routing tables.

Enabling RIPv2 route automatic summarization

You can disable RIPv2 route automatic summarization if you want to advertise all subnet routes.

Follow these steps to enable RIPv2 route automatic summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable RIPv2 automatic route summarization	**summary**	Optional Enabled by default

Advertising a summary route

You can configure RIPv2 to advertise a summary route on the specified interface.

To do so, use the following commands:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Disable RIPv2 automatic route summarization	**undo summary**	Required Enabled by default
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Advertise a summary route	**rip summary-address** *ip-address* { *mask* | *mask-length* }	Required

Note

You need to disable RIPv2 route automatic summarization before advertising a summary route on an interface.

Disabling Host Route Reception

Sometimes a router may receive from the same network many host routes, which are not helpful for routing and consume a large amount of network resources. In this case, you can disable RIP from receiving host routes to save network resources.

Follow these steps to disable RIP from receiving host routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Disable RIP from receiving host routes	**undo host-route**	Required Enabled by default

Note

RIPv2 can be disabled from receiving host routes, but RIPv1 cannot.

Advertising a Default Route

You can configure RIP to advertise a default route with a specified metric to RIP neighbors.

- In RIP view, you can configure all the interfaces of the RIP process to advertise a default route; in interface view, you can configure a RIP interface of the RIP process to advertise a default route. The latter takes precedence over the former on the interface.
- If a RIP process is enabled to advertise a default route, to disable an interface of the RIP process from default route advertisement, you can use the **rip default-route no-originate** command on the interface.

Follow these steps to configure RIP to advertise a default route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable RIP to advertise a default route	**default-route** { **only** | **originate** } [**cost** *cost*]	Optional Not enabled by default
Return to system view	**quit**	—

To do…	Use the command…	Remarks
Enter interface view	**interface** *interface-type interface-number*	—
Configure the RIP interface to advertise a default route	**rip default-route** { { **only** | **originate** } [**cost** *cost*] | **no-originate** }	Optional By default, a RIP interface can advertise a default route if the RIP process is configured with default route advertisement.

Note

The router enabled to advertise a default route does not receive default routes from RIP neighbors.

Configuring Inbound/Outbound Route Filtering

The device supports route filtering. You can filter routes by configuring the inbound and outbound route filtering policies by referencing an ACL or IP prefix list. You can also configure the router to receive only routes from a specified neighbor.

Follow these steps to configure route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure the filtering of incoming routes	**filter-policy** { *acl-number* | **gateway** *ip-prefix-name* | **ip-prefix** *ip-prefix-name* [**gateway** *ip-prefix-name*] } **import** [*interface-type interface-number*]	Required Not configured by default
Configure the filtering of outgoing routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*] | *interface-type interface-number*]	Required Not configured by default

Note

- Using the **filter-policy import** command filters incoming routes. Routes not passing the filtering will be neither installed into the routing table nor advertised to neighbors.
- Using the **filter-policy export** command filters outgoing routes, including routes redistributed with the **import-route** command.

Configuring a Priority for RIP

Multiple IGP protocols may run in a router. If you want RIP routes to have a higher priority than those learned by other routing protocols, you can assign RIP a smaller priority value to influence optimal route selection.

Follow these steps to configure a priority for RIP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure a priority for RIP	**preference** [**route-policy** *route-policy-name*] *value*	Optional 100 by default

Configuring RIP Route Redistribution

If a router runs RIP and other routing protocols, you can configure RIP to redistribute OSPF, IS-IS, BGP, static, or direct routes.

Follow these steps to configure RIP route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure a default metric for redistributed routes	**default cost** *value*	Optional The default metric of a redistributed route is 0 by default.
Redistribute routes from another protocol	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	Required No redistribution is configured by default.

Note

Only active routes can be redistributed. You can use the **display ip routing-table protocol** command to display route state information.

Configuring RIP Network Optimization

Complete the following tasks before configuring RIP network optimization:

- Configure network addresses for interfaces, and make neighboring nodes reachable to each other;
- Configure RIP basic functions.

Configuring RIP Timers

Follow these steps to configure RIP timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure values for RIP timers	**timers** { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *	Optional The default update timer, timeout timer, suppress timer, and garbage-collect timer are 30s, 180s, 120s and 120s respectively.

Note

Based on network performance, you need to make RIP timers of RIP routers identical to each other to avoid unnecessary traffic or route oscillation.

Configuring Split Horizon and Poison Reverse

Note

If both split horizon and poison reverse are configured, only the poison reverse function takes effect.

Enabling split horizon

The split horizon function disables an interface from sending routes received from the interface to prevent routing loops between adjacent routers.

Follow these steps to enable split horizon:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable split horizon	**rip split-horizon**	Optional Enabled by default

Enabling poison reverse

The poison reverse function allows an interface to advertise the routes received from it, but the metric of these routes is set to 16, making them unreachable.

Follow these steps to enable poison reverse:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable poison reverse	**rip poison-reverse**	Required Disabled by default

Configuring the Maximum Number of Load Balanced Routes

Follow these steps to configure the maximum number of load balanced routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure the maximum number of load balanced routes	**maximum load-balancing** *number*	Optional 4 by default

Enabling Zero Field Check on Incoming RIPv1 Messages

Some fields in the RIPv1 message must be zero. These fields are called zero fields. You can enable zero field check on received RIPv1 messages. If such a field contains a non-zero value, the RIPv1 message will not be processed. If you are sure that all messages are trusty, you can disable zero field check to save CPU resources.

This task does not apply to RIPv2 packets that have no zero fields.

Follow these steps to enable zero field check on incoming RIPv1 messages:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable zero field check on received RIPv1 messages	**checkzero**	Optional Enabled by default

Enabling Source IP Address Check on Incoming RIP Updates

You can enable source IP address check on incoming RIP updates.

For a message received, RIP compares the source IP address of the message with the IP address of the interface. If they are not in the same network segment, RIP discards the message.

Follow these steps to enable source IP address check on incoming RIP updates:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable source IP address check on incoming RIP messages	**validate-source-address**	Optional Enabled by default

Note

The source IP address check feature should be disabled if the RIP neighbor is not directly connected.

Configuring RIPv2 Message Authentication

RIPv2 supports two authentication modes: plain text and MD5.

In plain text authentication, the authentication information is sent with the RIP message, which however cannot meet high security needs.

Follow these steps to configure RIPv2 message authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure RIPv2 authentication	**rip authentication-mode** { **md5** { **rfc2082** *key-string key-id* | **rfc2453** *key-string* } | **simple** *password* }	Required

Note

This task does not apply to RIPv1 because RIPv1 does not support authentication. Although you can specify authentication modes for RIPv1 in interface view, the configuration does not take effect.

Specifying a RIP Neighbor

Usually, RIP sends messages to broadcast or multicast addresses. On non broadcast or multicast links, you need to manually specify RIP neighbors. If a specified neighbor is not directly connected, you must disable source address check on incoming updates.

Follow these steps to specify a RIP neighbor:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—

To do…	Use the command…	Remarks
Specify a RIP neighbor	**peer** *ip-address*	Required
Disable source address check on incoming RIP updates	**undo validate-source-address**	Required Not disabled by default

Note

- You need not use the **peer** *ip-address* command when the neighbor is directly connected; otherwise the neighbor may receive both the unicast and multicast (or broadcast) of the same routing information.
- If a specified neighbor is not directly connected, you need to disable source address check on incoming updates.

Configuring RIP-to-MIB Binding

This task allows you to enable a specific RIP process to receive SNMP requests.

Follow these steps to bind RIP to MIB:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Bind RIP to MIB	**rip mib-binding** *process-id*	Optional By default, MIB is bound to RIP process 1.

Configuring the RIP Packet Sending Rate

RIP periodically sends routing information in RIP packets to RIP neighbors.

Sending large numbers of RIP packets at the same time may affect device performance and consume large network bandwidth. To solve this problem, you can specify the maximum number of RIP packets that can be sent at the specified interval.

Follow these steps to configure the RIP packet sending rate:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable a RIP process and enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure the maximum number of RIP packets that can be sent at the specified interval	**output-delay** *time* **count** *count*	Optional By default, an interface sends up to three RIP packets every 20 milliseconds.

Displaying and Maintaining RIP

To do…	Use the command…	Remarks
Display RIP current status and configuration information	**display rip** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display all active routes in RIP database	**display rip** *process-id* **database**	
Display RIP interface information	**display rip** *process-id* **interface** [*interface-type interface-number*]	
Display routing information about a specified RIP process	**display rip** *process-id* **route** [*ip-address* { *mask* | *mask-length* } | **peer** *ip-address* | **statistics**]	
Clear the statistics of a RIP process	**reset rip** *process-id* **statistics**	Available in user view

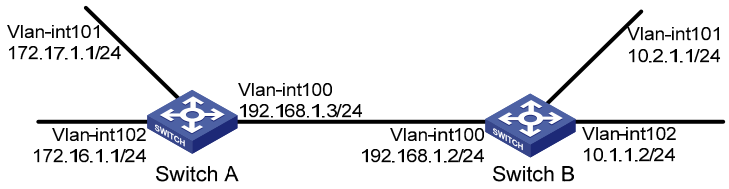
RIP Configuration Examples

Configuring RIP Version

Network requirements

As shown in Figure 1-4, enable RIPv2 on all interfaces on Switch A and Switch B.

Figure 1-4 Network diagram for RIP version configuration



Configuration procedure

1) Configure an IP address for each interface (only the IP address configuration for the VLAN interfaces is given in the following examples)

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.1.3 24
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ip address 172.17.1.1 24
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] ip address 172.16.1.1 24
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 100
```

```
[SwitchB-Vlan-interface100] ip address 192.168.1.2 24
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ip address 10.2.1.1 24
[SwitchB-Vlan-interface101] quit
```

2) Configure basic RIP functions

Configure Switch A.

```
[SwitchA] rip
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] network 172.16.0.0
[SwitchA-rip-1] network 172.17.0.0
```

Configure Switch B.

```
[SwitchB] rip
[SwitchB-rip-1] network 192.168.1.0
[SwitchB-rip-1] network 10.0.0.0
```

Display the RIP routing table of Switch A.

```
[SwitchA] display rip 1 route
Route Flags: R - RIP, T - TRIP
             P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------
 Peer 192.168.1.2  on Vlan-interface100
      Destination/Mask        Nexthop     Cost    Tag   Flags   Sec
         10.0.0.0/8        192.168.1.2      1       0    RA      11
```

From the routing table, you can find that RIPv1 uses a natural mask.

3) On Switch A and Switch B, specify the RIP version as RIPv2, and disable RIPv2 route automatic summarization to advertise all subnet routes.

Configure RIPv2 on Switch A.

```
[SwitchA] rip
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
```

Configure RIPv2 on Switch B.

```
[SwitchB] rip
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
```

Display the RIP routing table on Switch A.

```
[SwitchA] display rip 1 route
Route Flags: R - RIP, T - TRIP
             P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------
 Peer 192.168.1.2  on Vlan-interface100
      Destination/Mask        Nexthop     Cost    Tag   Flags   Sec
         10.0.0.0/8        192.168.1.2      1       0    RA      50
         10.2.1.0/24       192.168.1.2      1       0    RA      16
         10.1.1.0/24       192.168.1.2      1       0    RA      16
```

From the routing table, you can see RIPv2 uses classless subnet mask.

Note

Since the routing information advertised by RIPv1 has a long aging time, it will still exist until it ages out after RIPv2 is configured.

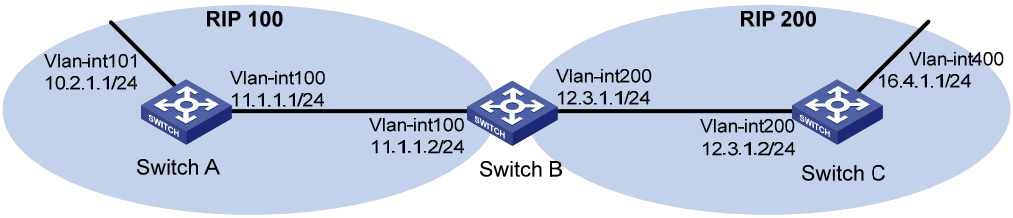
Configuring RIP Route Redistribution

Network requirements

As shown in the following figure:

- Two RIP processes are running on Switch B, which communicates with Switch A through RIP 100 and with Switch C through RIP 200.
- Configure route redistribution on Switch B to make RIP 200 redistribute direct routes and routes from RIP 100. Thus, Switch C can learn routes destined for 10.2.1.0/24 and 11.1.1.0/24, while Switch A cannot learn routes destined for 12.3.1.0/24 and 16.4.1.0/24.
- Configure a filtering policy on Switch B to filter out the route 10.2.1.1/24 from RIP 100, making the route not advertised to Switch C.

Figure 1-5 Network diagram for RIP route redistribution configuration



Configuration procedure

1) Configure an IP address for each interface (Omitted).
2) Configure basic RIP functions.

Enable RIP 100 and specify RIP version 2 on Switch A.

```
<SwitchA> system-view
[SwitchA] rip 100
[SwitchA-rip-100] network 10.0.0.0
[SwitchA-rip-100] network 11.0.0.0
[SwitchA-rip-100] version 2
[SwitchA-rip-100] undo summary
[SwitchA-rip-100] quit
```

Enable RIP 100 and RIP 200 and specify RIP version 2 on Switch B.

```
<SwitchB> system-view
[SwitchB] rip 100
[SwitchB-rip-100] network 11.0.0.0
[SwitchB-rip-100] version 2
```

```
[SwitchB-rip-100] undo summary
[SwitchB-rip-100] quit
[SwitchB] rip 200
[SwitchB-rip-200] network 12.0.0.0
[SwitchB-rip-200] version 2
[SwitchB-rip-200] undo summary
[SwitchB-rip-200] quit
```

Enable RIP 200 and specify RIP version 2 on Switch C.

```
<SwitchC> system-view
[SwitchC] rip 200
[SwitchC-rip-200] network 12.0.0.0
[SwitchC-rip-200] network 16.0.0.0
[SwitchC-rip-200] version 2
[SwitchC-rip-200] undo summary
```

Display the routing table of Switch C.

```
[SwitchC] display ip routing-table
Routing Tables: Public
         Destinations : 6        Routes : 6
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
12.3.1.0/24         Direct 0    0            12.3.1.2        Vlan200
12.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0            16.4.1.1        Vlan400
16.4.1.1/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

3) Configure route redistribution

On Switch B, configure RIP 200 to redistribute direct routes and routes from RIP 100.

```
[SwitchB] rip 200
[SwitchB-rip-200] import-route rip 100
[SwitchB-rip-200] import-route direct
[SwitchB-rip-200] quit
```

Display the routing table of Switch C.

```
[SwitchC] display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
10.2.1.0/24         RIP    100  1            12.3.1.1        Vlan200
11.1.1.0/24         RIP    100  1            12.3.1.1        Vlan200
12.3.1.0/24         Direct 0    0            12.3.1.2        Vlan200
12.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0            16.4.1.1        Vlan400
16.4.1.1/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

4) Configure an filtering policy to filter redistributed routes

Configure ACL 2000 to filter routes redistributed from RIP 100 on Switch B, making the route 10.2.1.0/24 not advertised to Switch C.

```
[SwitchB] acl number 2000
[SwitchB-acl-basic-2000] rule deny source 10.2.1.1 0.0.0.255
[SwitchB-acl-basic-2000] rule permit
[SwitchB-acl-basic-2000] quit
[SwitchB] rip 200
[SwitchB-rip-200] filter-policy 2000 export rip 100
```

Display the routing table of Switch C.

```
[SwitchC] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
11.1.1.0/24         RIP    100  1            12.3.1.1        Vlan200
12.3.1.0/24         Direct 0    0            12.3.1.2        Vlan200
12.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
16.4.1.0/24         Direct 0    0            16.4.1.1        Vlan400
16.4.1.1/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

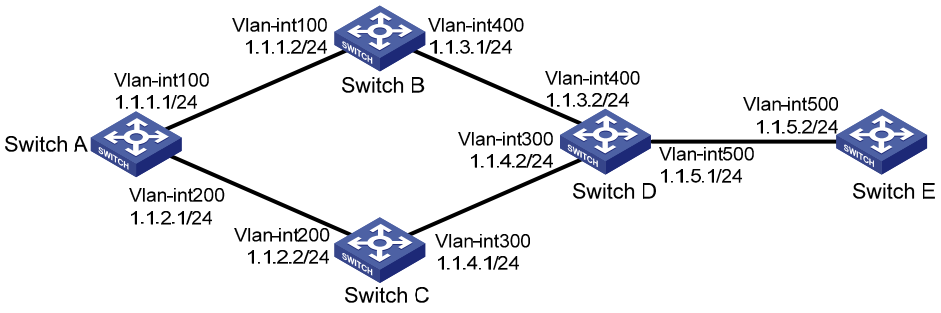
Configuring an Additional Metric for a RIP Interface

Network requirements

As shown in the following figure:

- RIP is enabled on all the interfaces of Switch A, Switch B, Switch C, Switch D, and Switch E. The switches are interconnected through RIPv2.
- Switch A has two links to Switch D. The link from Switch B to Switch D is more stable than that from Switch C to Switch D. Configure an additional metric for RIP routes received through VLAN-interface 200 on Switch A so that Switch A prefers the 1.1.5.0/24 network learned from Switch B.

Figure 1-6 Network diagram for RIP interface additional metric configuration



Configuration procedure

1) Configure IP addresses for the interfaces (omitted).
2) Configure RIP basic functions.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] rip 1
[SwitchA-rip-1] network 1.0.0.0
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
[SwitchA-rip-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] rip 1
[SwitchB-rip-1] network 1.0.0.0
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchB] rip 1
[SwitchC-rip-1] network 1.0.0.0
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] network 1.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] rip 1
[SwitchE-rip-1] network 1.0.0.0
[SwitchE-rip-1] version 2
[SwitchE-rip-1] undo summary
```

Display the IP routing table of Switch A.

```
[SwitchA] display rip 1 database
   1.0.0.0/8, cost 0, ClassfulSumm
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, Rip-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, Rip-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 1, nexthop 1.1.2.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.2.2
```

The display shows that there are two RIP routes to network 1.1.5.0/24. Their next hops are Switch B (1.1.1.2) and Switch C (1.1.2.2) respectively, with the same cost of 2. Switch C is the next hop router to reach network 1.1.4.0/24, with a cost of 1.

3) Configure an additional metric for the RIP interface.

Configure an additional metric of 3 for VLAN-interface 200 on Switch A.

```
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] rip metricin 3
[SwitchA-Vlan-interface200] display rip 1 database
   1.0.0.0/8, cost 0, ClassfulSumm
       1.1.1.0/24, cost 0, nexthop 1.1.1.1, Rip-interface
       1.1.2.0/24, cost 0, nexthop 1.1.2.1, Rip-interface
       1.1.3.0/24, cost 1, nexthop 1.1.1.2
       1.1.4.0/24, cost 2, nexthop 1.1.1.2
       1.1.5.0/24, cost 2, nexthop 1.1.1.2
```

The display shows that there is only one RIP route to network 1.1.5.0/24, with the next hop as Switch B (1.1.1.2) and a cost of 2.

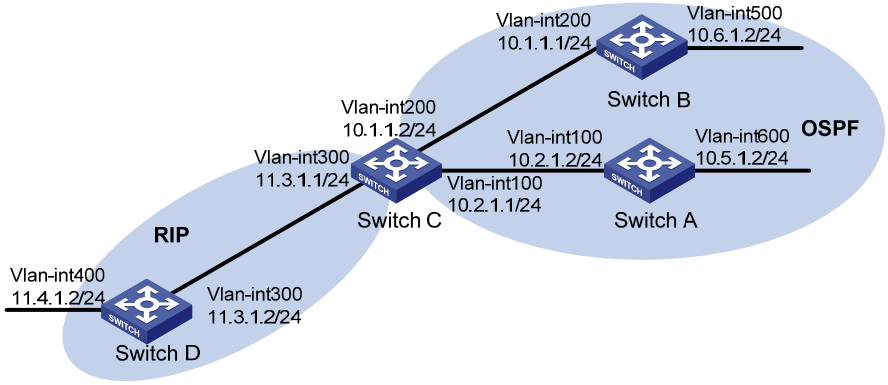
Configuring RIP to Advertise a Summary Route

Network requirements

As shown in the following figure:

- Switch A and Switch B run OSPF, Switch D runs RIP, and Switch C runs OSPF and RIP.
- Configure RIP to redistribute OSPF routes on Switch C so that Switch D has routes destined for networks 10.1.1.0/24, 10.2.1.0/24, 10.5.1.0/24, and 10.6.1.0/24.
- Route summarization is configured on Switch C and only the summary route 10.0.0.0/8 is advertised, thus reducing the routing table size of Switch D.

Figure 1-7 Network diagram for RIP summary route advertisement



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.5.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
```

Configure Switch B.

```
<SwitchB> system-view
```

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 10.6.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
```

3) Configure RIP basic functions.

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] rip 1
[SwitchC-rip-1] network 11.3.1.0
[SwitchC-rip-1] version 2
[SwitchC-rip-1] undo summary
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] network 11.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
[SwitchD-rip-1] quit
```

Configure RIP to redistribute the routes from OSPF process 1 and direct routes on Switch C.

```
[SwitchC-rip-1] import-route direct
[SwitchC-rip-1] import-route ospf 1
```

Display the routing table information of Switch D.

```
[SwitchD] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.2.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.5.1.0/24         RIP    100  1            11.3.1.1        Vlan300
10.6.1.0/24         RIP    100  1            11.3.1.1        Vlan300
11.3.1.0/24         Direct 0    0            11.3.1.2        Vlan300
11.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0            11.4.1.2        Vlan400
11.4.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
```

```
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
```

4) Configure route summarization on Switch C and advertise only the summary route 10.0.0.0/8.

```
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] rip summary-address 10.0.0.0 8
```

Display the routing table information of Switch D.

```
[SwitchD] display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.0.0.0/8          RIP    100  1            11.3.1.1        Vlan300
11.3.1.0/24         Direct 0    0            11.3.1.2        Vlan300
11.3.1.2/32         Direct 0    0            127.0.0.1       InLoop0
11.4.1.0/24         Direct 0    0            11.4.1.2        Vlan400
11.4.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

Troubleshooting RIP

No RIP Updates Received

Symptom:

No RIP updates are received when the links work well.

Analysis:

After enabling RIP, you must use the **network** command to enable corresponding interfaces. Make sure no interfaces are disabled from handling RIP messages.

If the peer is configured to send multicast messages, the same should be configured on the local end.

Solution:

- Use the **display current-configuration** command to check RIP configuration
- Use the **display rip** command to check whether some interface is disabled

Route Oscillation Occurred

Symptom:

When all links work well, route oscillation occurs on the RIP network. After displaying the routing table, you may find some routes appear and disappear in the routing table intermittently.

Analysis:

In the RIP network, make sure all the same timers within the whole network are identical and relationships between timers are reasonable. For example, the timeout timer value should be greater than the update timer value.

Solution:

- Use the **display rip** command to check the configuration of RIP timers
- Use the **timers** command to adjust timers properly.

Table of Contents

1 OSPF Configuration

Open Shortest Path First (OSPF) is a link state interior gateway protocol developed by the OSPF working group of the Internet Engineering Task Force (IETF). At present, OSPF version 2 (RFC2328) is used.

When configuring OSPF, go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

Introduction to OSPF

Note

Unless otherwise noted, OSPF refers to OSPFv2 throughout this document.

OSPF has the following features:

- Wide scope: Supports networks of various sizes and up to several hundred routers in an OSPF routing domain.
- Fast convergence: Transmits updates instantly after network topology changes for routing information synchronization in the AS.
- Loop-free: Computes routes with the shortest path first (SPF) algorithm according to collected link states, so no route loops are generated.

- Area partition: Allows an AS to be split into different areas for ease of management and routing information transmitted between areas is summarized to reduce network bandwidth consumption.
- Equal-cost multi-route: Supports multiple equal-cost routes to a destination.
- Routing hierarchy: Supports a four-level routing hierarchy that prioritizes routes into intra-area, inter-area, external Type-1, and external Type-2 routes.
- Authentication: Supports interface-based packet authentication to ensure the security of packet exchange.
- Multicast: Supports multicasting protocol packets on some types of links.

Basic Concepts

Autonomous System

A set of routers using the same routing protocol to exchange routing information constitute an Autonomous System (AS).

OSPF route computation

OSPF route computation in an area is described as follows:

- Based on the network topology around itself, each router generates Link State Advertisements (LSA) and sends them to other routers in update packets.
- Each OSPF router collects LSAs from other routers to compose a LSDB (Link State Database). An LSA describes the network topology around a router, so the LSDB describes the entire network topology of the AS.
- Each router transforms the LSDB to a weighted directed graph, which actually reflects the topology architecture of the entire network. All the routers have the same graph.
- Each router uses the SPF algorithm to compute a Shortest Path Tree that shows the routes to the nodes in the autonomous system. The router itself is the root of the tree.

Router ID

An OSPF process running on a router must have its own router ID, which is a 32-bit unsigned integer, the unique identifier of the router in the AS.

OSPF packets

OSPF uses five types of packets:

- Hello packet: Periodically sent to find and maintain neighbors, containing the values of some timers, information about the DR, BDR and known neighbors.
- DD packet (database description packet): Describes the digest of each LSA in the LSDB, exchanged between two routers for data synchronization.
- LSR (link state request) packet: Requests needed LSAs from the neighbor. After exchanging the DD packets, the two routers know which LSAs of the neighbor are missing from the local LSDBs. Then, they send an LSR packet to each other, requesting the missing LSAs. The LSA packet contains the digest of the missing LSAs.
- LSU (link state update) packet: Transmits the needed LSAs to the neighbor.
- LSAck (link state acknowledgment) packet: Acknowledges received LSU packets. It contains the headers of received LSAs (a packet can acknowledge multiple LSAs).

LSA types

OSPF sends routing information in LSAs, which, as defined in RFC 2328, have the following types:

- Router LSA: Type-1 LSA, originated by all routers, flooded throughout a single area only. This LSA describes the collected states of the router's interfaces to an area.
- Network LSA: Type-2 LSA, originated for broadcast and NBMA networks by the designated router, flooded throughout a single area only. This LSA contains the list of routers connected to the network.
- Network Summary LSA: Type-3 LSA, originated by ABRs (Area Border Routers), and flooded throughout the LSA's associated area. Each summary-LSA describes a route to a destination outside the area, yet still inside the AS (an inter-area route).
- ASBR Summary LSA: Type-4 LSA, originated by ABRs and flooded throughout the LSA's associated area. Type 4 summary-LSAs describe routes to ASBR (Autonomous System Boundary Router).
- AS External LSA: Type-5 LSA, originated by ASBRs, and flooded throughout the AS (except stub and NSSA areas). Each AS-external-LSA describes a route to another AS.
- NSSA LSA: Type-7 LSA, as defined in RFC 1587, originated by ASBRs in NSSAs (Not-So-Stubby Areas) and flooded throughout a single NSSA. NSSA LSAs describe routes to other ASs.
- Opaque LSA: A proposed type of LSA, the format of which consists of a standard LSA header and application specific information. Opaque LSAs are used by the OSPF protocol or by some application to distribute information into the OSPF routing domain. The opaque LSA includes three types, Type 9, Type 10 and Type 11, which are used to flood into different areas. The Type 9 opaque LSA is flooded into the local subnet, the Type 10 is flooded into the local area, and the Type 11 is flooded throughout the whole AS.

Neighbor and Adjacency

In OSPF, the “Neighbor” and”Adjacency” are two different concepts.

Neighbor: Two routers that have interfaces to a common network. Neighbor relationships are maintained by, and usually dynamically discovered by, OSPF's hello packets. When a router starts, it sends a hello packet via the OSPF interface, and the router that receives the hello packet checks parameters carried in the packet. If parameters of the two routers match, they become neighbors.

Adjacency: A relationship formed between selected neighboring routers for the purpose of exchanging routing information. Not every pair of neighboring routers become adjacent, which depends on network types. Only by synchronizing the LSDB via exchanging DD packets and LSAs can two routers become adjacent.

OSPF Area Partition

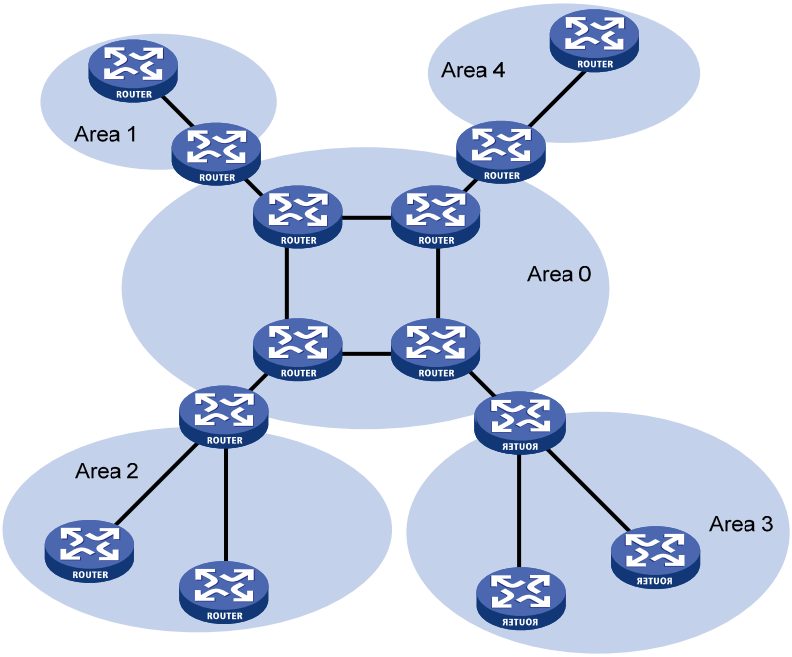
Area partition

When a large number of OSPF routers are present on a network, LSDBs may become so large that a great amount of storage space is occupied and CPU resources are exhausted by performing SPF computation.

In addition, as the topology of a large network is prone to changes, enormous OSPF packets may be created, reducing bandwidth utilization. Each topology change makes all routers perform route calculation.

To solve this problem, OSPF splits an AS into multiple areas, which are identified by area ID. The boundaries between areas are routers rather than links. A network segment (or a link) can only reside in one area, in other words, an OSPF interface must be specified to belong to its attached area, as shown in the figure below.

Figure 1-1 OSPF area partition



After area partition, area border routers perform route summarization to reduce the number of LSAs advertised to other areas and minimize the effect of topology changes.

Backbone area and virtual links

Each AS has a backbone area, which is responsible for distributing routing information between none-backbone areas. Routing information between non-backbone areas must be forwarded by the backbone area. Therefore, OSPF requires that:

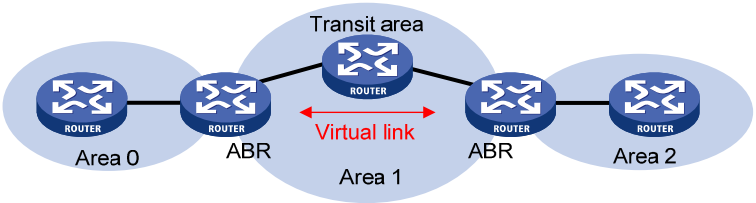
- All non-backbone areas must maintain connectivity to the backbone area.
- The backbone area itself must maintain connectivity.

In practice, due to physical limitations, the requirements may not be satisfied. In this case, configuring OSPF virtual links is a solution.

A virtual link is established between two area border routers via a non-backbone area and is configured on both ABRs to take effect. The area that provides the non-backbone area internal route for the virtual link is a “transit area”.

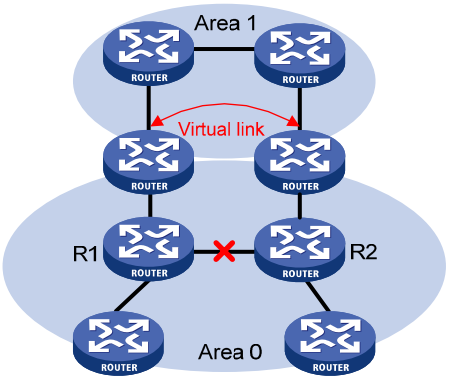
In the following figure, Area 2 has no direct physical link to the backbone area 0. Configuring a virtual link between ABRs can connect Area 2 to the backbone area.

Figure 1-2 Virtual link application 1



Another application of virtual links is to provide redundant links. If the backbone area cannot maintain internal connectivity due to a physical link failure, configuring a virtual link can guarantee logical connectivity in the backbone area, as shown below.

Figure 1-3 Virtual link application 2



The virtual link between the two ABRs acts as a point-to-point connection. Therefore, you can configure interface parameters such as hello packet interval on the virtual link as they are configured on physical interfaces.

The two ABRs on the virtual link exchange OSPF packets with each other directly, and the OSPF routers in between simply convey these OSPF packets as normal IP packets.

Stub area

The ABR in a stub area does not distribute Type-5 LSAs into the area, so the routing table size and amount of routing information in this area are reduced significantly.

You can configure the stub area as a totally stub area, where the ABR advertises neither the destinations to other areas nor external routes.

Stub area configuration is optional, and not every area is eligible to be a stub area. In general, a stub area resides on the border of the AS.

The ABR in a stub area generates a default route into the area.

Note the following when configuring a (totally) stub area:

- The backbone area cannot be a (totally) stub area.
- To configure an area as a stub area, the **stub** command must be configured on routers in the area.
- To configure an area as a totally stub area, the **stub** command must be configured on routers in the area, and the ABR of the area must be configured with the **stub** [**no-summary**] command.
- A (totally) stub area cannot have an ASBR because AS external routes cannot be distributed into the stub area.
- Virtual links cannot transit (totally) stub areas.

NSSA area

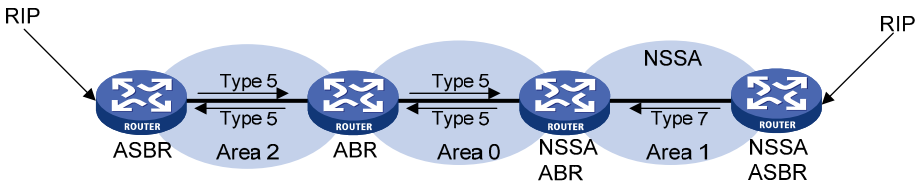
Similar to a stub area, an NSSA area imports no AS external LSA (Type-5 LSA) but can import Type-7 LSAs that are generated by the ASBR and distributed throughout the NSSA area. When traveling to the NSSA ABR, Type-7 LSAs are translated into Type-5 LSAs by the ABR for advertisement to other areas.

In the following figure, the OSPF AS contains three areas: Area 1, Area 2 and Area 0. The other two ASs employ the RIP protocol. Area 1 is an NSSA area, and the ASBR in it translates RIP routes into Type-7 LSAs and advertises them throughout Area 1. When these LSAs travel to the NSSA ABR, the ABR translates Type-7 LSAs to Type-5 LSAs for advertisement to Area 0 and Area 2.

On the left of the figure, RIP routes are translated into Type-5 LSAs by the ASBR of Area 2 and distributed into the OSPF AS. However, Area 1 is an NSSA area, so these Type-5 LSAs cannot travel to Area 1.

Like stub areas, virtual links cannot transit NSSA areas.

Figure 1-4 NSSA area



Comparsion between the areas

Figure 1-5 Comparison between the areas

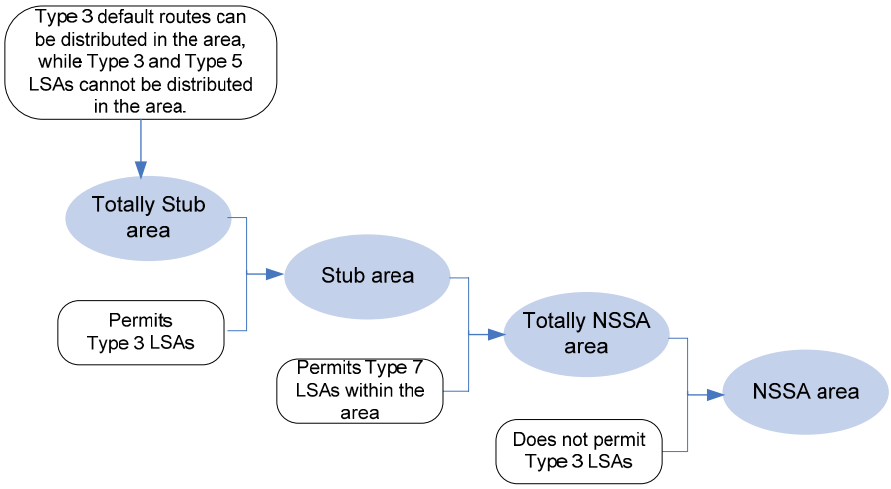


Figure 1-5 shows the comparison of the areas:

- A totally stub area can import Type 3 default routes advertised by the ABR, while it does not import external routes and inter-area routes.
- Compared with a totally stub area, a stub area can import inter-area routes.
- Compared with a stub area, an NSSA area can import external routes through Type 7 LSAs advertised by the ASBR to the area.
- Compared with an NSSA area, a totally NSSA area does not import inter-area routes.

Classification of Routers

Router types

The OSPF routers fall into four types according to the position in the AS:

1) Internal Router

All interfaces on an internal router belong to one OSPF area.

2) Area Border Router (ABR)

An area border router belongs to more than two areas, one of which must be the backbone area. It connects the backbone area to a non-backbone area. The connection between an area border router and the backbone area can be physical or logical.

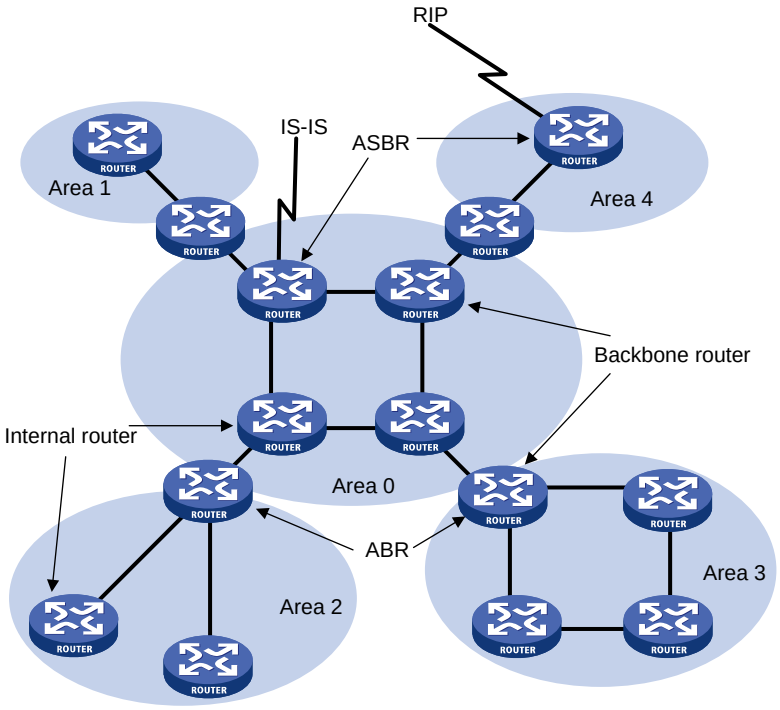
3) Backbone Router

At least one interface of a backbone router must be attached to the backbone area. Therefore, all ABRs and internal routers in area 0 are backbone routers.

4) Autonomous System Border Router (ASBR)

The router exchanging routing information with another AS is an ASBR, which may not reside on the boundary of the AS. It can be an internal router or area border router.

Figure 1-6 OSPF router types



Route types

OSPF prioritize routes into four levels:

- Intra-area route
- Inter-area route
- Type-1 external route
- Type-2 external route

The intra-area and inter-area routes describe the network topology of the AS, while external routes describe routes to destinations outside the AS.

OSPF classifies external routes into two types: Type-1 and Type-2. A Type-1 external route is an IGP route, such as a RIP or static route, which has high credibility and whose cost is comparable with the cost of an OSPF internal route. The cost from a router to the destination of the Type-1 external route= the cost from the router to the corresponding ASBR+ the cost from the ASBR to the destination of the external route.

A Type-2 external route is an EGP route, which has low credibility, so OSPF considers the cost from the ASBR to the destination of the Type-2 external route is much greater than the cost from the ASBR to an OSPF internal router. Therefore, the cost from the internal router to the destination of the Type-2 external route= the cost from the ASBR to the destination of the Type-2 external route. If two routes to

the same destination have the same cost, then take the cost from the router to the ASBR into consideration.

Classification of OSPF Networks

OSPF network types

OSPF classifies networks into four types upon the link layer protocol:

- Broadcast: When the link layer protocol is Ethernet or FDDI, OSPF considers the network type broadcast by default. On Broadcast networks, hello packets, LSU packets and LSAck packets are sent to multicast addresses 224.0.0.5 (reserved for OSPF routers) and 224.0.0.6 (reserved for OSPF DRs), while DD packets and LSR packets are sent in unicast.
- NBMA (Non-Broadcast Multi-Access): When the link layer protocol is Frame Relay, ATM or X.25, OSPF considers the network type as NBMA by default. Packets on these networks are sent to unicast addresses.
- P2MP (point-to-multipoint): By default, OSPF considers no link layer protocol as P2MP, which is a conversion from other network types such as NBMA in general. On P2MP networks, packets are sent to multicast addresses (224.0.0.5).
- P2P (point-to-point): When the link layer protocol is PPP or HDLC, OSPF considers the network type as P2P. On P2P networks, packets are sent to multicast addresses (224.0.0.5).

NBMA network configuration principle

Typical NBMA networks are ATM and Frame Relay networks.

You need to perform some special configuration on NBMA interfaces. Since these interfaces cannot broadcast hello packets for neighbor location, you need to specify neighbors manually and configure whether the neighbors have the DR election right.

An NBMA network is fully meshed, which means any two routers in the NBMA network have a direct virtual link for communication. If direct connections are not available between some routers, the type of interfaces associated should be configured as P2MP, or as P2P for interfaces with only one neighbor.

Differences between NBMA and P2MP networks:

- NBMA networks are fully meshed, non-broadcast and multi access. P2MP networks are not required to be fully meshed.
- It is required to elect the DR and BDR on NBMA networks, while DR and BDR are not available on P2MP networks.
- NBMA is the default network type, while P2MP is a conversion from other network types, such as NBMA in general.
- On NBMA networks, packets are unicast, and neighbors are configured manually on routers. On P2MP networks, packets are multicast.

DR and BDR

DR/BDR introduction

On broadcast or NBMA networks, any two routers exchange routing information with each other. If n routers are present on a network, n(n-1)/2 adjacencies are required. Any change on a router in the network generates traffic for routing information synchronization, consuming network resources. The Designated Router is defined to solve the problem. All other routers on the network send routing information to the DR, which is responsible for advertising link state information.

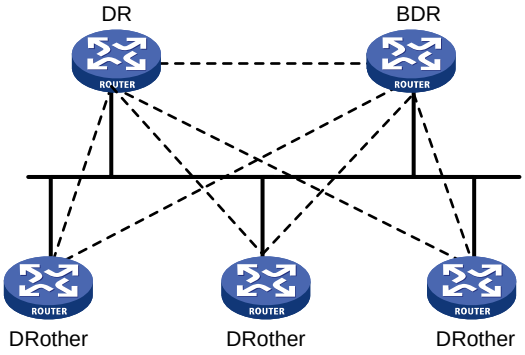
If the DR fails to work, routers on the network have to elect another DR and synchronize information with the new DR. It is time-consuming and prone to routing calculation errors. The Backup Designated Router (BDR) is introduced to reduce the synchronization period.

The BDR is elected along with the DR and establishes adjacencies for routing information exchange with all other routers. When the DR fails, the BDR will become the new DR in a very short period by avoiding adjacency establishment and DR reelection. Meanwhile, other routers elect another BDR, which requires a relatively long period but has no influence on routing calculation.

Other routers, also known as DRothers, establish no adjacency and exchange no routing information with each other, thus reducing the number of adjacencies on broadcast and NBMA networks.

In the following figure, real lines are Ethernet physical links, and dashed lines represent adjacencies. With the DR and BDR in the network, only seven adjacencies are enough.

Figure 1-7 DR and BDR in a network



DR/BDR election

The DR and BDR in a network are elected by all routers rather than configured manually. The DR priority of an interface determines its qualification for DR/BDR election. Interfaces attached to the network and having priorities higher than 0 are election candidates.

The election votes are hello packets. Each router sends the DR elected by itself in a hello packet to all the other routers. If two routers on the network declare themselves as the DR, the router with the higher DR priority wins. If DR priorities are the same, the router with the higher router ID wins. In addition, a router with the priority 0 cannot become the DR/BDR.

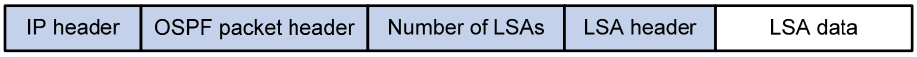
Note that:

- The DR election is available on broadcast, NBMA interfaces rather than P2P, or P2MP interfaces.
- A DR is an interface of a router and belongs to a single network segment. The router's other interfaces may be a BDR or DRother.
- After DR/BDR election and then a new router joins, it cannot become the DR immediately even if it has the highest priority on the network.
- The DR may not be the router with the highest priority in a network, and the BDR may not be the router with the second highest priority.

OSPF Packet Formats

OSPF packets are directly encapsulated into IP packets. OSPF has the IP protocol number 89. The OSPF packet format is shown below (taking a LSU packet as an example).

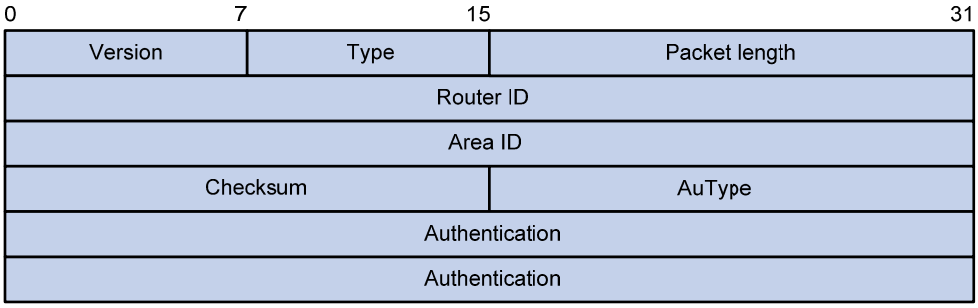
Figure 1-8 OSPF packet format



OSPF packet header

OSPF packets are classified into five types that have the same packet header, as shown below.

Figure 1-9 OSPF packet header



- Version: OSPF version number, which is 2 for OSPFv2.
- Type: OSPF packet type from 1 to 5, corresponding with hello, DD, LSR, LSU and LSAck respectively.
- Packet length: Total length of the OSPF packet in bytes, including the header.
- Router ID: ID of the advertising router.
- Area ID: ID of the area where the advertising router resides.
- Checksum: Checksum of the message.
- Autype: Authentication type from 0 to 2, corresponding with non-authentication, simple (plaintext) authentication and MD5 authentication respectively.
- Authentication: Information determined by authentication type. It is not defined for authentication type 0. It is defined as password information for authentication type 1, and defined as Key ID, MD5 authentication data length and sequence number for authentication type 2.

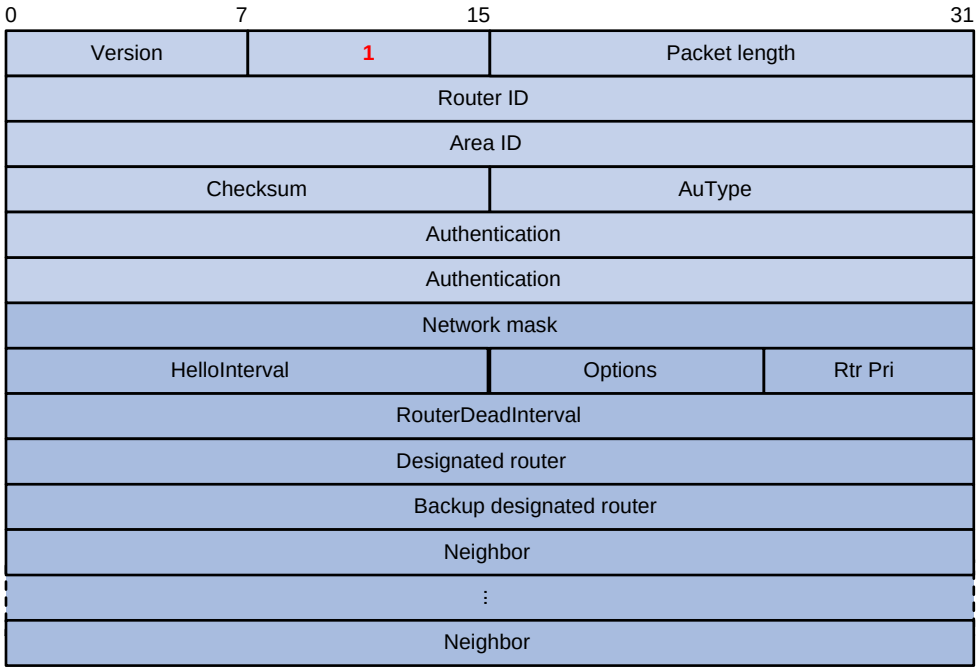
Note

MD5 authentication data is added following an OSPF packet rather than contained in the Authentication field.

Hello packet

A router sends hello packets periodically to neighbors to find and maintain neighbor relationships and to elect the DR/BDR, including information about values of timers, DR, BDR and neighbors already known. The format is shown below:

Figure 1-10 Hello packet format



Major fields:

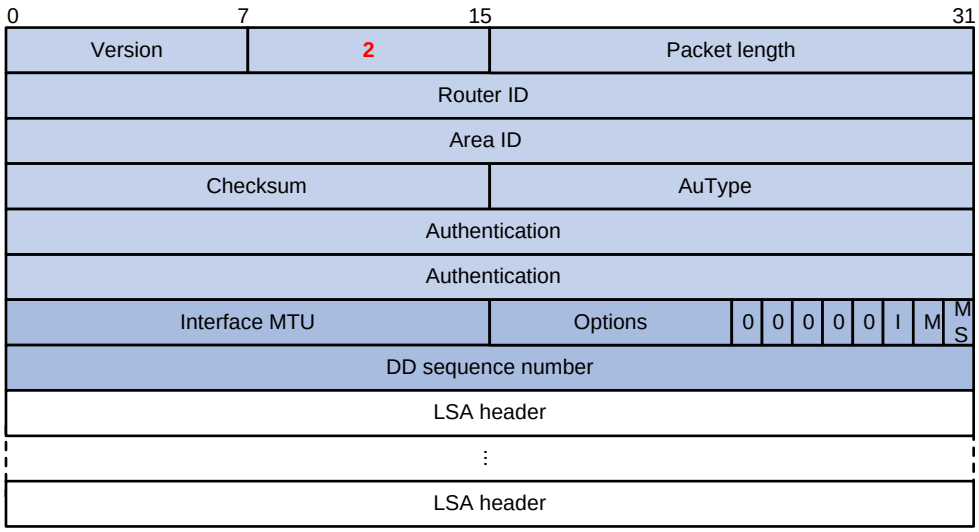
- Network mask: Network mask associated with the router's sending interface. If two routers have different network masks, they cannot become neighbors.
- HelloInterval: Interval for sending hello packets. If two routers have different intervals, they cannot become neighbors.
- Rtr Pri: Router priority. A value of 0 means the router cannot become the DR/BDR.
- RouterDeadInterval: Time before declaring a silent router down. If two routers have different time values, they cannot become neighbors.
- Designated router: IP address of the DR interface.
- Backup designated router: IP address of the BDR interface
- Neighbor: Router ID of the neighbor router.

DD packet

Two routers exchange database description (DD) packets describing their LSDBs for database synchronization, contents in DD packets including the header of each LSA (uniquely representing a LSA). The LSA header occupies small part of an LSA to reduce traffic between routers. The recipient checks whether the LSA is available using the LSA header.

The DD packet format:

Figure 1-11 DD packet format



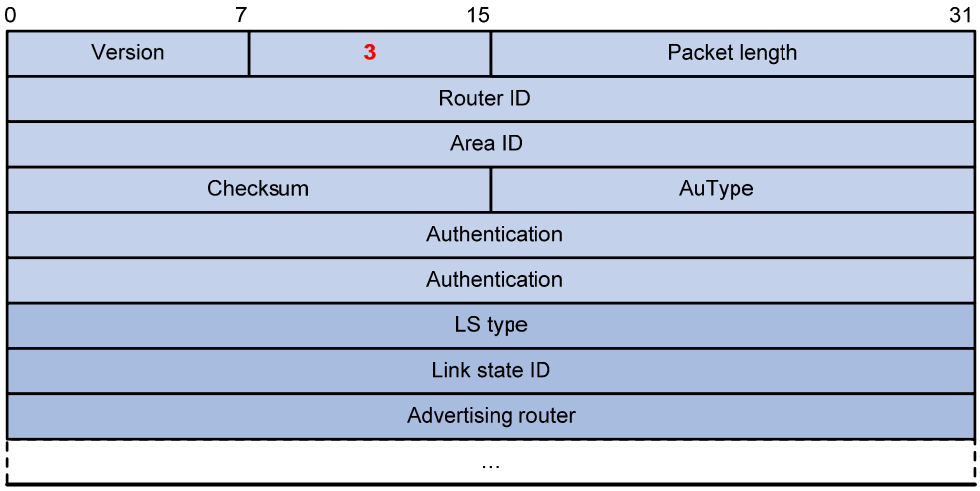
Major fields:

- Interface MTU: Size in bytes of the largest IP datagram that can be sent out the associated interface, without fragmentation.
- I (Initial) The Init bit, which is set to 1 if the packet is the first packet of database description packets, and set to 0 if not.
- M (More): The More bit, which is set to 0 if the packet is the last packet of DD packets, and set to 1 if more DD Packets are to follow.
- MS (Master/Slave): The Master/Slave bit. When set to 1, it indicates that the router is the master during the database exchange process. Otherwise, the router is the slave.
- DD Sequence Number: Used to sequence the collection of database description packets for ensuring reliability and intactness of DD packets between the master and slave. The initial value is set by the master. The DD sequence number then increments until the complete database description has been sent.

LSR packet

After exchanging DD packets, any two routers know which LSAs of the peer routers are missing from the local LSDBs. In this case, they send LSR (link state request) packets, requesting the missing LSAs. The packets contain the digests of the missing LSAs. The following figure shows the LSR packet format.

Figure 1-12 LSR packet format



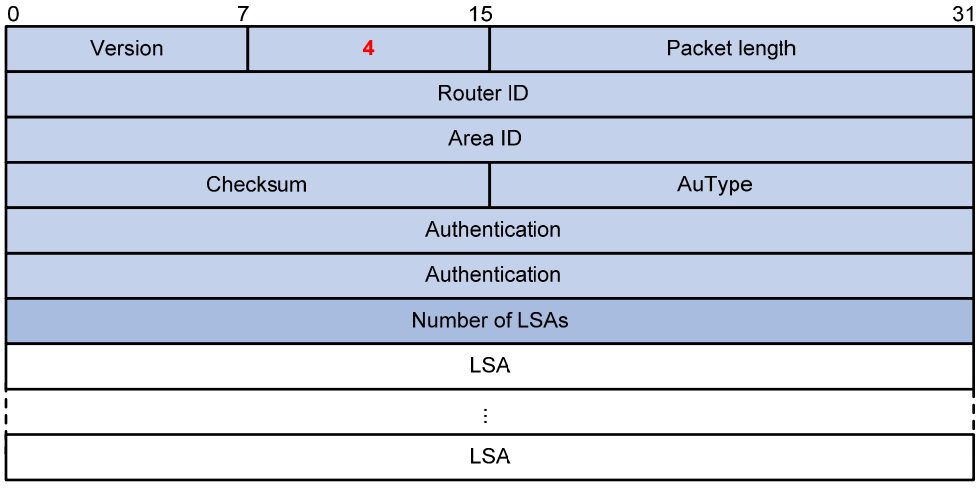
Major fields:

- LS type: Type number of the LSA to be requested. Type 1 for example indicates the Router LSA.
- Link State ID: Determined by LSA type.
- Advertising Router: ID of the router that sent the LSA.

LSU packet

LSU (Link State Update) packets are used to send the requested LSAs to peers, and each packet carries a collection of LSAs. The LSU packet format is shown below.

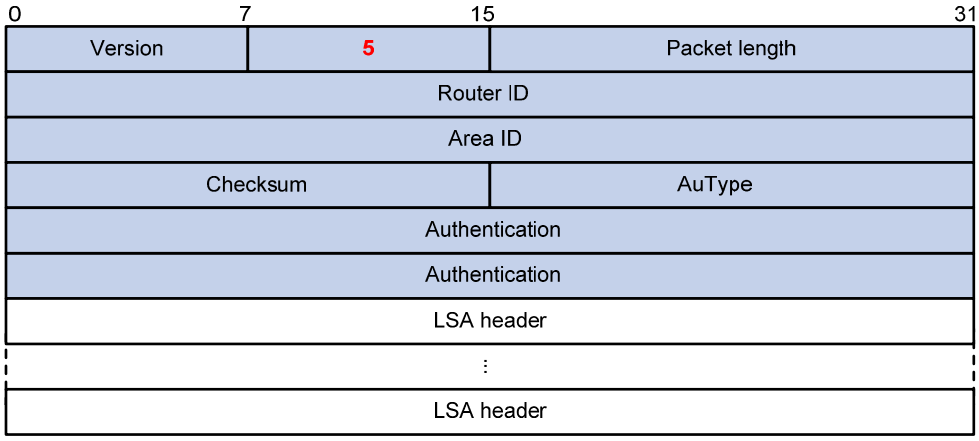
Figure 1-13 LSU packet format



LSAck packet

LSAack (Link State Acknowledgment) packets are used to acknowledge received LSU packets, contents including LSA headers to describe the corresponding LSAs. Multiple LSAs can be acknowledged in a single Link State Acknowledgment packet. The following figure gives its format.

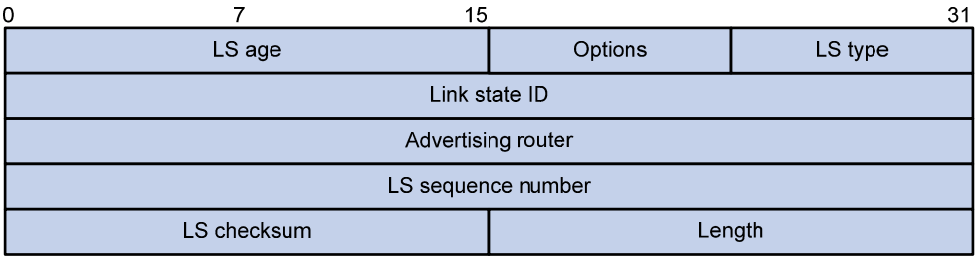
Figure 1-14 LSAck packet format



LSA header format

All LSAs have the same header, as shown in the following figure.

Figure 1-15 LSA header format



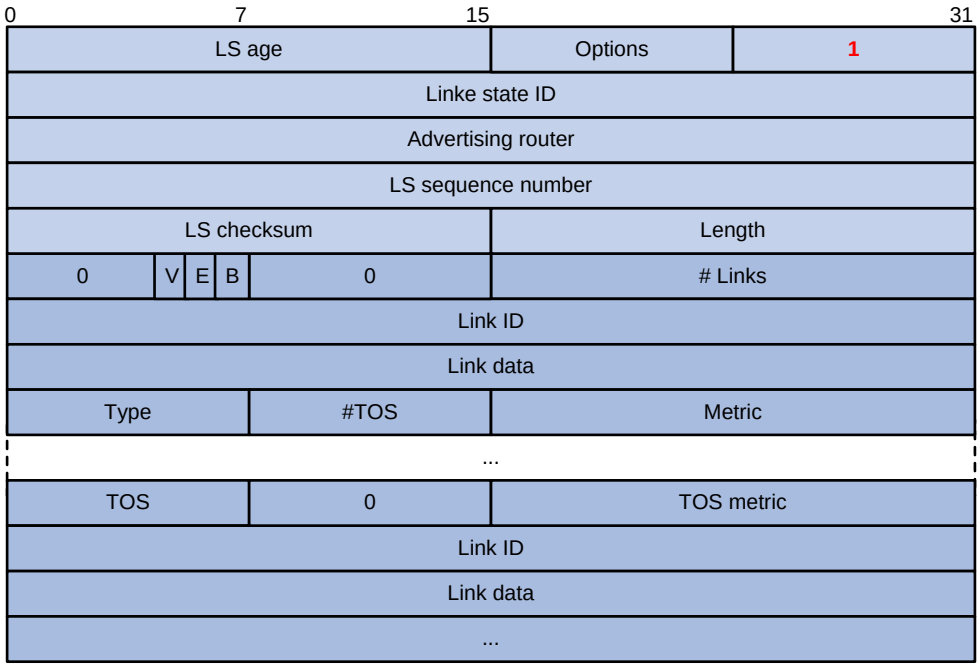
Major fields:

- LS age: Time in seconds elapsed since the LSA was originated. A LSA ages in the LSDB (added by 1 per second), but does not in transmission.
- LS type: Type of the LSA.
- Link State ID: The contents of this field depend on the LSA's type
- LS sequence number: Used by other routers to judge new and old LSAs.
- LS checksum: Checksum of the LSA except the LS age field.
- Length: Length in bytes of the LSA, including the LSA header.

Formats of LSAs

1) Router LSA

Figure 1-16 Router LSA format



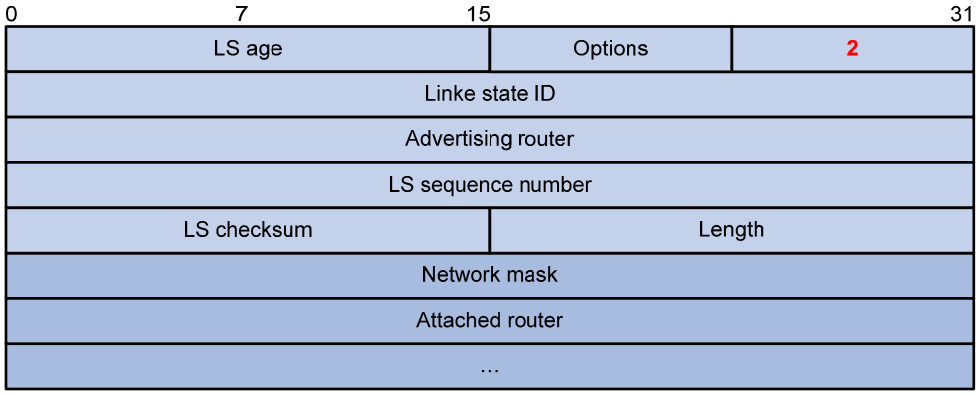
Major fields:

- Link State ID: ID of the router that originated the LSA.
- V (Virtual Link): Set to 1 if the router that originated the LSA is a virtual link endpoint.
- E (External): Set to 1 if the router that originated the LSA is an ASBR.
- B (Border): Set to 1 if the router that originated the LSA is an ABR.
- # Links: Number of router links (interfaces) to the area, described in the LSA.
- Link ID: Determined by Link type.
- Link data: Determined by Link type.
- Type: Link type. A value of 1 indicates a point-to-point link to a remote router; a value of 2 indicates a link to a transit network; a value of 3 indicates a link to a stub network; a value of 4 indicates a virtual link.
- #TOS: Number of different TOS metrics given for this link.
- Metric: Cost of using this router link.
- TOS: IP Type of Service that this metric refers to.
- TOS metric: TOS-specific metric information.

2) Network LSA

A Network LSA is originated by the DR on a broadcast or NBMA network. The LSA describes all routers attached to the network.

Figure 1-17 Network LSA format



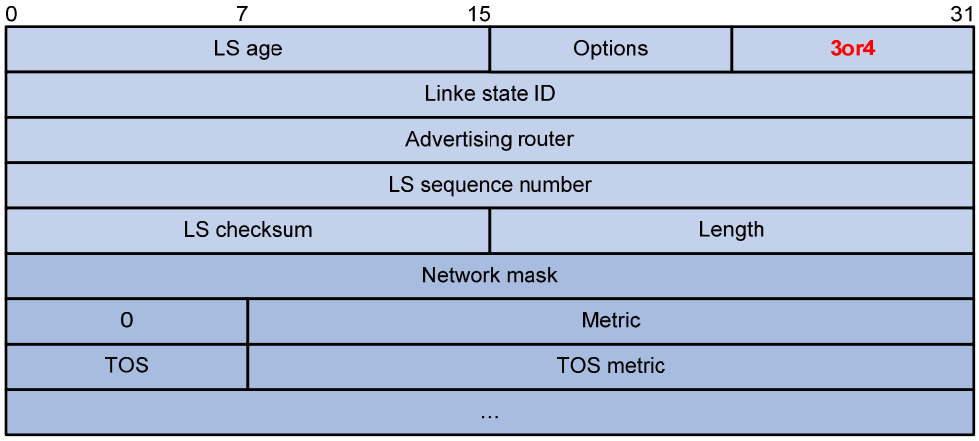
Major fields:

- Link State ID: The interface address of the DR
- Network mask: The mask of the network (a broadcast or NBMA network)
- Attached router: The IDs of the routers, which are adjacent to the DR, including the DR itself

3) Summary LSA

Network summary LSAs (Type-3 LSAs) and ASBR summary LSAs (Type-4 LSAs) are originated by ABRs. Other than the difference in the Link State ID field, the format of type 3 and 4 summary-LSAs is identical.

Figure 1-18 Summary LSA format



Major fields:

- Link State ID: For a Type-3 LSA, it is an IP address outside the area; for a type 4 LSA, it is the router ID of an ASBR outside the area.
- Network mask: The network mask for the type 3 LSA; set to 0.0.0.0 for the Type-4 LSA
- Metric: The metric to the destination

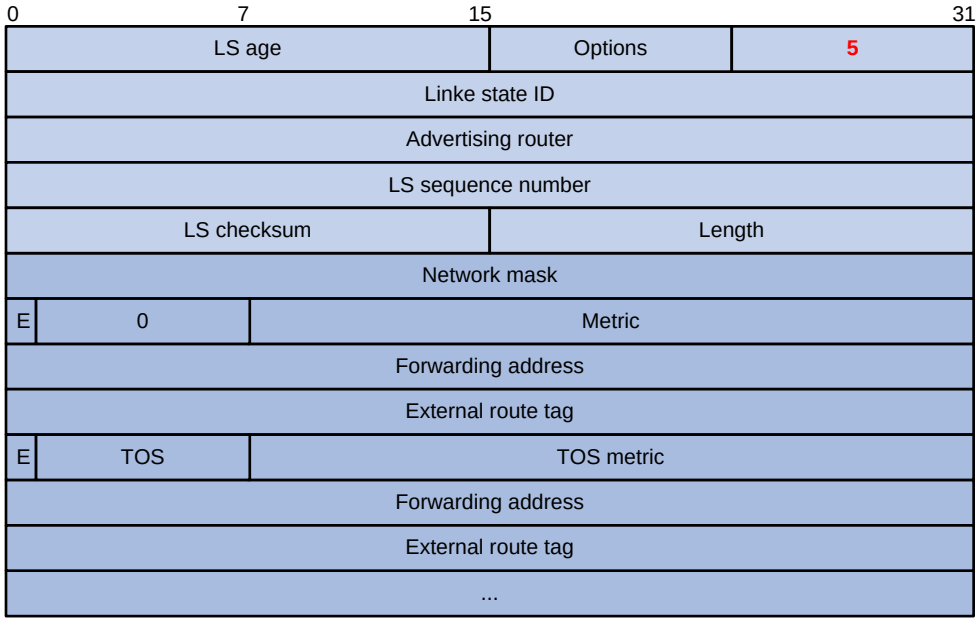
Note

A Type-3 LSA can be used to advertise a default route, having the Link State ID and Network Mask set to 0.0.0.0.

4) AS external LSA

An AS external LSA originates from an ASBR, describing routing information to a destination outside the AS.

Figure 1-19 AS external LSA format



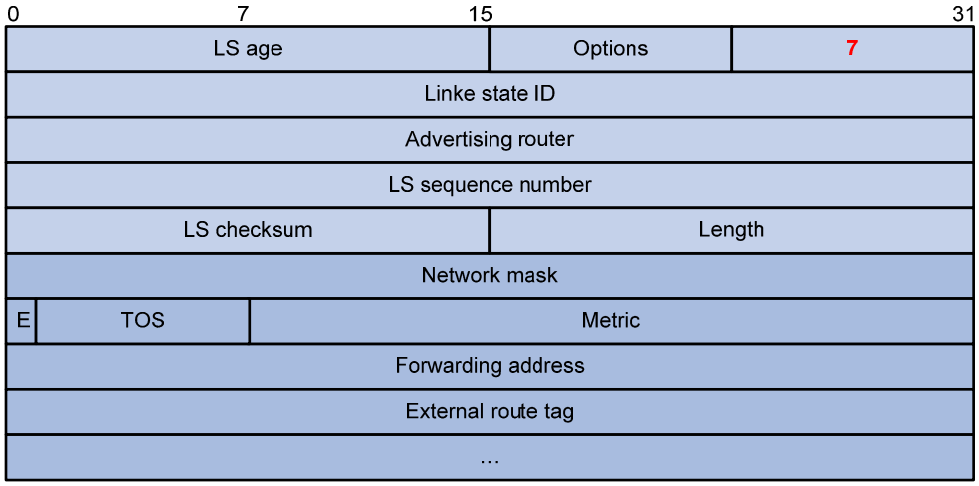
Major fields:

- Link State ID: The IP address of another AS to be advertised. When describing a default route, the Link State ID is always set to Default Destination (0.0.0.0) and the Network Mask is set to 0.0.0.0
- Network mask: The IP address mask for the advertised destination
- E (External Metric): The type of the external metric value, which is set to 1 for type 2 external routes, and set to 0 for type 1 external routes. Refer to Route types for description about external route types
- Metric: The metric to the destination
- Forwarding Address: Data traffic for the advertised destination will be forwarded to this address
- External Route Tag: A tag attached to each external route. This is not used by the OSPF protocol itself. It may be used to manage external routes.

5) NSSA external LSA

An NSSA external LSA originates from the ASBR in a NSSA and is flooded in the NSSA area only. It has the same format as the AS external LSA.

Figure 1-20 NSSA external LSA format



Supported OSPF Features

Multi-process

With multi-process support, multiple OSPF processes can run on a router simultaneously and independently. Routing information interactions between different processes seem like interactions between different routing protocols. Multiple OSPF processes can use the same RID.

An interface of a router can only belong to a single OSPF process.

Authentication

OSPF supports authentication on packets. Only packets that pass the authentication are received. If hello packets cannot pass authentication, no neighbor relationship can be established.

The authentication type for interfaces attached to a single area must be identical. Authentication types include non-authentication, plaintext authentication and MD5 ciphertext authentication. The authentication password for interfaces attached to a network segment must be identical.

OSPF Graceful Restart

Note

For GR information, refer to *GR Overview* in the *System Volume*.

After an OSPF GR Restarter restarts, it needs to perform the following two tasks in order to re-synchronize its LSDB with its neighbors.

- To obtain once again effective OSPF neighbor information (assume the adjacencies are not changed).
- To obtain once again the LSDB.

After restart, the GR Restarter negotiates GR capability with its neighbors and sends an OSPF GR signal to its GR-capable neighbors so that they will not remove their adjacencies with it and advertise the adjacencies. The GR Restarter re-establishes neighborships and updates its own routing table and

forwarding table based on the new routing information received from neighbors and removes the stale routes.

VPN

OSPF supports multi-instance, which can run in VPN networks.

In BGP MPLS VPN networks, multiple sites in the same VPN can use OSPF as the internal routing protocol, but they are treated as different ASs. An OSPF route learned by a site will be forwarded to another site as an external route, which leads to heavy OSPF routing traffic and management issues.

 **Note**

For configuration of this feature, refer to *MCE Configuration* in the *IP Routing Volume*.

OSPF sham link

An OSPF sham link is a point-to-point link between two PE routers on the MPLS VPN backbone.

In general, BGP peers exchange routing information on the MPLS VPN backbone using the BGP extended community attribute. OSPF running on a PE at the other end utilizes this information to originate a Type-3 summary LSA as an inter-area route between the PE and CE.

If a router connects to a PE router in the same area and establishes an internal route (backdoor route) to a destination, in this case, since an OSPF intraarea route has a higher priority than a backbone route, VPN traffic will always travel on the backdoor route rather than the backbone route. To avoid this, an unnumbered sham link can be configured between PE routers, connecting the router to another PE router via an intraarea route with a lower cost.

Protocols and Standards

- RFC 1765: OSPF Database Overflow
- RFC 2328: OSPF Version 2
- RFC 3101: OSPF Not-So-Stubby Area (NSSA) Option
- RFC 3137: OSPF Stub Router Advertisement
- RFC 3630: Traffic Engineering Extensions to OSPF Version 2
- RFC 4811: OSPF Out-of-Band LSDB Resynchronization
- RFC 4812: OSPF Restart Signaling
- RFC 4813: OSPF Link-Local Signaling

OSPF Configuration Task List

An OSPF routing domain has different types of routers, such as intra-area routers, ABR, and ASBR.

OSPF can work normally only after being enabled on a router, regardless of the router's type. On an OSPF-enabled router, you can use the default values of parameters, such as the transmit interval of OSPF packets, LSA delay timer, and SPF calculation interval, or configure them as required.

Network planning is needed before OSPF configuration on routers. The configurations for routers in an area are performed on the area basis. Wrong configurations may cause communication failures, even routing information block or routing loops between neighboring routers..

Complete the following tasks to configure OSPF:

Task		Remarks
Enabling OSPF		Required
Configuring OSPF Areas	Configuring a Stub Area	Optional
	Configuring an NSSA Area	
	Configuring a Virtual Link	
Configuring OSPF Network Types	Configuring the OSPF Network Type for an Interface as Broadcast	Optional
	Configuring the OSPF Network Type for an Interface as NBMA	Optional
	Configuring the OSPF Network Type for an Interface as P2MP	Optional
	Configuring the OSPF Network Type for an Interface as P2P	Optional
Configuring OSPF Route Control	Configuring OSPF Route Summarization	Optional
	Configuring OSPF Inbound Route Filtering	Optional
	Configuring ABR Type-3 LSA Filtering	Optional
	Configuring an OSPF Cost for an Interface	Optional
	Configuring the Maximum Number of OSPF Routes	Optional
	Configuring the Maximum Number of Load-balanced Routes	Optional
	Configuring a Priority for OSPF	Optional
	Configuring OSPF Route Redistribution	Optional
Configuring OSPF Network Optimization	Configuring OSPF Packet Timers	Optional
	Specifying an LSA Transmission Delay	Optional
	Specifying SPF Calculation Interval	Optional
	Specifying the LSA Minimum Repeat Arrival Interval	Optional
	Specifying the LSA Generation Interval	Optional
	Disabling Interfaces from Sending OSPF Packets	Optional
	Configuring Stub Routers	Optional
	Configuring OSPF Authentication	Optional
	Adding the Interface MTU into DD Packets	Optional
	Configuring the Maximum Number of External LSAs in LSDB	Optional
	Making External Route Selection Rules Defined in RFC1583 Compatible	Optional
	Logging Neighbor State Changes	Optional
	Configuring OSPF Network Management	Optional
	Enabling Message Logging	Optional
	Enabling the Advertisement and Reception of Opaque LSAs	Optional
	Configuring OSPF to Give Priority to Receiving and Processing Hello Packets	Optional
	Configuring the LSU Transmit Rate	Optional

Task		Remarks
Configuring OSPF Sham Link	Configuration Prerequisites	Optional
	Configuring a Loopback Interface	Optional
	Advertising Routes of a Loopback Interface	Optional
	Creating a Sham Link	Optional
Configuring OSPF Graceful Restart	Configuring the OSPF GR Restarter	Optional
	Configuring the OSPF GR Helper	Optional
	Triggering OSPF Graceful Restart	Optional

Enabling OSPF

You need to enable OSPF before you can perform other OSPF configuration tasks.

Prerequisites

Before configuring OSPF, you have configured the link layer protocol, and IP addresses for interfaces, making neighboring nodes accessible with each other at the network layer.

Configuration Procedure

To enable OSPF on a router, you need to create an OSPF process and specify areas with which the process is associated, and the network segments contained in each area. If an interface's IP address resides on a network segment of an area, the interface belongs to the area and is enabled with OSPF, and OSPF advertises the direct route of the interface.

To run OSPF, a router must have a Router ID, which is the unique identifier of the router in the AS.

- You can specify a Router ID when creating the OSPF process. Any two routers in an AS must have different Router IDs. In practice, the ID of a router is the IP address of one of its interfaces.
- If you specify no Router ID when creating the OSPF process, the global Router ID will be used. For details about global Router ID, refer to *IP Routing Overview* in the *IP Routing Volume*. You are recommended to specify a Router ID when creating the OSPF process.

The system supports OSPF multi-process and OSPF multi-instance:

- When a router runs multiple OSPF processes, you need to specify a Router ID for each process, which takes effect locally and has no influence on packet exchange between routers. Therefore, two routers having different process IDs can exchange packets.
- You can configure an OSPF process to run in a specified VPN instance.

Follow these steps to enable OSPF:

To do…	Use the command…	Remarks
Enter system view	**System-view**	—
Enable an OSPF process and enter its view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Not enabled by default.
Configure a description for the OSPF process	**description** *description*	Optional Not configured by default.

To do…	Use the command…	Remarks
Configure an OSPF area and enter OSPF area view	**area** *area-id*	Required Not configured by default.
Configure a description for the area	**description** *description*	Optional Not configured by default.
Specify a network to enable OSPF on the interface attached to the network	**network** *ip-address wildcard-mask*	Required Not configured by default.

Note

- A network segment can only belong to one area.
- It is recommended to configure a description for each OSPF process to help identify purposes of processes and for ease of management and memorization.
- It is recommended to configure a description for each area to help identify purposes of areas and for ease of management and memorization.

Configuring OSPF Areas

After splitting an OSPF AS into multiple areas, you can further configure some areas as stub areas or NSSA areas as needed.

If connectivity between the backbone and a non-backbone area or within the backbone itself cannot be achieved, you can configure virtual links to solve it.

Prerequisites

Before configuring an OSPF area, you have configured:

- IP addresses for interfaces, making neighboring nodes accessible with each other at the network layer.
- OSPF basic functions.

Configuring a Stub Area

You can configure a non-backbone area at the AS edge as a stub area by configuring the **stub** command on all the routers attached to the area. In this way, Type-5 LSAs, which describe AS external routes, will not be flooded within the stub area, reducing the routing table size. The ABR generates a default route into the stub area so that all packets destined outside of the AS are sent through the default route.

To further reduce the routing table size and routing information exchanged in the stub area, you can configure it as a totally stub area by using the **stub** [**no-summary**] command on the ABR. In this way, neither AS external routes nor inter-area routing information will be distributed into the area. All the packets destined outside of the AS or area will be sent to the ABR for forwarding.

Follow these steps to configure OSPF areas:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Configure the area as a stub area	**stub** [**no-summary**]	Required Not configured by default.
Specify a cost for the default route advertised to the stub area	**default-cost** *cost*	Optional Defaults to 1.

Note

- It is required to use the **stub** command on routers attached to a stub area.
- Using the **default-cost** command only takes effect on the ABR of a stub area.
- The backbone area cannot be a (totally) stub area.
- A (totally) stub area cannot have an ASBR because AS external routes cannot be distributed into the stub area.
- Virtual links cannot transit (totally) stub areas.

Configuring an NSSA Area

A stub area cannot redistribute routes. You can configure the area as an NSSA area to allow for route redistribution while keeping other characteristics of a stub area.

Follow these steps to configure an NSSA area:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Configure the area as an NSSA area	**nssa** [**default-route-advertise** | **no-import-route** | **no-summary**] *	Required Not configured by default.
Specify a cost for the default route advertised to the NSSA area	**default-cost** *cost*	Optional Defaults to 1.

Note

- It is required to use the **nssa** command on all the routers attached to an NSSA area.
- Using the **default-cost** command only takes effect on the ABR/ASBR of an NSSA area.

Configuring a Virtual Link

Non-backbone areas exchange routing information via the backbone area. Therefore, connectivity between the backbone and non-backbone areas and within the backbone itself must be maintained.

If necessary physical links are not available for this connectivity maintenance, you can configure virtual links to solve it.

Follow these steps to configure a virtual link:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Configure a virtual link	**vlink-peer** *router-id* [**hello** *seconds* | **retransmit** *seconds* | **trans-delay** *seconds* | **dead** *seconds* | **simple** [**plain** | **cipher**] *password* | { **md5** | **hmac-md5** } *key-id* [**plain** | **cipher**] *password*] *	Required You need to configure this command on both ends of a virtual link. Note that **hello** and **dead** intervals must be identical on both ends of the virtual link.

Configuring OSPF Network Types

OSPF classifies networks into four types: broadcast, NBMA, P2MP, and P2P, upon the link layer protocol.

You can change the network type of an interface as needed. For example:

- When an NBMA network becomes fully meshed through address mapping, namely, when any two routers in the network have a direct virtual link in between, you can change the network type to broadcast, without manually configuring the neighbors.
- When some routers in the broadcast network do not support multicast, you can change the network type to NBMA.
- An NBMA network is fully meshed, which means any two routers in the NBMA network have a direct virtual link for communication. If direct connections are not available between some routers, the type of interfaces associated should be configured as P2MP, or as P2P for interfaces with only one neighbor.

If two interfaces on a link are both configured as the broadcast, NBMA or P2MP type, they cannot establish a neighbor relationship unless they are on the same network segment.

Prerequisites

Before configuring OSPF network types, you have configured:

- IP addresses for interfaces, making neighboring nodes accessible with each other at network layer.
- OSPF basic functions.

Configuring the OSPF Network Type for an Interface as Broadcast

Follow these steps to configure the OSPF network type for an interface as broadcast:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the OSPF network type for the interface as broadcast	**ospf network-type broadcast**	Required By default, the network type is broadcast.
Configure a DR priority for the interface	**ospf dr-priority** *priority*	Optional The default DR priority is 1.

Configuring the OSPF Network Type for an Interface as NBMA

After configuring the network type of an interface as NBMA, you need to make some special configurations.

Because NBMA interfaces cannot find neighbors via broadcasting Hello packets, you need to specify neighbors and neighbor DR priorities. (A DR priority of 0 means the router does not have the DR election right; a DR priority greater than 0 means the router has the DR election right).

Follow these steps to configure the OSPF network type for an Interface as NBMA:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the OSPF network type for the interface as NBMA	**ospf network-type nbma**	Required By default, the network type is broadcast.
Configure a DR priority for the interface	**ospf dr-priority** *priority*	Optional The default DR priority is 1
Exit to system view	**quit**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Specify a neighbor and its DR priority	**peer** *ip-address* [**dr-priority** *dr-priority*]	Required

Note

The DR priority configured with the **ospf dr-priority** command and the one configured with the **peer** command have the following differences:

- The former is for actual DR election.
- The latter is to indicate whether a neighbor has the election right or not. If you configure the DR priority for a neighbor as 0, the local router will consider the neighbor has no election right, and thus no hello packet is sent to this neighbor, reducing the number of hello packets for DR/BDR election on networks. However, if the local router is the DR or BDR, it sends hello packets to the neighbor with priority 0 for adjacency establishment.

Configuring the OSPF Network Type for an Interface as P2MP

Follow these steps to configure the OSPF network type for an interface as P2MP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the OSPF network type for the interface as P2MP	**ospf network-type p2mp**	Required By default, the network type is broadcast.

Configuring the OSPF Network Type for an Interface as P2P

Follow these steps to configure the OSPF network type for an interface as P2P:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the OSPF network type for the interface as P2P.	**ospf network-type p2p**	Required By default, the network type is broadcast.

Configuring OSPF Route Control

This section covers how to control OSPF routing information advertisement and reception, and route redistribution from other protocols.

Prerequisites

Before configuring this task, you have configured:

- IP addresses for interfaces

- OSPF basic functions
- Corresponding filters if routing information filtering is needed.

Configuring OSPF Route Summarization

Route summarization: An ABR or ASBR summarizes routes with the same prefix into a single route and distribute it to other areas.

Through route summarization, routing information across areas and the size of routing tables on routers will be reduced, improving calculation speed of routers.

Assume in an area are three internal routes 19.1.1.0/24, 19.1.2.0/24, and 19.1.3.0/24. By configuring route summarization on the ABR, the three routes are summarized into the route 19.1.0.0/16 that is advertised into other areas.

Configuring route summarization on an ABR

If contiguous network segments are available in the area, you can summarize them into a single network segment. An ABR generates Type-3 LSAs on a per network segment basis for an attached non-backbone area.

In this way, the ABR in the area distributes only the summary LSA to reduce the scale of LSDBs on routers in other areas and the influence of topological changes.

Follow these steps to configure route summarization on an ABR:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter OSPF area view	**area** *area-id*	—
Configure ABR route summarization	**abr-summary** *ip-address* { *mask* | *mask-length* } [**advertise** | **not-advertise**] [**cost** *cost*]	Required The command is available on an ABR only. Not configured by default.

Configuring route summarization when redistributing routes into OSPF on an ASBR

If summarization for redistributed routes is not configured on an ASBR, it will advertise each redistributed route in a separate ASE LSA. After a summary is configured on the ASBR, it advertises only the summary route in an ASE LSA instead of more specific routes, thus reducing the number of LSAs in the LSDBs.

If summarization for redistributed routes is configured on an ASBR, it will summarize redistributed Type-5 LSAs that fall into the specified address range. If the ASBR is in an NSSA area, it also summarizes Type-7 LSAs that fall into the specified address range. If the ASBR is also the ABR, it will summarize Type-5 LSAs translated from Type-7 LSAs.

Follow these steps to configure route summarization when redistributing routes into OSPF on an ASBR:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*]*	—
Configure ASBR route summarization	**asbr-summary** *ip-address* { *mask* | *mask-length* } [**tag** *tag* | **not-advertise** | **cost** *cost*] *	Required The command is available on an ASBR only. Not configured by default.

Configuring OSPF Inbound Route Filtering

Note

- For details about IP prefix list, refer to *Route Policy Configuration* in the *IP Routing Volume*.
- For details about route policy, refer to *Route Policy Configuration* in the *IP Routing Volume*.

Since OSPF is a link state-based interior gateway protocol, routing information is contained in LSAs. Routes computed by OSPF can be filtered and only permitted routes are installed into the routing table.

There are four filtering methods:

- Filtering routing information by destination address through ACLs and IP address prefixes.
- Filtering routing information by next hop through the filtering criteria configured with the **gateway** keyword.
- Filtering routing information by destination address through ACLs and IP address prefixes and by next hop through the filtering criteria configured with the **gateway** keyword.
- Filtering routing information by route policies specified by the **route-policy** keyword.

Follow these steps to configure inbound route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure inbound route filtering	**filter-policy** { *acl-number* [**gateway** *ip-prefix-name*] | **gateway** *ip-prefix-name* | **ip-prefix** *ip-prefix-name* [**gateway** *ip-prefix-name*] **| route-policy** *route-policy-name* } **import**	Required Not configured by default.

Configuring ABR Type-3 LSA Filtering

This task is configured on an ABR to filter Type-3 LSAs to be advertised in the attached non-backbone area and the Type-3 LSAs to be advertised to other areas.

Follow these steps to configure Type-3 LSA filtering on an ABR:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Configure ABR Type-3 LSA filtering	**filter** { *acl-number* | **ip-prefix** *ip-prefix-name* } { **import** | **export** }	Required Not configured by default.

Configuring an OSPF Cost for an Interface

You can configure an OSPF cost for an interface with one of the following two methods:

- Configure the cost value in interface view.
- Configure a bandwidth reference value for the interface, and OSPF computes the cost automatically based on the bandwidth reference value: Interface OSPF cost= Bandwidth reference value/Interface bandwidth. If the calculated cost is greater than 65535, the value of 65535 is used; if the calculated cost is less than 1, the value of 1 is used.

If no cost is configured for an interface, OSPF computes the interface cost automatically.

Follow these steps to configure an OSPF cost for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an OSPF cost for the interface	**ospf cost** *value*	Optional By default, an interface computes its cost according to the bandwidth. The cost value defaults to 0 on loopback interfaces.

Follow these steps to configure a bandwidth reference value:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure a bandwidth reference value	**bandwidth-reference** *value*	Optional The value defaults to 100 Mbps.

Configuring the Maximum Number of OSPF Routes

Follow these steps to configure the maximum number of routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the maximum number of OSPF routes	**maximum-routes** { **external** | **inter** | **intra** } *number*	Optional By default, the maximum number of AS external routes, inter-area routes and intra-area routes is 12288, 10240 and 1024, respectively.

Configuring the Maximum Number of Load-balanced Routes

If several routes with the same cost to the same destination are available, configuring them as load-balanced routes can improve link utilization.

Follow these steps to configure the maximum number of load-balanced routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the maximum number of equivalent load-balanced routes	**maximum load-balancing** *maximum*	Optional The default number is 4.

Configuring a Priority for OSPF

A router may run multiple routing protocols, and it sets a priority for each protocol. When a route found by several routing protocols, the route found by the protocol with the highest priority will be selected.

Follow these steps to configure a priority for OSPF:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure a priority for OSPF	**preference** [**ase**] [**route-policy** *route-policy-name*] *value*	Optional The priority of OSPF internal routes defaults to 10. The priority of OSPF external routes defaults to 150.

Configuring OSPF Route Redistribution

Configure route redistribution into OSPF

If the router runs OSPF and other routing protocols, you can configure OSPF to redistribute RIP, IS-IS, BGP, static, or direct routes and advertise these routes in Type-5 LSAs or Type-7 LSAs.

By filtering redistributed routes, OSPF translates only routes not filtered out into Type-5 LSAs or Type-7 LSAs for advertisement.

Follow these steps to configure OSPF route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure OSPF to redistribute routes from another protocol	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] *	Required Not configured by default.
Configure OSPF to filter redistributed routes before advertisement	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*]]	Optional Not configured by default.

Note

Only active routes can be redistributed. You can use the **display ip routing-table protocol** command to display route state information.

Configure OSPF to redistribute a default route

Using the **import-route** command cannot redistribute a default external route. To do so, you need to use the **default-route-advertise** command.

Follow these steps to configure OSPF to redistribute a default external route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Redistribute a default route	**default-route-advertise** [**always** | **cost** *cost* | **type** *type* | **route-policy** *route-policy-name*] *	Optional Not redistributed by default.
	default-route-advertise summary cost *cost*	

Note

The **default-route-advertise summary cost** command is applicable only to VPN, and the default route is redistributed in a Type-3 LSA. The PE router will advertise the default route to the CE router.

Configure the default parameters for redistributed routes

You can configure default parameters such as the cost, upper limit, tag and type for redistributed routes. Tags are used to indicate information related to protocols. For example, when redistributing BGP routes, OSPF uses tags to identify AS IDs.

Follow these steps to configure the default parameters for redistributed routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the default parameters for redistributed routes (cost, route number, tag and type)	**default** { **cost** *cost* | **limit** *limit* | **tag** *tag* | **type** *type* } *	Optional By default, the default cost is 1, default upper limit of routes redistributed per time is 1000, default tag is 1 and default type of redistributed routes is Type-2.

Advertising a Host Route

Follow these steps to advertise a host route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Advertise a host route	**host-advertise** *ip-address cost*	Optional Not advertised by default.

Configuring OSPF Network Optimization

You can optimize your OSPF network in the following ways:

- Change OSPF packet timers to adjust the OSPF network convergence speed and network load. On low speed links, you need to consider the delay time for sending LSAs on interfaces.
- Change the interval for SPF calculation to reduce resource consumption caused by frequent network changes.

- Configure OSPF authentication to meet high security requirements of some mission-critical networks.
- Configure OSPF network management functions, such as binding OSPF MIB with a process, sending trap information and collecting log information.

Prerequisites

Before configuring OSPF network optimization, you have configured:

- IP addresses for interfaces;
- OSPF basic functions.

Configuring OSPF Packet Timers

You can configure the following timers on OSPF interfaces as needed:

- Hello timer: Interval for sending hello packets. It must be identical on OSPF neighbors. The longer the interval, the lower convergence speed and smaller network load.
- Poll timer: Interval for sending hello packets to the neighbor that is down on the NBMA network.
- Dead timer: Interval within which if the interface receives no hello packet from the neighbor, it declares the neighbor is down.
- LSA retransmission timer: Interval within which if the interface receives no acknowledgement packets after sending a LSA to the neighbor, it will retransmit the LSA.

Follow these steps to configure timers for OSPF packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the hello interval	**ospf timer hello** *seconds*	Optional The hello interval on P2P, Broadcast interfaces defaults to 10 seconds and defaults to 30 seconds on P2MP and NBMA interfaces.
Specify the poll interval	**ospf timer poll** *seconds*	Optional The poll interval defaults to 120 seconds.
Specify the dead interval	**ospf timer dead** *seconds*	Optional The default dead interval is 40 seconds on P2P, Broadcast interfaces and 120 seconds on P2MP and NBMA interfaces.
Specify the retransmission interval	**ospf timer retransmit** *interval*	Optional The retransmission interval defaults to 5 seconds.

Note

- The hello and dead intervals restore to default values after you change the network type for an interface.
- The dead interval should be at least four times the hello interval on an interface.
- The poll interval is at least four times the hello interval.
- The retransmission interval should not be so small for avoidance of unnecessary LSA retransmissions. In general, this value is bigger than the round-trip time of a packet between two adjacencies.

Specifying an LSA Transmission Delay

Since OSPF packets need time for traveling on links, extending LSA age time with a delay is necessary, especially for low speed links.

Follow these steps to specify an LSA transmission delay on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify an LSA transmission delay	**ospf trans-delay** *seconds*	Optional 1 second by default

Specifying SPF Calculation Interval

The LSDB changes lead to SPF calculations. When an OSPF network changes frequently, a large amount of network resources will be occupied, reducing the working efficiency of routers. You can adjust the SPF calculation interval for the network to reduce negative influence.

Follow these steps to configure SPF calculation interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Specify SPF calculation interval(s)	**spf-schedule-interval** *maximum-interval* [*minimum-interval* [*incremental-interval*]]	Optional By default, the interval is 5 seconds.

Note

With this task configured, when network changes are not frequent, SPF calculation applies at the *minimum-interval*. If network changes become frequent, SPF calculation interval is incremented by *incremental-interval* × 2^{n-2} (n is the number of calculation times) each time a calculation occurs, up to the *maximum-interval*.

Specifying the LSA Minimum Repeat Arrival Interval

After receiving the same LSA as the previously received LSA within the LSA minimum repeat arrival interval, an interface discards the LSA.

Follow these steps to configure the LSA minimum repeat arrival interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the LSA minimum repeat arrival interval	**lsa-arrival-interval** *interval*	Optional Defaults to 1000 milliseconds.

Note

The interval set with the **lsa-arrival-interval** command should be smaller or equal to the interval set with the **lsa-generation-interva**l command.

Specifying the LSA Generation Interval

With this feature configured, you can protect network resources and routers from being over consumed due to frequent network changes.

Follow these steps to configure the LSA generation interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required
Configure the LSA generation interval	**lsa-generation-interval** *maximum-interval* [*initial-interval* [*incremental-interval*]]	Optional By default, the maximum interval is 5 seconds, the minimum interval is 0 milliseconds and the incremental interval is 5000 milliseconds.

Note

With this command configured, when network changes are not frequent, LSAs are generated at the *minimum-interval*. If network changes become frequent, LSA generation interval is incremented by *incremental-interval*•2n-2 (n is the number of generation times) each time a generation occurs, up to the *maximum-interval*.

Disabling Interfaces from Sending OSPF Packets

Follow these steps to disable interfaces from sending routing information:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Disable interfaces from sending OSPF packets	**silent-interface** { **all** | *interface-type interface-number* }	Optional Not disabled by default

Note

- Different OSPF processes can disable the same interface from sending OSPF packets. Use of the **silent-interface** command disables only the interfaces associated with the current process rather than interfaces associated with other processes.
- After an OSPF interface is set to silent, other interfaces on the router can still advertise direct routes of the interface in Router LSAs, but no OSPF packet can be advertised for the interface to find a neighbor. This configuration can enhance adaptability of OSPF networking and reduce resource consumption.

Configuring Stub Routers

A stub router is used for traffic control. It tells other OSPF routers not to use it to forward data, but they can have a route to it.

The Router LSAs from the stub router may contain different link type values. A value of 3 means a link to the stub network, so the cost of the link remains unchanged. A value of 1, 2 or 4 means a point-to-point link, a link to a transit network or a virtual link. In such cases, a maximum cost value of 65535 is used. Thus, other neighbors find the links to the stub router have such big costs, they will not send packets to the stub router for forwarding as long as there is a route with a smaller cost.

Follow these steps to configure a router as a stub router:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the router as a stub router	**stub-router**	Required Not configured by default.

Note

A stub router has nothing to do with a stub area.

Configuring OSPF Authentication

OSPF supports packet authentication to ensure the security of packet exchange.

After authentication is configured, OSPF only receives packets that pass authentication, so failed packets cannot establish neighboring relationships.

To configure OSPF authentication, you need to configure the same area authentication mode on all the routers in the area. In addition, the authentication mode and password for all interfaces attached to the same area must be identical.

Follow these steps to configure OSPF authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enter area view	**area** *area-id*	—
Configure the authentication mode	**authentication-mode** { **md5** | **simple** }	Required Not configured by default.
Exit to OSPF view	**quit**	—
Exit to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the authentication mode (simple authentication) for the interface	**ospf authentication-mode simple** [**cipher** | **plain**] *password*	Either is required. Not configured by default.
Configure the authentication mode (MD5 authentication) for the interface	**ospf authentication-mode** { **hmac-md5** | **md5** } *key-id* [**cipher** | **plain**] *password*	

Adding the Interface MTU into DD Packets

Generally, when an interface sends a DD packet, it adds 0 into the Interface MTU field of the DD packet rather than the interface MTU.

Follow these steps to add the interface MTU into DD packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable OSPF to add the interface MTU into DD packets	**ospf mtu-enable**	Optional Not enabled by default.; that is, the interface fills in a value of 0.

Configuring the Maximum Number of External LSAs in LSDB

Follow these steps to configure the maximum number of external LSAs in the Link State Database:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Specify the maximum number of external LSAs in the LSDB	**lsdb-overflow-limit** *number*	Optional Not specified by default

Making External Route Selection Rules Defined in RFC1583 Compatible

The selection of an external route from multiple LSAs defined in RFC2328 is different from the one defined in RFC1583. If RFC1583 is made compatible with RFC 2328, the routes in the backbone area are preferred; if not, the routes in the non-backbone area are preferred to reduce the burden of the backbone area.

Follow these steps to make them compatible:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required
Make RFC1583 compatible	**rfc1583 compatible**	Optional Compatible by default

Note

To avoid routing loops, it is recommended to configure all the routers to be either compatible or incompatible with the external route selection rules defined in RFC 1583.

Logging Neighbor State Changes

Follow these steps to enable the logging of neighbor state changes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enable the logging of neighbor state changes	**log-peer-change**	Optional Enabled by default

Configuring OSPF Network Management

After trap generation is enabled for OSPF, OSPF generates traps to report important events. Traps fall into the following levels:

- Level-3, for fault traps
- Level-4, for alarm traps
- Level-5, for normal but important traps
- Level-6, for notification traps

The generated traps are sent to the Information Center of the device. The output rules of the traps, namely, whether to output the traps and the output direction, are determined according to the Information Center configuration. (For Information Center configuration, refer to *Information Center Configuration* in the *System Volume*.)

Follow these steps to configure OSPF network management:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Bind OSPF MIB to an OSPF process	**ospf mib-binding** *process-id*	Optional The first OSPF process is bound with OSPF MIB by default.
Enable OSPF trap generation	**snmp-agent trap enable ospf** [*process-id*] [**ifauthfail** | **ifcfgerror** | **ifrxbadpkt** | **ifstatechange** | **iftxretransmit** | **lsdbapproachoverflow** | **lsdboverflow** | **maxagelsa** | **nbrstatechange** | **originatelsa** | **vifcfgerror** | **virifauthfail** | **virifrxbadpkt** | **virifstatechange** | **viriftxretransmit** | **virnbrstatechange**] *	Optional Enabled by default

Enabling Message Logging

Follow these steps to enable message logging:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enable message logging	**enable log** [**config** | **error** | **state**]	Required Not enabled by default.

Enabling the Advertisement and Reception of Opaque LSAs

With this feature enabled, the OSPF router can receive and advertise Type 9, Type 10 and Type 11 opaque LSAs.

Follow these steps to enable the advertisement and reception of opaque LSAs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Enable the advertisement and reception of opaque LSAs	**opaque-capability enable**	Optional Disabled by default

Configuring OSPF to Give Priority to Receiving and Processing Hello Packets

To ensure OSPF runs normally, a router receives and processes Hello packets and other protocol packets at the same time. When the router has established neighbor relationships with multiple neighboring routers and the routing table size is big, the router will need to receive and process large numbers of packets. Configuring OSPF to give priority to receiving and processing Hello packets helps ensure stable neighbor relationships.

Follow these steps to configure OSPF to give priority to receiving and processing Hello packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure OSPF to give priority to receiving and processing Hello packets	**ospf packet-process prioritized-treatment**	Required Not configured by default.

Configuring the LSU Transmit Rate

Sending large numbers of LSU packets for LSDB synchronization with neighbors may affect router performance and consume large network bandwidths. Therefore, you can configure the router to send LSU packets at a proper interval and limit the maximum number of LSU packets sent out an OSPF interface each time.

Follow these steps to configure the LSU transmit rate:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	—
Configure the LSU transmit rate	**transmit-pacing interval** *interval* **count** *count*	Optional By default, an OSPF interface sends up to three LSU packets every 20 milliseconds.

Configuring OSPF Sham Link

The sham link is considered an OSPF intra-area route. It is used to ensure that the VPN traffic is transmitted over the backbone instead of the backdoor link between two CEs.

The source and destination addresses of the sham link must be loopback interface addresses with 32-bit masks. Besides, the loopback interfaces must be bound to the VPN instances and be advertised through BGP.

Configuration Prerequisites

Before configuring OSPF sham link, be sure to configure OSPF in the LAN where CEs reside

Configuring a Loopback Interface

Follow these steps to configure a loopback interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a loopback interface and enter loopback interface view	**interface loopback** *interface-number*	Required
Bind the loopback interface to VPN instance	**ip binding vpn-instance** *vpn-instance-name*	Required By default, an interface is associated with no VPN instance.
Configure the address of the loopback interface	**ip address** *ip-address* { *mask* | *mask-length* }	Required

Advertising Routes of a Loopback Interface

Follow these steps to advertise routes of a loopback interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	Required

To do…	Use the command…	Remarks
Enter BGP VPN instance view	**ipv4-family vpn-instance** *vpn-instance-name*	Required
Inject direct routes, that is, loopback host routes	**import-route direct**	Required

Note

For BGP VPN information, refer to *MCE Configuration* in the *IP Routing Volume*.

Creating a Sham Link

Follow these steps to create a sham link:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *	—
Configure the route tag	**route-tag** *tag-value*	Required
Enter OSPF area view	**area** *area-id*	Required
Configure a sham link	**sham-link** *source-ip-address destination-ip-address* [**cost** *cost* | **dead** *dead-interval* | **hello** *hello-interval* | **retransmit** *retrans-interval* | **trans-delay** *delay* | **simple** [**cipher** | **plain**] *password* | { **md5** | **hmac-md5** } *key-id* [**cipher** | **plain**] *password*]*	Required By default, no sham link is configured.

Note

- If you start OSPF but do not configure the router ID, the system will automatically elect one. However, the same election rules produce the same router ID. Therefore, you are recommended to configure the router ID when starting an OSPF process. For the election rules, refer to *OSPF Configuration* in the *IP Routing Volume*.
- If you configure multiple OSPF VPN instances but do not configure the route tag, the system will automatically create one based on the AS number configured. If you do not configure BGP, the tag will be 0.

Configuring OSPF Graceful Restart

Note

One device can act as both a GR Restarter and GR Helper at the same time.

Configuring the OSPF GR Restarter

You can configure the IETF standard or non IETF standard OSPF GR Restarter.

Configure the IETF standard OSPF GR Restarter

Follow these steps to configure the standard IETF OSPF GR Restarter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF and enter its view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Disabled by default
Enable opaque LSA advertisement capability	**opaque-capability enable**	Required Disabled by default
Enable the IETF standard Graceful Restart capability for OSPF	**graceful-restart ietf**	Required Disabled by default
Configure the Graceful Restart interval for OSPF	**graceful-restart interval** *timer*	Optional 120 seconds by default

Configure the non-IETF standard OSPF GR Restarter

Follow these steps to configure non-IETF standard OSPF GR Restarter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF and enter its view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Disabled by default
Enable the link-local signaling capability	**enable link-local-signaling**	Required Disabled by default
Enable the out-of-band re-synchronization capability	**enable out-of-band-resynchronization**	Required Disabled by default
Enable non IETF standard Graceful Restart capability for OSPF	**graceful-restart** [**nonstandard**]	Required Disabled by default

To do…	Use the command…	Remarks
Configure Graceful Restart interval for OSPF	**graceful-restart interval** *timer*	Optional 120 seconds by default

Configuring the OSPF GR Helper

You can configure the IETF standard or non IETF standard OSPF GR Helper.

Configuring the IETF standard OSPF GR Helper

Follow these steps to configure the IETF standard OSPF GR Helper:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF and enter its view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Disabled by default
Enable opaque LSA reception and advertisement	**opaque-capability enable**	Required Not enabled by default.
Configure the neighbors for which the router can serve as a GR Helper	**graceful-restart help** { *acl-number* | **prefix** *prefix-list* }	Optional The router can server as a GR Helper for any OSPF neighbor by default.

Configuring the non IETF standard OSPF GR Helper

Follow these steps to configure the non IETF standard OSPF GR Helper:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF and enter its view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Disabled by default
Enable the link-local signaling capability	**enable link-local-signaling**	Required Disabled by default
Enable the out-of-band re-synchronization capability	**enable out-of-band-resynchronization**	Required Disabled by default
Configure the neighbors for which the router can serve as a GR Helper	**graceful-restart help** { *acl-number* | **prefix** *prefix-list* }	Optional The router can server as a GR Helper for any OSPF neighbor by default.

Triggering OSPF Graceful Restart

Performing an active/standby switchover on a distributed device, or performing the following configuration on an OSPF router will trigger an OSPF Graceful Restart process.

For the IETF standard GR capable routers, ensure they have the following capabilities enabled:

- Opaque LSA advertisement
- IETF standard GR

For the non IETF standard GR capable routers, ensure they have the following capabilities enabled:

- link local signaling
- out of band re-synchronization
- Non IETF standard GR

Follow these steps to trigger OSPF Graceful Restart:

To do…	Use the command…	Remarks
Trigger OSPF Graceful Restart	**reset ospf** [*process-id*] **process graceful-restart**	Required Available in user view

Displaying and Maintaining OSPF

To do…	Use the command…	Remarks
Display OSPF brief information	**display ospf** [*process-id*] **brief**	Available in any view
Display OSPF statistics	**display ospf** [*process-id*] **cumulative**	
Display Link State Database information	**display ospf** [*process-id*] **lsdb** [**brief** | [{ **ase** | **router** | **network** | **summary** | **asbr** | **nssa** | **opaque-link** | **opaque-area** | **opaque-as** } [*link-state-id*]] [**originate-router** *advertising-router-id* | **self-originate**]]	
Display OSPF neighbor information	**display ospf** [*process-id*] **peer** [**verbose |** [*interface-type interface-number*] [*neighbor-id*]]	
Display neighbor statistics of OSPF areas	**display ospf** [*process-id*] **peer statistics**	
Display next hop information	**display ospf** [*process-id*] **nexthop**	
Display routing table information	**display ospf** [*process-id*] **routing** [**interface** *interface-type interface-number*] [**nexthop** *nexthop-address*]	
Display virtual link information	**display ospf** [*process-id*] **vlink**	
Display information about OSPF sham links	**display ospf** [*process-id*] **sham-link** [**area** *area-id*]	
Display OSPF request queue information	**display ospf** [*process-id*] **request-queue** [*interface-type interface-number*] [*neighbor-id*]	
Display OSPF retransmission queue information	**display ospf** [*process-id*] **retrans-queue** [*interface-type interface-number*] [*neighbor-id*]	
Display OSPF ABR and ASBR information	**display ospf** [*process-id*] **abr-asbr**	
Display OSPF interface information	**display ospf** [*process-id*] **interface** [**all** | *interface-type interface-number*]	
Display OSPF error information	**display ospf** [*process-id*] **error**	
Display OSPF ASBR summarization information	**display ospf** [*process-id*] **asbr-summary** [*ip-address* { *mask* | *mask-length* }]	

To do…	Use the command…	Remarks
Reset OSPF counters	**reset ospf** [*process-id*] **counters** [**neighbor** [*interface-type interface-number*] [*router-id*]]	Available in user view
Reset an OSPF process	**reset ospf** [*process-id*] **process** [**graceful-restart**]	
Re-enable OSPF route redistribution	**reset ospf** [*process-id*] **redistribution**	

OSPF Configuration Examples

Caution

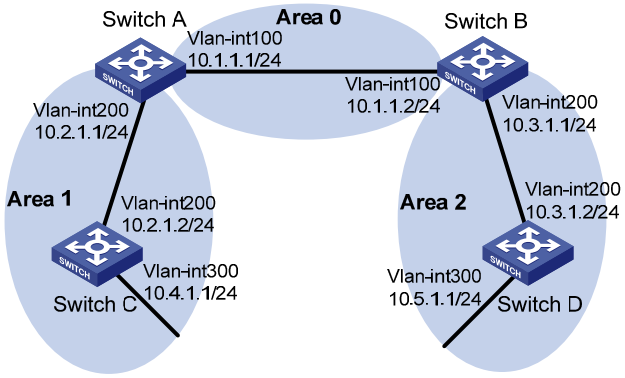
These examples only cover commands for OSPF configuration.

Configuring OSPF Basic Functions

Network requirements

- As shown in the following figure, all switches run OSPF. The AS is split into three areas, in which, Switch A and Switch B act as ABRs to forward routing information between areas.
- After configuration, all switches can learn routes to every network segment in the AS.

Figure 1-21 Network diagram for OSPF basic configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
```

```
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] area 2
[SwitchB-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.2] quit
[SwitchB-ospf-1] quit
```

Configure Switch C

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] network 10.4.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Configure Switch D

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] network 10.5.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] quit
[SwitchD-ospf-1] quit
```

3) Verify the configuration

Display information about neighbors on Switch A.

```
[SwitchA] display ospf peer verbose

          OSPF Process 1 with Router ID 10.2.1.1
                 Neighbors

 Area 0.0.0.0 interface 10.1.1.1(Vlan-interface100)'s neighbors
 Router ID: 10.3.1.1         Address: 10.1.1.2         GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 10.1.1.1  BDR: 10.1.1.2  MTU: 0
   Dead timer due in 37  sec
   Neighbor is up for 06:03:59
   Authentication Sequence: [ 0 ]
```

```
  Neighbor state change count: 5

                Neighbors

 Area 0.0.0.1 interface 10.2.1.1(Vlan-interface200)'s neighbors
 Router ID: 10.4.1.1         Address: 10.2.1.2         GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 10.2.1.1  BDR: 10.2.1.2  MTU: 0
   Dead timer due in 32  sec
   Neighbor is up for 06:03:12
   Authentication Sequence: [ 0 ]
   Neighbor state change count: 5
```

Display OSPF routing information on Switch A.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 10.2.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        10       Transit 10.2.1.1        10.2.1.1        0.0.0.1
 10.3.1.0/24        4        Inter   10.1.1.2        10.3.1.1        0.0.0.0
 10.4.1.0/24        13       Stub    10.2.1.2        10.4.1.1        0.0.0.1
 10.5.1.0/24        14       Inter   10.1.1.2        10.3.1.1        0.0.0.0
 10.1.1.0/24        2        Transit 10.1.1.1        10.2.1.1        0.0.0.0

 Total Nets: 5
 Intra Area: 3  Inter Area: 2  ASE: 0  NSSA: 0
```

Display the Link State Database on Switch A.

```
[SwitchA] display ospf lsdb

          OSPF Process 1 with Router ID 10.2.1.1
                  Link State Database

                          Area: 0.0.0.0
 Type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    10.2.1.1        10.2.1.1          1069  36    80000012       0
 Router    10.3.1.1        10.3.1.1           780  36    80000011       0
 Network   10.1.1.1        10.2.1.1          1069  32    80000010       0
 Sum-Net   10.5.1.0        10.3.1.1           780  28    80000003      12
 Sum-Net   10.2.1.0        10.2.1.1          1069  28    8000000F      10
 Sum-Net   10.3.1.0        10.3.1.1           780  28    80000014       2
 Sum-Net   10.4.1.0        10.2.1.1           769  28    8000000F      13
                          Area: 0.0.0.1
 Type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    10.2.1.1        10.2.1.1           769  36    80000012       0
 Router    10.4.1.1        10.4.1.1          1663  48    80000012       0
```

```
 Network   10.2.1.1        10.2.1.1          769  32    80000010       0
 Sum-Net   10.5.1.0        10.2.1.1          769  28    80000003      14
 Sum-Net   10.3.1.0        10.2.1.1         1069  28    8000000F       4
 Sum-Net   10.1.1.0        10.2.1.1         1069  28    8000000F       2
 Sum-Asbr  10.3.1.1        10.2.1.1         1069  28    8000000F       2
```

Display OSPF routing information on Switch D.

```
[SwitchD] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2

 Total Nets: 5
 Intra Area: 2  Inter Area: 3  ASE: 0  NSSA: 0
```

On Switch D, ping the IP address 10.4.1.1 to check connectivity.

```
[SwitchD] ping 10.4.1.1
  PING 10.4.1.1: 56  data bytes, press CTRL_C to break
    Request time out
    Reply from 10.4.1.1: bytes=56 Sequence=2 ttl=253 time=2 ms
    Reply from 10.4.1.1: bytes=56 Sequence=3 ttl=253 time=1 ms
    Reply from 10.4.1.1: bytes=56 Sequence=4 ttl=253 time=1 ms
    Reply from 10.4.1.1: bytes=56 Sequence=5 ttl=253 time=1 ms

  --- 10.4.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 1/1/2 ms
```

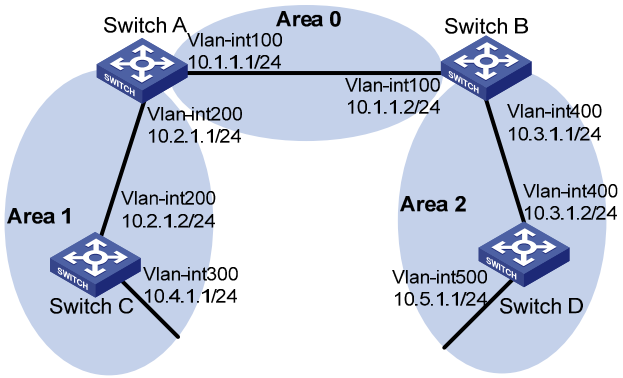
Configuring OSPF Route Redistribution

Network requirements

As shown in the following figure:

- All the switches run OSPF, and the AS is divided into three areas.
- Switch A and Switch B act as ABRs to forward routes between areas.
- Switch C is configured as an ASBR to redistribute external routes (static routes). Routing information is propagated properly in the AS.

Figure 1-22 Network diagram for OSPF redistributing routes from outside of an AS



Configuration procedure

1) Configure IP addresses for interfaces (omitted).
2) Configure OSPF basic functions (Refer to Configuring OSPF Basic Functions).
3) Configure OSPF to redistribute routes.

On Switch C, configure a static route destined for network 3.1.2.0/24.

```
<SwitchC> system-view
[SwitchC] ip route-static 3.1.2.1 24 10.4.1.2
```

On Switch C, configure OSPF to redistribute static routes.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route static
```

4) Verify the configuration.

Display the ABR/ASBR information of Switch D.

```
<SwitchD> display ospf abr-asbr

          OSPF Process 1 with Router ID 10.5.1.1
                 Routing Table to ABR and ASBR

 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.3.1.1        0.0.0.2         10    10.3.1.1        ABR
 Inter       10.4.1.1        0.0.0.2         22    10.3.1.1        ASBR
```

Display the OSPF routing table of Switch D.

```
<SwitchD> display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
```

```
10.1.1.0/24        12        Inter    10.3.1.1        10.3.1.1        0.0.0.2

Routing for ASEs
Destination        Cost      Type     Tag             NextHop         AdvRouter
3.1.2.0/24         1         Type2    1               10.3.1.1        10.4.1.1

Total Nets: 6
Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

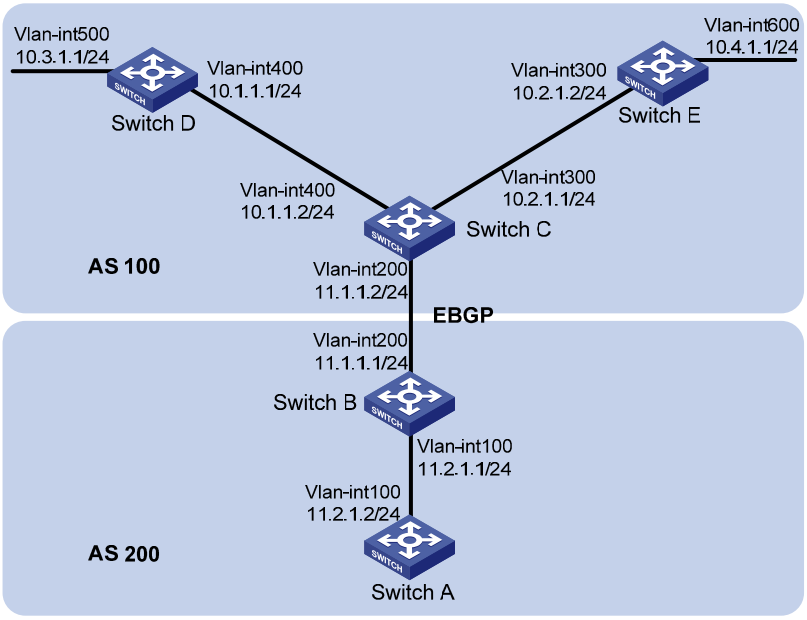
Configuring OSPF to Advertise a Summary Route

Network requirements

As shown in the following figure:

- Switch A and Switch B are in AS 200, which runs OSPF.
- Switch C, Switch D, and Switch E are in AS 100, which runs OSPF.
- An eBGP connection is established between Switch B and Switch C. Switch C is configured to redistribute OSPF routes into BGP.
- Switch B is configured to redistribute BGP routes into OSPF. Switch B is configured with route summarization and advertises only the summary route 10.0.0.0/8 to reduce Switch A's routing table size.

Figure 1-23 Network diagram for OSPF summary route advertisement (on switches)



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
```

```
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 11.2.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] ospf
[SwitchE-ospf-1] area 0
[SwitchE-ospf-1-area-0.0.0.0] network 10.2.1.0 0.0.0.255
[SwitchE-ospf-1-area-0.0.0.0] network 10.4.1.0 0.0.0.255
[SwitchE-ospf-1-area-0.0.0.0] quit
[SwitchE-ospf-1] quit
```

3) Configure BGP

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp] peer 11.1.1.2 as 100
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 100
[SwitchC-bgp] peer 11.1.1.1 as 200
[SwitchC-bgp] import-route ospf
```

4) Configure route redistribution on Switch B.

Configure OSPF to redistribute routes from BGP on Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route bgp
```

Display the OSPF routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 8        Routes : 8

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         O_ASE  150  1            11.2.1.1        Vlan100
10.2.1.0/24         O_ASE  150  1            11.2.1.1        Vlan100
10.3.1.0/24         O_ASE  150  1            11.2.1.1        Vlan100
10.4.1.0/24         O_ASE  150  1            11.2.1.1        Vlan100
11.2.1.0/24         Direct 0    0            11.2.1.2        Vlan100
11.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

5) # Configure summary route 10.0.0.0/8 on Switch B and advertise it.

```
[SwitchB-ospf-1] asbr-summary 10.0.0.0 8
```

Display the OSPF routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.0.0.0/8          O_ASE  150  2            11.2.1.1        Vlan100
11.2.1.0/24         Direct 0    0            11.2.1.2        Vlan100
11.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

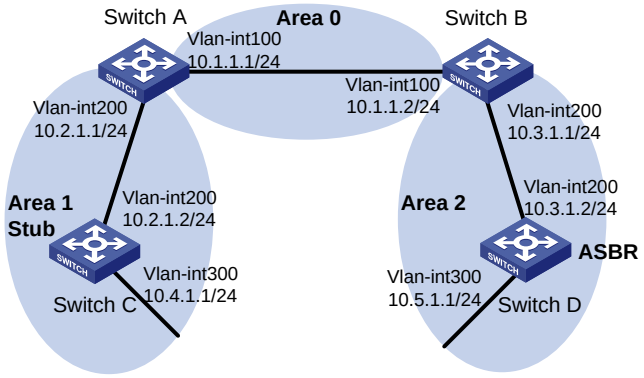
Configuring an OSPF Stub Area

Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. Switch A and Switch B act as ABRs to forward routing information between areas. Switch D acts as the ASBR to redistribute routes (static routes).

It is required to configure Area 1 as a Stub area, reducing LSAs to this area without affecting route reachability.

Figure 1-24 Network diagram for OSPF Stub area configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted).
2) Configure OSPF basic functions (refer to Configuring OSPF Basic Functions).
3) Configure Switch D to redistribute static routes.

```
[SwitchD] ip route-static 3.1.2.1 24 10.5.1.2
[SwitchD] ospf
[SwitchD-ospf-1] import-route static
[SwitchD-ospf-1] quit
```

Display ABR/ASBR information on Switch C.

```
[SwitchC] display ospf abr-asbr

          OSPF Process 1 with Router ID 10.4.1.1
                 Routing Table to ABR and ASBR

 Type        Destination     Area            Cost  Nexthop         RtType
 Intra       10.2.1.1        0.0.0.1         3     10.2.1.1        ABR
 Inter       10.3.1.1        0.0.0.1         5     10.2.1.1        ABR
 Inter       10.5.1.1        0.0.0.1         7     10.2.1.1        ASBR
```

Display OSPF routing table information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        3        Transit 10.2.1.2        10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Routing for ASEs
```

```
Destination        Cost     Type    Tag         NextHop         AdvRouter
3.1.2.0/24         1        Type2   1           10.2.1.1        10.5.1.1

Total Nets: 6
Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

Note

In the above output, since Switch C resides in a normal OSPF area, its routing table contains an external route.

---

4) Configure Area 1 as a Stub area.

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] stub
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Display OSPF routing information on Switch C

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.2.1.0/24        3        Transit 10.2.1.2        10.2.1.1        0.0.0.1
 10.3.1.0/24        7        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1
 10.5.1.0/24        17       Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.1.1.0/24        5        Inter   10.2.1.1        10.2.1.1        0.0.0.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 4  ASE: 0  NSSA: 0
```

Note

When Switch C resides in the Stub area, a default route takes the place of the external route.

Filter Type-3 LSAs out the stub area

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub no-summary
[SwitchA-ospf-1-area-0.0.0.1] quit
```

Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 0.0.0.0/0          4        Inter   10.2.1.1        10.2.1.1        0.0.0.1
 10.2.1.0/24        3        Transit 10.2.1.2        10.4.1.1        0.0.0.1
 10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

Note

After this configuration, routing entries on the stub router are further reduced, containing only one default external route.

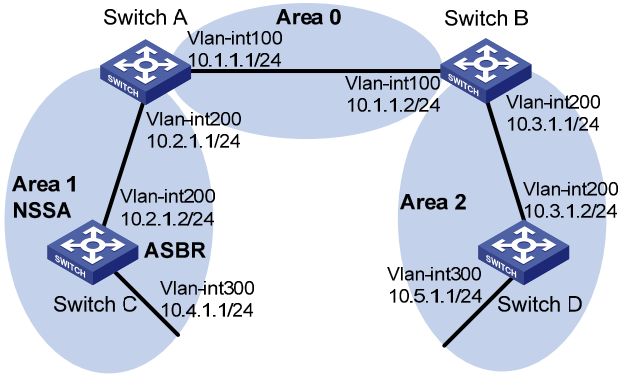
Configuring an OSPF NSSA Area

Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. Switch A and Switch B act as ABRs to forward routing information between areas.

It is required to configure Area 1 as an NSSA area, and configure Router C as the ASBR to redistribute static routes into the AS.

Figure 1-25 Network diagram for OSPF NSSA area configuration



Configuration procedure

1) Configure IP addresses for interfaces.
2) Configure OSPF basic functions (refer to Configuring OSPF Basic Functions).
3) Configure Area 1 as an NSSA area.

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] nssa default-route-advertise no-summary
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] nssa
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Note

It is recommended to configure the **nssa** command with the keyword **default-route-advertise no-summary** on Switch A (an ABR) to reduce the routing table size on NSSA routers. On other NSSA routers, use the **nssa** command.

Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
```

```
0.0.0.0/0          65536    Inter   10.2.1.1        10.2.1.1        0.0.0.1
10.2.1.0/24        65535    Transit 10.2.1.2        10.4.1.1        0.0.0.1
10.4.1.0/24        3        Stub    10.4.1.1        10.4.1.1        0.0.0.1

Total Nets: 3
Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

4) Configure Switch C to redistribute static routes.

```
[SwitchC] ip route-static 3.1.3.1 24 11.1.1.1
[SwitchC] ospf
[SwitchC-ospf-1] import-route static
[SwitchC-ospf-1] quit
```

Display OSPF routing information on Switch D.

```
[SwitchD-ospf-1] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        22       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.3.1.0/24        10       Transit 10.3.1.2        10.3.1.1        0.0.0.2
 10.4.1.0/24        25       Inter   10.3.1.1        10.3.1.1        0.0.0.2
 10.5.1.0/24        10       Stub    10.5.1.1        10.5.1.1        0.0.0.2
 10.1.1.0/24        12       Inter   10.3.1.1        10.3.1.1        0.0.0.2

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop         AdvRouter
 3.1.3.0/24         1        Type2   1           10.3.1.1        10.2.1.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

Note

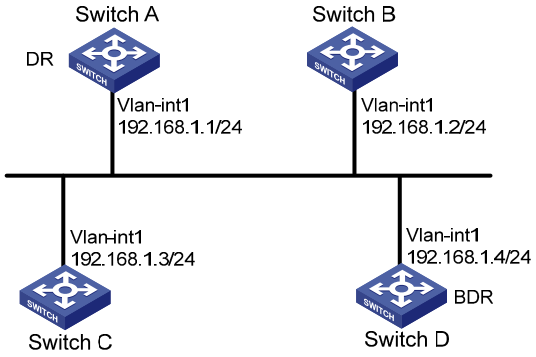
You can see on Switch D an external route imported from the NSSA area.

Configuring OSPF DR Election

Network requirements

- In the following figure, OSPF Switches A, B, C and D reside on the same network segment.
- It is required to configure Switch A as the DR, and configure Switch C as the BDR.

Figure 1-26 Network diagram for OSPF DR election configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] router id 1.1.1.1
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] router id 2.2.2.2
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] router id 3.3.3.3
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] router id 4.4.4.4
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
```

```
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

Display OSPF neighbor information on Switch A.

```
[SwitchA] display ospf peer verbose

          OSPF Process 1 with Router ID 1.1.1.1
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.1(Vlan-interface1)'s neighbors
 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 38  sec
   Neighbor is up for 00:01:31
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:01:28
   Authentication Sequence: [ 0 ]

 Router ID: 4.4.4.4          Address: 192.168.1.4      GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:01:28
   Authentication Sequence: [ 0 ]
```

Switch D becomes the DR, and Switch C is the BDR.

3) Configure router priorities on interfaces

Configure Switch A.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ospf dr-priority 100
[SwitchA-Vlan-interface1] quit
```

Configure Switch B.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ospf dr-priority 0
[SwitchB-Vlan-interface1] quit
```

Configure Switch C.

```
[SwitchC] interface vlan-interface 1
[SwitchC-Vlan-interface1] ospf dr-priority 2
[SwitchC-Vlan-interface] quit
```

Display neighbor information on Switch D.

```
[SwitchD] display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
 Router ID: 1.1.1.1      Address: 192.168.1.1      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 100
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:11:17
   Authentication Sequence: [ 0 ]

 Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 0
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 35  sec
   Neighbor is up for 00:11:19
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
   State: Full  Mode:Nbr is  Slave  Priority: 2
   DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 33  sec
   Neighbor is up for 00:11:15
   Authentication Sequence: [ 0 ]
```

The DR and BDR have no change.

 **Note**

In the above output, you can find the priority configuration does not take effect immediately.

---

4) Restart OSPF process

Restart the OSPF process of Switch D.

```
<SwitchD> reset ospf 1 process
Warning : Reset OSPF process? [Y/N]:y
```

Display neighbor information on Switch D.

```
[SwitchD] display ospf peer verbose

          OSPF Process 1 with Router ID 4.4.4.4
                  Neighbors

 Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
```

```
 Router ID: 1.1.1.1          Address: 192.168.1.1      GR State: Normal
   State: Full  Mode: Nbr is Slave  Priority: 100
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 39  sec
   Neighbor is up for 00:01:40
   Authentication Sequence: [ 0 ]

 Router ID: 2.2.2.2          Address: 192.168.1.2      GR State: Normal
   State: 2-Way  Mode: None  Priority: 0
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 35  sec
   Neighbor is up for 00:01:44
   Authentication Sequence: [ 0 ]

 Router ID: 3.3.3.3          Address: 192.168.1.3      GR State: Normal
   State: Full  Mode: Nbr is Slave  Priority: 2
   DR: 192.168.1.1  BDR: 192.168.1.3  MTU: 0
   Dead timer due in 39  sec
   Neighbor is up for 00:01:41
   Authentication Sequence: [ 0 ]
```

Switch A becomes the DR, and Switch C is the BDR.

Note

If the neighbor state is *full,* it means Switch D has established the adjacency with the neighbor. If the neighbor state is *2-way*, it means the two switches are neither the DR nor the BDR, and they do not exchange LSAs.

Display OSPF interface information.

```
[SwitchA] display ospf interface

          OSPF Process 1 with Router ID 1.1.1.1
                  Interfaces

 Area: 0.0.0.0
 IP Address      Type      State   Cost  Pri   DR             BDR
 192.168.1.1     Broadcast DR      1     100   192.168.1.1    192.168.1.3

[SwitchB] display ospf interface

          OSPF Process 1 with Router ID 2.2.2.2
                  Interfaces

 Area: 0.0.0.0
 IP Address      Type      State    Cost  Pri   DR           BDR
```

```
192.168.1.2     Broadcast DROther  1     0     192.168.1.1   192.168.1.3
```

Note

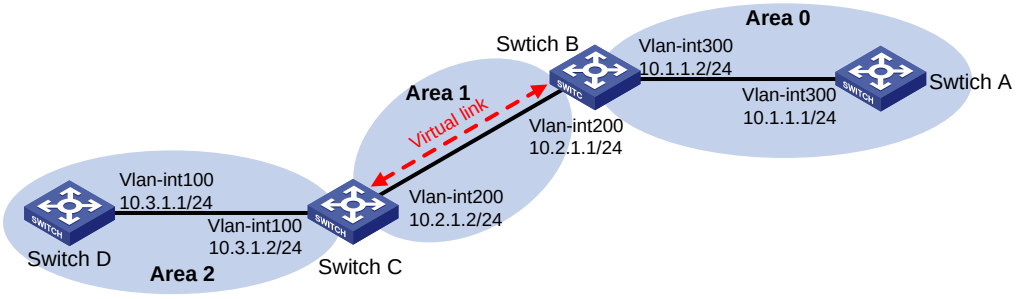
The interface state *DROther* means the interface is not the DR/BDR.

Configuring OSPF Virtual Links

Network requirements

- In the following figure, Area 2 has no direct connection to Area 0, and Area 1 acts as the Transit Area to connect Area 2 to Area 0 via a configured virtual link between Switch B and Switch C.
- After configuration, Switch B can learn routes to Area 2.

Figure 1-27 Network diagram for OSPF virtual link configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf 1 router-id 1.1.1.1
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1 router-id 2.2.2.2
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] area 1
[SwitchB–ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchB–ospf-1-area-0.0.0.1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf 1 router-id 3.3.3.3
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] area 2
[SwitchC–ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchC–ospf-1-area-0.0.0.2] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf 1 router-id 4.4.4.4
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] quit
```

Display the OSPF routing table of Switch B.

```
[SwitchB] display ospf routing

          OSPF Process 1 with Router ID 2.2.2.2
                   Routing Tables
 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0
 Total Nets: 2
 Intra Area: 2  Inter Area: 0  ASE: 0  NSSA: 0
```

Note

Since Area 0 has no direct connection to Area 2, the routing table of Switch B has no route to Area 2.

---

3) Configure a virtual link

Configure Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] vlink-peer 3.3.3.3
[SwitchB-ospf-1-area-0.0.0.1] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] vlink-peer 2.2.2.2
[SwitchC-ospf-1-area-0.0.0.1] quit
```

Display the OSPF routing table of Switch B.

```
[SwitchB] display ospf routing
          OSPF Process 1 with Router ID 2.2.2.2
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter       Area
 10.2.1.0/24        2        Transit 10.2.1.1        3.3.3.3         0.0.0.1
 10.3.1.0/24        5        Inter   10.2.1.2        3.3.3.3         0.0.0.0
 10.1.1.0/24        2        Transit 10.1.1.2        2.2.2.2         0.0.0.0

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

Switch B has learned the route 10.3.1.0/24 to Area 2.

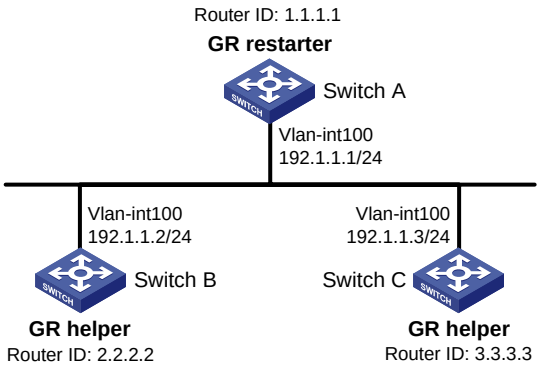
OSPF Graceful Restart Configuration Example

Network requirements

As shown in the following figure:

- Switch A, Switch B and Switch C that belong to the same autonomous system and the same OSPF routing domain are GR capable.
- Switch A acts as the non IETF standard GR Restarter whereas Switch B and Switch C are the GR Helpers and re-synchronize their LSDB with Switch A through OOB communication of GR.

Figure 1-28 Network diagram for OSPF GR configuration



Configuration procedure

1) Configure Switch A

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
[SwitchA] router id 1.1.1.1
[SwitchA] ospf 100
[SwitchA-ospf-100] enable link-local-signaling
[SwitchA-ospf-100] enable out-of-band-resynchronization
[SwitchA-ospf-100] graceful-restart
[SwitchA-ospf-100] area 0
[SwitchA-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
```

```
[SwitchA-ospf-100-area-0.0.0.0] return
```

2) Configure Switch B

```
<SwitchB> system-view
[SwitchB] acl number 2000
[SwitchB-acl-basic-2000] rule 10 permit source 192.1.1.1 0.0.0.0
[SwitchB-acl-basic-2000] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.1.1.2 255.255.255.0
[SwitchB-Vlan-interface100] quit
[SwitchB] router id 2.2.2.2
[SwitchB] ospf 100
[SwitchB-ospf-100] graceful-restart help 2000
[SwitchB-ospf-100] area 0
[SwitchB-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
```

3) Configure Switch C

```
<SwitchC> system-view
[SwitchC] acl number 2000
[SwitchC-acl-basic-2000] rule 10 permit source 192.1.1.1 0.0.0.0
[SwitchC-acl-basic-2000] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ip address 192.1.1.3 255.255.255.0
[SwitchC-Vlan-interface100] quit
[SwitchC] router id 3.3.3.3
[SwitchC] ospf 100
[SwitchC-ospf-100] graceful-restart help 2000
[SwitchC-ospf-100] area 0
[SwitchC-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
```

4) Verify the configuration

After the configurations on Switch A, Switch B and Switch C are completed and the switches are running steadily, enable OSPF Graceful Restart event debugging and then perform OSPF GR on Switch A.

```
<SwitchA> debugging ospf event graceful-restart
<SwitchA> terminal monitor
<SwitchA> terminal debugging
<SwitchA> reset ospf 100 process graceful-restart
Warning : Reset OSPF process? [Y/N]:y
%Dec 12 09:36:12:500 2006 RouterA RM/3/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
192.1.1.1(Ethernet1/1) from Full to Down
OSPF 1: Intf 192.1.1.1 Rcv InterfaceDown State BackupDR -> Down.
OSPF 1 nonstandard GR Started for OSPF Router
OSPF 1 notify RM that OSPF process will enter GR.
OSPF 1 created GR wait timer, timeout interval is 40(s).
OSPF 1 created GR Interval timer,timeout interval is 120(s).
OSPF 1: Intf 192.1.1.1 Rcv InterfaceUp State Down -> Waiting.
OSPF 1: Intf 192.1.1.1 Rcv BackupSeen State Waiting -> BackupDR.
OSPF 1 created OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
```

```
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
%Oct 22 09:36:12:566 2008 RouterA RM/3/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour
192.1.1.2(Ethernet1/1) from Loading to Full
OSPF 1 restarted OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 deleted OOB Progress timer for neighbor 192.1.1.2.
OSPF 1 Gr Wait Timeout timer fired.
OSPF 1 deleted GR wait timer.
OSPF 1 deleted GR Interval timer.
OSPF 1 GR Completed for OSPF Router
OSPF 1 notified RM that OSPF process left GR.
RM notified that all protocol left GR.
OSPF 1 started flushing STALE LSA after all protocol left GR.
OSPF 1: Flush Stale Area LSAs
OSPF 1: Start Flush Stale ASE + NSSA LSAs
OSPF 1: End Flush Stale ASE + NSSA LSAs
```

Switch A completes GR with the help of Switch B.

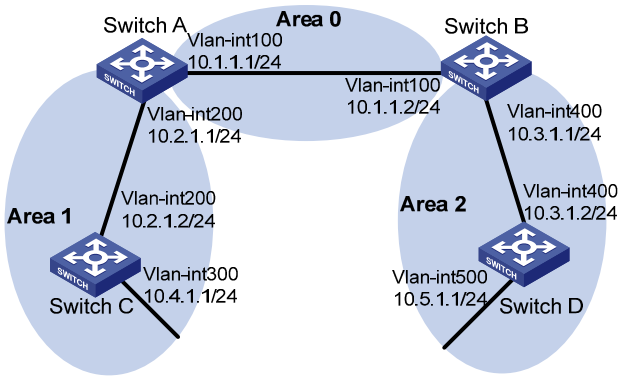
Configuring Route Filtering

Network requirements

As shown in the following figure:

- All the switches in the network run OSPF. The AS is divided into three areas.
- Switch A and Switch B work as ABRs.
- Configure Switch C as an ASBR to redistribute external routes (static routes), and configure a filter policy on Switch C to filter out redistributed route 3.1.3.0/24.
- Configure a route policy on Switch A to filter route 10.5.1.0/24.

Figure 1-29 Network diagram for OSPF route filtering configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF basic functions (Refer to Configuring OSPF Basic Functions).
3) Configure OSPF to redistribute routes.

On Switch C, configure a static route destined for network 3.1.1.0/24.

```
<SwitchC> system-view
[SwitchC] ip route-static 3.1.1.0 24 10.4.1.2
```

On Switch C, configure a static route destined for network 3.1.2.0/24.

```
[SwitchC] ip route-static 3.1.2.0 24 10.4.1.2
```

On Switch C, configure a static route destined for network 3.1.3.0/24.

```
[SwitchC] ip route-static 3.1.3.0 24 10.4.1.2
```

On Switch C, configure OSPF to redistribute static routes.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route static
[SwitchC-ospf-1] quit
```

Display the OSPF routing table of Switch A.

```
<SwitchA> display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE  150  1            10.2.1.2        Vlan200
3.1.2.0/24          O_ASE  150  1            10.2.1.2        Vlan200
3.1.3.0/24          O_ASE  150  1            10.2.1.2        Vlan200
10.1.1.0/24         Direct 0    0            10.1.1.1        Vlan200
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan200
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         OSPF   10   4            10.1.1.2        Vlan100
10.4.1.0/24         OSPF   10   13           10.2.1.2        Vlan200
10.5.1.0/24         OSPF   10   14           10.1.1.2        Vlan100
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

4) On Switch C, filter out route 3.1.3.0/24.

Configure the IPv4 prefix list.

```
[SwitchC] ip ip-prefix prefix1 index 1 deny 3.1.3.0 24
[SwitchC] ip ip-prefix prefix1 index 2 permit 3.1.1.0 24
[SwitchC] ip ip-prefix prefix1 index 3 permit 3.1.2.0 24
```

Reference the prefix list to filter out route 3.1.3.0/24.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] filter-policy ip-prefix prefix1 export static
```

Display the OSPF routing table of Switch A.

```
<SwitchA> display ip routing-table
Routing Tables: Public
         Destinations : 11       Routes : 11

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE  150  1            10.2.1.2        Vlan200
3.1.2.0/24          O_ASE  150  1            10.2.1.2        Vlan200
10.1.1.0/24         Direct 0    0            10.1.1.1        Vlan100
```

```
10.1.1.1/32         Direct 0    0           127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0           10.2.1.1        Vlan200
10.2.1.1/32         Direct 0    0           127.0.0.1       InLoop0
10.3.1.0/24         OSPF   10   4           10.1.1.2        Vlan100
10.4.1.0/24         OSPF   10   13          10.2.1.2        Vlan200
10.5.1.0/24         OSPF   10   14          10.1.1.2        Vlan100
127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
```

The route destined for network 3.1.3.0/24 is filtered out.

5) On Switch A, filter out the route 10.5.1.1/24.

Configure the ACL on Switch A.

```
<SwitchA> system-veiw
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule 0 deny source 10.5.1.0 0.0.0.255
[SwitchA-acl-basic-2000] rule 1 permit source any
[SwitchA-acl-basic-2000] quit
```

Use the ACL to filter route 10.5.1.0/24.

```
[SwitchA] ospf 1
[SwitchA-ospf-1] filter-policy 2000 import
[SwitchA-ospf-1] quit
```

Display the OSPF routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 10       Routes : 10

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

3.1.1.0/24          O_ASE  150  1            10.2.1.2        Vlan200
3.1.2.0/24          O_ASE  150  1            10.2.1.2        Vlan200
10.1.1.0/24         Direct 0    0            10.1.1.1        Vlan100
10.1.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.2.1.0/24         Direct 0    0            10.2.1.1        Vlan200
10.2.1.1/32         Direct 0    0            127.0.0.1       InLoop0
10.3.1.0/24         OSPF   10   4            10.1.1.2        Vlan100
10.4.1.0/24         OSPF   10   13           10.2.1.2        Vlan200
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The route destined for 10.5.1.1/24 is filtered out.

Troubleshooting OSPF Configuration

No OSPF Neighbor Relationship Established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower layer protocols work well, check OSPF parameters configured on interfaces. Two neighbors must have the same parameters, such as the area ID, network segment and mask (a P2P or virtual link may have different network segments and masks).

Processing steps

1) Display OSPF neighbor information using the **display ospf peer** command.
2) Display OSPF interface information using the **display ospf interface** command.
3) Ping the neighbor router's IP address to check connectivity.
4) Check OSPF timers. The dead interval on an interface must be at least four times the hello interval.
5) On an NBMA network, using the **peer ip-address** command to specify the neighbor manually is required.
6) On an NBMA or a broadcast network, at least one connected interface must have a DR priority higher than 0.

Incorrect Routing Information

Symptom

OSPF cannot find routes to other areas.

Analysis

The backbone area must maintain connectivity to all other areas. If a router connects to more than one area, at least one area must be connected to the backbone. The backbone cannot be configured as a Stub area.

In a Stub area, all routers cannot receive external routes, and all interfaces connected to the Stub area must belong to the Stub area.

Solution

1) Use the **display ospf peer** command to display neighbors.
2) Use the **display ospf interface** command to display OSPF interface information.
3) Use the **display ospf lsdb** command to display the Link State Database to check its integrity.
4) Display information about area configuration using the **display current-configuration configuration ospf** command. If more than two areas are configured, at least one area is connected to the backbone.
5) In a Stub area, all routers attached are configured with the **stub** command. In an NSSA area, all interface connected to which are configured with the **nssa** command.
6) If a virtual link is configured, use the **display ospf vlink** command to check the state of the virtual link.

Table of Contents

1 IS-IS Configuration

When configuring IS-IS, go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

IS-IS Overview

Intermediate System-to-Intermediate System (IS-IS) is a dynamic routing protocol designed by the International Organization for Standardization (ISO) to operate on the connectionless network protocol (CLNP).

The IS-IS routing protocol was modified and extended in RFC 1195 by the International Engineer Task Force (IETF) for application in both TCP/IP and OSI reference models, and the new one is called Integrated IS-IS or Dual IS-IS.

IS-IS is an Interior Gateway Protocol (IGP) used within an Autonomous System. It adopts the Shortest Path First (SPF) algorithm for route calculation.

Basic Concepts

IS-IS terminology

- Intermediate system (IS). An IS, similar to a router in TCP/IP, is the basic unit in IS-IS to generate and propagate routing information. In the following text, an IS refers to a router.
- End system (ES). An ES refers to a host system in TCP/IP. ISO defines the ES-IS protocol for communication between an ES and an IS, and therefore an ES does not participate in the IS-IS processing.

- Routing domain (RD). A group of ISs exchanges routing information with each other using the same routing protocol in a routing domain.
- Area. An area is a unit in a routing domain. The IS-IS protocol allows a routing domain to be divided into multiple areas.
- Link State Database (LSDB). All link states in the network forms the LSDB. There is at least one LSDB in each IS. The IS uses the SPF algorithm and LSDB to generate its own routes.
- Link State Protocol Data Unit (LSPDU) or Link State Packet (LSP). Each IS can generate an LSP which contains all the link state information of the IS.
- Network Protocol Data Unit (NPDU). An NPDU is a network layer protocol packet in OSI, which is equivalent to an IP packet in TCP/IP.
- Designated IS. On a broadcast network, the designated router is also known as the designated IS.
- Network service access point (NSAP). An NSAP is an OSI network layer address. It identifies an abstract network service access point and describes the network address in the OSI reference model.

IS-IS address format

1) NSAP

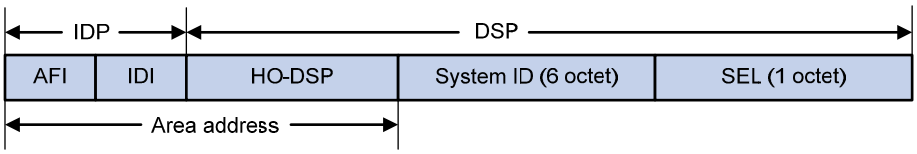
As shown in Figure 1-1, an NSAP address consists of the Initial Domain Part (IDP) and the Domain Specific Part (DSP). The IDP is equal to the network ID of an IP address, and the DSP is equal to the subnet and host ID.

The IDP includes the Authority and Format Identifier (AFI) and the Initial Domain Identifier (IDI).

The DSP includes the High Order Part of DSP (HO-DSP), System ID and SEL, where the HO-DSP identifies the area, the System ID identifies the host, and the SEL identifies the type of service.

The IDP and DSP are variable in length. The length of an NSAP address varies from 8 bytes to 20 bytes.

Figure 1-1 NSAP address format



2) Area address

The area address comprises the IDP and the HODSP of the DSP, which identify the area and the routing domain. Different routing domains cannot have the same area address.

Generally, a router only needs one area address, and all nodes in the same routing domain must share the same area address. However, a router can have three area addresses at most to support smooth area merging, partitioning and switching.

3) System ID

A system ID identifies a host or router uniquely. It has a fixed length of 48 bits (6 bytes).

The system ID of a device can be generated from the Router ID. For example, a router uses the IP address 168.10.1.1 of Loopback 0 as the Router ID, and the system ID in IS-IS can be obtained in the following way:

- Extend each decimal number of the IP address to 3 digits by adding 0s from the left, like 168.010.001.001;

- Divide the extended IP address into 3 sections with 4 digits in each section to get the system ID 1680.1000.1001.

There are other methods to define a system ID. The principle is to make sure it can uniquely identify a host or router.

4) SEL

The NSAP Selector (SEL), or the N-SEL, is similar to the protocol identifier in IP. Different transport layer protocols correspond to different SELs. All SELs in IP are 00.

NET

A network entity title (NET) indicates the network layer information of an IS and does not include transport layer information. It is a special NSAP address with the SEL being 0. Therefore, the length of the NET is equal to the NSAP and is in the range 8 bytes to 20 bytes.

Generally, a router only needs one NET, but it can have three NETs at most for smooth area merging and partitioning. When you configure multiple NETs, make sure their system IDs are the same.

For example, a NET is ab.cdef.1234.5678.9abc.00, where,

Area = ab.cdef, System ID = 1234.5678.9abc, and SEL = 00.

IS-IS Area

Two-level hierarchy

IS-IS has a two-level hierarchy to support large scale networks. A large scale routing domain is divided into multiple Areas. Typically, a Level-1 router is deployed within an area, a Level-2 router is deployed between areas, and a Level-1-2 router is deployed between Level-1 and Level-2 routers.

Level-1 and Level-2

1) Level-1 router

A Level-1 router establishes neighbor relationships with Level-1 and Level-1-2 routers in the same area. The LSDB maintained by the Level-1 router contains the local area routing information. It directs the packets destined for an outside area to the nearest Level-1-2 router.

2) Level-2 router

A Level-2 router establishes neighbor relationships with the Level-2 and Level-1-2 routers in the same or in different areas. It maintains a Level-2 LSDB which contains inter-area routing information. All the Level-2 and Level-1-2 routers must be contiguous to form the backbone of a routing domain.

3) Level-1-2 router

A router with both Level-1 and Level-2 router functions is a Level-1-2 router. It can establish Level-1 neighbor relationships with the Level-1 and Level-1-2 routers in the same area, or establish Level-2 neighbor relationships with the Level-2 and Level-1-2 routers in different areas. A Level-1 router must be connected to other areas through a Level-1-2 router. The Level-1-2 router maintains two LSDBs, where the Level-1 LSDB is for routing within the area, and the Level-2 LSDB is for routing between areas.

Note

- The Level-1 routers in different areas can not establish neighbor relationships.
- The neighbor relationship establishment of Level-2 routers has nothing to do with area.

Figure 1-2 shows an IS-IS network topology. Area 1 comprises a set of Level-2 routers and is the backbone. The other four areas are non-backbone areas connected to the backbone through Level-1-2 routers.

Figure 1-2 IS-IS topology

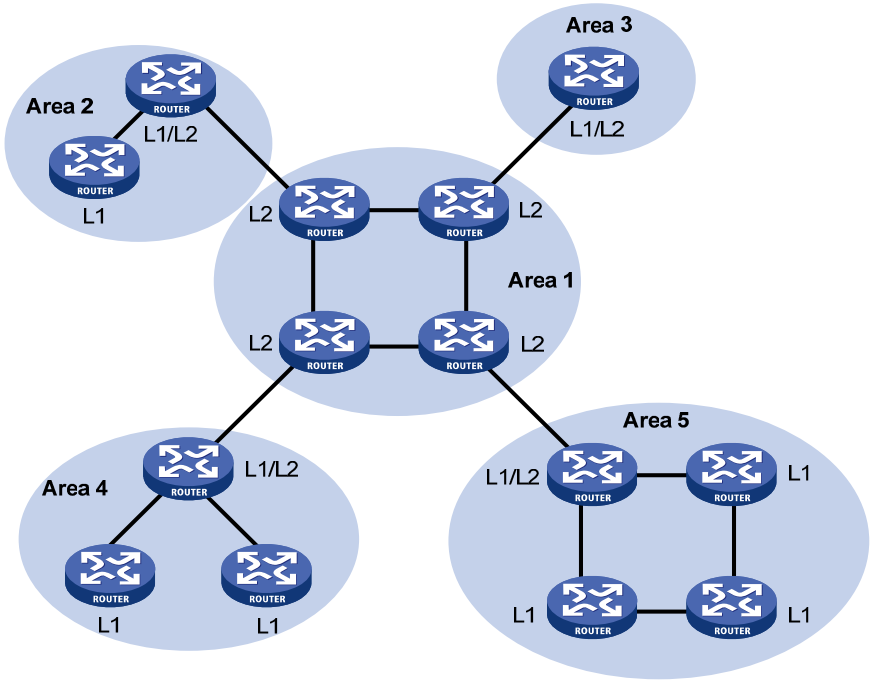
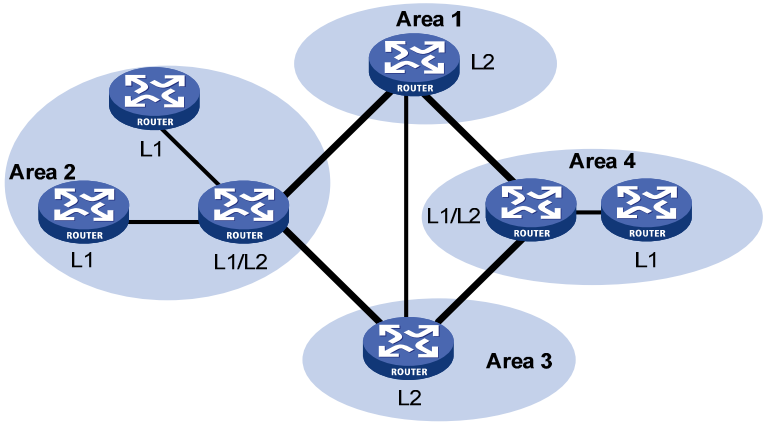


Figure 1-3 shows another IS-IS topology. The Level-1-2 routers connect to the Level-1 and Level-2 routers, and form the IS-IS backbone together with the Level-2 routers. There is no area defined as the backbone in this topology. The backbone comprises all contiguous Level-2 and Level-1-2 routers which can reside in different areas.

Figure 1-3 IS-IS topology



Note

The IS-IS backbone does not need to be a specific Area.

Both the IS-IS Level-1 and Level-2 routers use the SPF algorithm to generate the shortest path tree (SPT).

Routing method

A Level-1 router makes routing decisions based on the system ID. If the destination is not in the area, the packet is forwarded to the nearest Level-1-2 router.

A Level-2 router routes packets across areas according to the area address.

Route leaking

An IS-IS routing domain is comprised of only one Level-2 area and multiple Level-1 areas. A Level-1 area consists of a group of Level-1 routers and is connected with a Level-2 area rather than other Level-1 areas.

The routing information of a Level-1 area is sent to the Level-2 area through the Level-1-2 router. Therefore, the Level-2 router knows the routing information of the entire IS-IS routing domain but does not share the information of other Level-1 areas and the Level-2 area with the Level-1 area by default.

Since a Level-1 router simply sends packets destined for other areas to the nearest Level-1-2 router, this may cause that the best paths cannot be selected.

To solve this problem, route leaking was introduced. A Level-2 router can advertise Level-2 routing information to a specified Level-1 area. By having the routing information of other areas, a Level-1 router in the area can make a better routing decision for a packet to another area.

IS-IS Network Type

Network type

IS-IS supports two network types:

- Broadcast network, such as Ethernet, Token-Ring.
- Point-to-point network, such as PPP, HDLC.

Note

For a Non-Broadcast Multi-Access (NBMA) interface, such as an ATM interface, you need to configure subinterfaces for it and configure the interface type for the subinterfaces as point-to-point or broadcast. IS-IS cannot run on point to multipoint (P2MP) links.

DIS and pseudonodes

On an IS-IS broadcast network, a router is elected as the Designated Intermediate System (DIS).

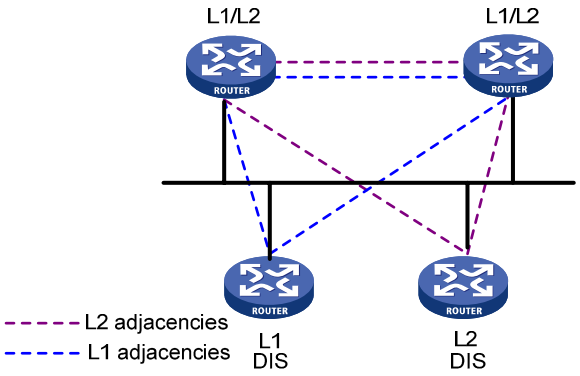
The Level-1 and Level-2 DISs are elected respectively. You can assign different priorities for different level DIS elections. The higher a router's priority is, the more likelihood the router becomes the DIS. If there are multiple routers with the same highest DIS priority, the one with the highest SNPA (Subnetwork Point of Attachment) address (MAC address on a broadcast network) will be elected. A router can be the DIS for different levels.

IS-IS DIS election differs from OSPF DIS election in that:

- A router with priority 0 can also participate in the DIS election.
- When a router is added to the network and becomes the new DIS, an LDP flooding is triggered.

As shown in Figure 1-4, the same level routers on a network including non-DIS routers establish adjacencies with each other.

Figure 1-4 DIS in the IS-IS broadcast network



The DIS creates and updates pseudonodes as well as generates their LSPs to describe all routers on the network.

A pseudonode represents a virtual node on the broadcast network. It is not a real router. In IS-IS, it is identified by the system ID of the DIS and a one-byte Circuit ID (a non zero value).

Using pseudonodes can reduce the resources consumed by SPF and simplify network topology.

Note

On IS-IS broadcast networks, all routers are adjacent with each other. However, the DIS is responsible for the synchronization of their LSDBs.

IS-IS PDU Format

PDU header format

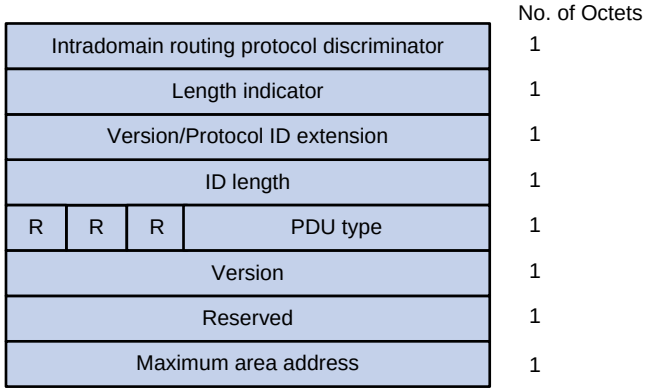
IS-IS packets are encapsulated into link layer frames. The Protocol Data Unit (PDU) consists of two parts, the headers and the variable length fields, where the headers comprise the PDU common header and the PDU specific header. All PDUs have the same PDU common header, while the specific headers vary by PDU type. The following figure shows the PDU format.

Figure 1-5 PDU format

Common header format

Figure 1-6 shows the PDU common header format.

Figure 1-6 PDU common header format



- Intradomain Routing Protocol Discriminator: Set to 0x83.
- Length Indicator: Length of the PDU header in bytes, including both common and specific headers.
- Version/Protocol ID Extension: Set to 1(0x01).
- ID Length: Length of the NSAP address and NET ID.
- R(Reserved): Set to 0.
- PDU Type: For details, refer to Table 1-1.
- Version: Set to 1(0x01).
- Maximum Area Address: Maximum number of area addresses supported.

Table 1-1 PDU type

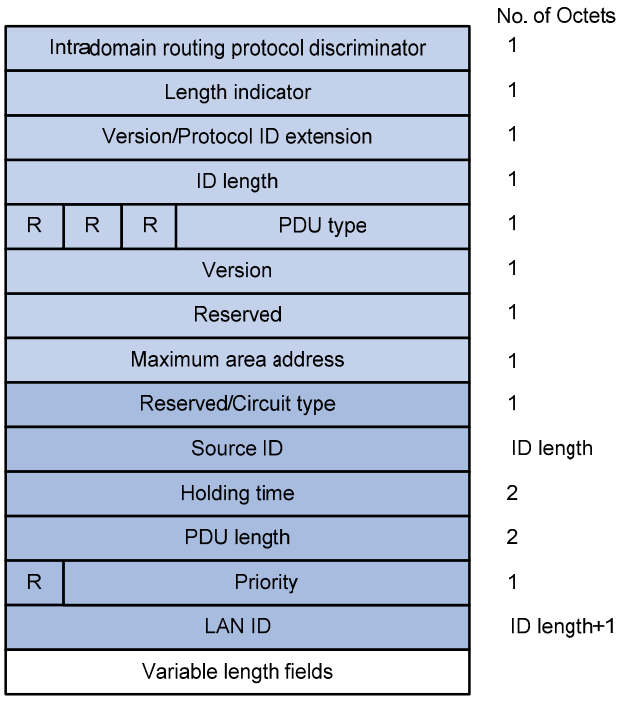
Type	PDU Type	Acronym
15	Level-1 LAN IS-IS hello PDU	L1 LAN IIH
16	Level-2 LAN IS-IS hello PDU	L2 LAN IIH
17	Point-to-Point IS-IS hello PDU	P2P IIH
18	Level-1 Link State PDU	L1 LSP
20	Level-2 Link State PDU	L2 LSP
24	Level-1 Complete Sequence Numbers PDU	L1 CSNP
25	Level-2 Complete Sequence Numbers PDU	L2 CSNP
26	Level-1 Partial Sequence Numbers PDU	L1 PSNP
27	Level-2 Partial Sequence Numbers PDU	L2 PSNP

Hello

Hello packets are used by routers to establish and maintain neighbor relationships. A hello packet is also called an IS-to-IS hello PDU (IIH). For broadcast networks, the Level-1 routers use the Level-1 LAN IIHs; and the Level-2 routers use the Level-2 LAN IIHs. The P2P IIHs are used on point-to-point networks.

Figure 1-7 illustrates the hello packet format in broadcast networks, where the blue fields are the common header.

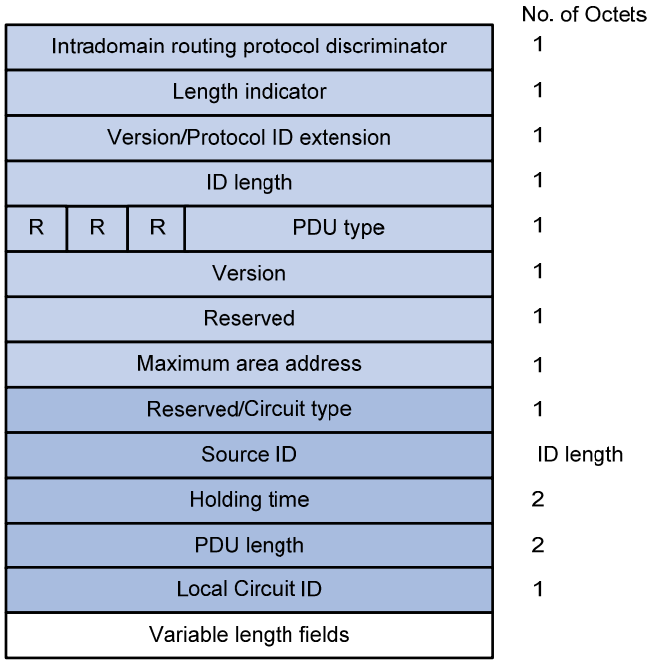
Figure 1-7 L1/L2 LAN IIH format



- Reserved/Circuit Type: The first 6 bits are reserved with a value of 0. The last 2 bits indicate the router type. 00 means reserved, 01 indicates L1, 10 indicates L2, and 11 indicates L1/2.
- Source ID: System ID of the router advertising the hello packet.
- Holding Time: If no hello packets are received from the neighbor within the holding time, the neighbor is considered down.
- PDU Length: Total length of the PDU in bytes.
- Priority: DIS priority.
- LAN ID: Includes the system ID and a one-byte pseudonode ID.

Figure 1-8 shows the hello packet format on the point-to-point networks.

Figure 1-8 P2P IIH format



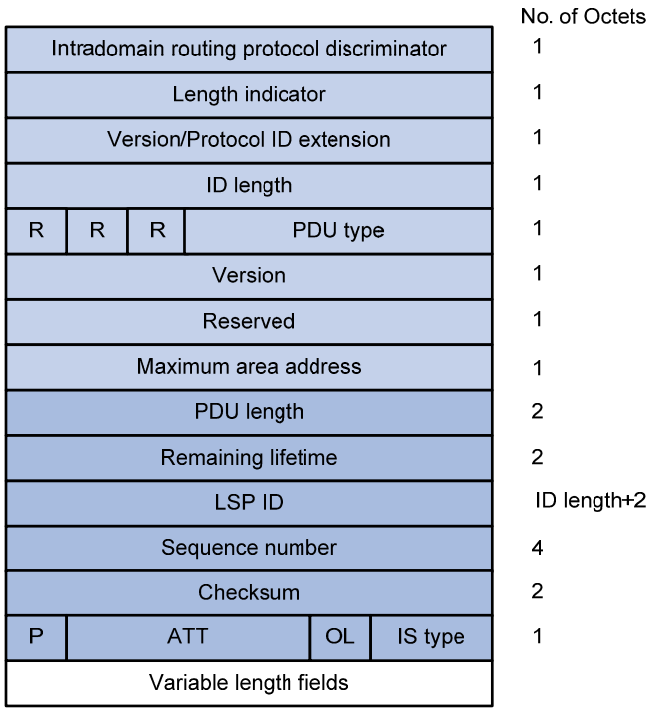
Instead of the priority and LAN ID fields in the LAN IIH, the P2P IIH has a Local Circuit ID field.

LSP packet format

The Link State PDUs (LSP) carry link state information. LSP involves two types: Level-1 LSP and Level-2 LSP. The Level-2 LSPs are sent by the Level-2 routers, and the Level-1 LSPs are sent by the Level-1 routers. The level-1-2 router can send both types of LSPs.

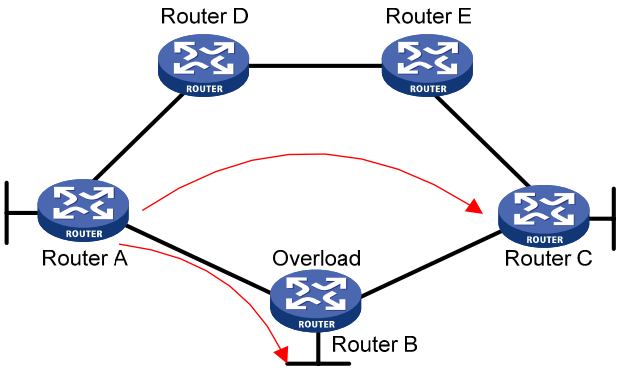
The two types of LSPs have the same format, as shown in Figure 1-9.

Figure 1-9 L1/L2 LSP format



- PDU Length: Total length of the PDU in bytes.
- Remaining Lifetime: LSP remaining lifetime in seconds.
- LSP ID: Consists of the system ID, the pseudonode ID (one byte) and the LSP fragment number (one byte).
- Sequence Number: LSP sequence number.
- Checksum: LSP checksum.
- P (Partition Repair): Only for L2 LSPs. It indicates whether the router supports partition repair.
- ATT (Attachment): Generated by a L1/L1 router for L1 LSPs only. It indicates that the router generating the LSP is connected to multiple areas.
- OL (LSDB Overload): Indicates that the LSDB is not complete because the router runs out of memory. In this case, other routers will not send packets to the overloaded router, except packets destined to the networks directly connected to the router. For example, in Figure 1-10, Router A forwards packets to Router C through Router B. Once other routers know the OL field of LSPs from Router B is set to 1, Router A will send packets to Router C via Router D and Router E, but still send to Router B packets destined to the network directly connected to Router B.

Figure 1-10 LSDB overload



- IS Type: Type of the router generating the LSP.

SNP format

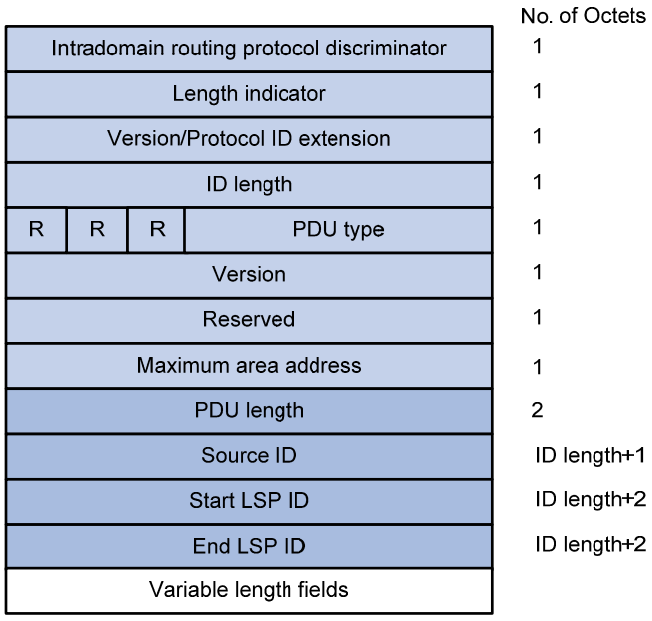
A sequence number PDU (SNP) acknowledges the latest received LSPs. It is similar to an Acknowledge packet, but more efficient.

SNP involves Complete SNP (CSNP) and Partial SNP (PSNP), which are further divided into Level-1 CSNP, Level-2 CSNP, Level-1 PSNP and Level-2 PSNP.

CSNP covers the summary of all LSPs in the LSDB to synchronize the LSDB between neighboring routers. On broadcast networks, CSNP is sent by the DIS periodically (10s by default). On point-to-point networks, CSNP is only sent during the first adjacency establishment.

The CSNP packet format is shown in Figure 1-11.

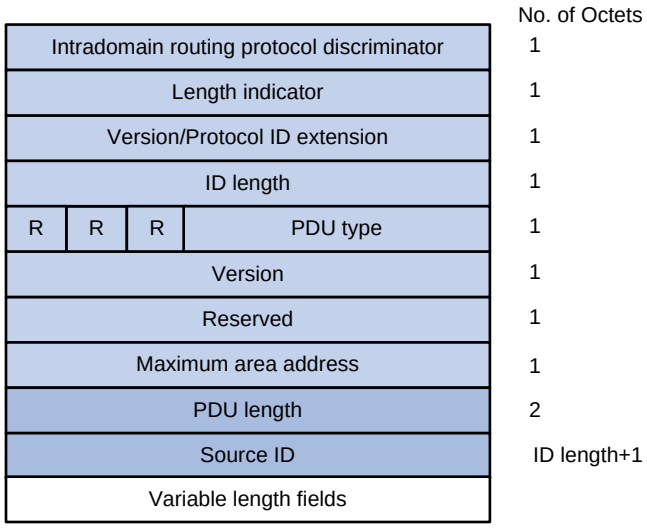
Figure 1-11 L1/L2 CSNP format



PSNP only contains the sequence numbers of one or multiple latest received LSPs. It can acknowledge multiple LSPs at one time. When LSDBs are not synchronized, a PSNP is used to request new LSPs from neighbors.

Figure 1-12 shows the PSNP packet format.

Figure 1-12 L1/L2 PSNP format



CLV

The variable fields of PDU comprise multiple Code-Length-Value (CLV) triplets. Figure 1-13 shows the CLV format.

Figure 1-13 CLV format

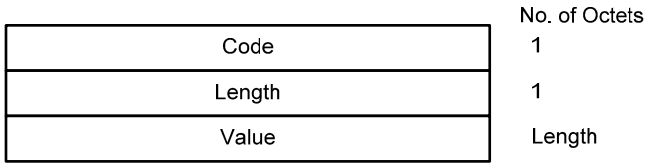


Table 1-2 shows that different PDUs contain different CLVs.

Table 1-2 CLV name and the corresponding PDU type

CLV Code	Name	PDU Type
1	Area Addresses	IIH, LSP
2	IS Neighbors (LSP)	LSP
4	Partition Designated Level2 IS	L2 LSP
6	IS Neighbors (MAC Address)	LAN IIH
7	IS Neighbors (SNPA Address)	LAN IIH
8	Padding	IIH
9	LSP Entries	SNP
10	Authentication Information	IIH, LSP, SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH, LSP
130	IP External Reachability Information	L2 LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH, LSP

Code 1 to 10 of CLV are defined in ISO 10589 (code 3 and 5 are not shown in the table), and others are defined in RFC 1195.

Supported IS-IS Features

Multiple instances and processes

IS-IS supports multiple instances and processes. Multiple processes allow a IS-IS process to work in concert with a group of interfaces. This means that a router can run multiple IS-IS processes, and each process corresponds to a unique group of interfaces.

For routers supporting VPN, each IS-IS process is associated with a VPN instance. Thus, the VPN instance is also associated with interfaces corresponding to the process.

IS-IS Graceful Restart

Note

For detailed GR information, refer to *GR Overview* in the *System Volume*.

After an IS-IS GR Restarter restarts IS-IS, it needs to complete the following two tasks to synchronize the LSDB with its neighbors.

- To obtain effective IS-IS neighbor information without changing adjacencies.
- To obtain the LSDB contents.

After the restart, the GR Restarter will send an OSPF GR signal to its neighbors to keep the adjacencies. After receiving the responses from neighbors, the GR Restarter can restore the neighbor table.

After reestablishing neighborships, the GR Restarter will synchronize the LSDB and exchange routing information with all adjacent GR capable neighbors. After that, the GR Restarter will update its own routing table and forwarding table based on the new routing information and remove the stale routes. In this way, the IS-IS routing convergence is complete.

Management tag

Management tag simplifies routing information management by carrying the management information of the IP address prefixes (to control route redistribution from other routing protocols) and BGP community and extended community attributes.

LSP fragment extension

IS-IS advertises link state information by flooding LSPs. One LSP carries a limited amount of link state information; therefore, IS-IS fragments LSPs. Each LSP fragment is uniquely identified by a combination of the System ID, Pseudonode ID (0 for a common LSP or a non-zero value for a Pseudonode LSP), and LSP Number (LSP fragment number) of the node or pseudo node that generated the LSP. The one-byte LSP Number field, allowing a maximum of only 256 fragments to be generated by an IS-IS router, limits the amount of link information that the IS-IS router can advertise.

The LSP fragment extension feature allows an IS-IS router to generate more LSP fragments. Up to 50 additional virtual systems can be configured on the router, and each virtual system is capable of generating 256 LSP fragments to enable the IS-IS router to generate up to 13056 LSP fragments.

1) Terms

- Originating System

It is the router actually running IS-IS. After LSP fragment extension is enabled, additional virtual systems can be configured for the router. Originating system is the actual IS-IS process that originally runs.

- System ID: System ID of the originating system.
- Additional System ID

Additional virtual system IDs are configured for the IS-IS router after LSP fragment extension is enabled. Each additional system ID can generate 256 LSP fragments. Both the additional system ID and the system ID must be unique in the entire routing domain.

- Virtual System

A virtual system is identified by an additional system ID and generates extended LSP fragments.

- Original LSP

It is the LSP generated by the originating system. The system ID in its LSP ID field is the system ID of the originating system.

- Extended LSP

Extended LSPs are generated by virtual systems. The system ID in its LSP ID field is the virtual system ID.

After additional system IDs are configured, an IS-IS router can advertise more link state information in extended LSP fragments. Each virtual system can be considered a virtual router. An extended LSP fragment is advertised by a virtual system identified by an additional system ID.

2) Operation modes

The LSP fragment extension feature operates in two modes:

- Mode-1: Applicable to a network where some routers do not support LSP fragment extension. In this mode, adjacencies are formed between the originating system and virtual systems, with the link cost from the originating system to each virtual system as 0. Thus, each virtual system acts as a router connected to the originating system in the network, but the virtual systems are reachable through the originating system only. Therefore, the IS-IS routers not supporting LSP fragment extension can operate normally without modifying the extended LSP fragments received, but some limitation is imposed on the link state information in the extended LSP fragments advertised by the virtual systems.
- Mode-2: Applicable to a network where all the routers support LSP fragment extension. In this mode, all the IS-IS routers know which virtual system belongs to which originating system; therefore, no limitation is imposed on the link state information of the extended LSP fragments advertised by the virtual systems.

The operation mode of LSP fragment extension is configured based on area and routing level. Mode-1 allows the routers supporting and not supporting LSP fragment extension to interoperate with each other, but it restricts the link state information in the extended fragments. Mode-2 does not restrict the link state information in the extended fragments, and is recommended for an area where all the routers are at the same routing level and support LSP fragment extension.

Dynamic host name mapping mechanism

The dynamic host name mapping mechanism provides the mappings between the host names and the system IDs for the IS-IS routers. The dynamic host name information is announced in the dynamic host name CLV of an LSP.

This mechanism also provides the mapping between a host name and the DIS of a broadcast network, which is announced in the dynamic host name TLV of a pseudonode LSP.

A host name is easier to remember than a system ID. After enabling this feature on the router, you can see the host names instead of system IDs using the **display** command.

Protocols and Standards

- ISO 10589 ISO IS-IS Routing Protocol
- ISO 9542 ES-IS Routing Protocol
- ISO 8348/Ad2 Network Services Access Points
- RFC 1195 - Use of OSI IS-IS for Routing in TCP/IP and Dual Environments

- RFC 2763 - Dynamic Hostname Exchange Mechanism for IS-IS
- RFC 2966 - Domain-wide Prefix Distribution with Two-Level IS-IS
- RFC 2973 - IS-IS Mesh Groups
- RFC 3277 - IS-IS Transient Blackhole Avoidance
- RFC 3358 - Optional Checksums in ISIS
- RFC 3373 - Three-Way Handshake for IS-IS Point-to-Point Adjacencies
- RFC 3567 - Intermediate System to Intermediate System (IS-IS) Cryptographic Authentication
- RFC 3719 - Recommendations for Interoperable Networks using IS-IS
- RFC 3786 - Extending the Number of IS-IS LSP Fragments Beyond the 256 Limit
- RFC 3787 - Recommendations for Interoperable IP Networks using IS-IS
- RFC 3847 - Restart signaling for IS-IS

IS-IS Configuration Task List

Complete the following tasks to configure IS-IS:

Task		Remarks
Configuring IS-IS Basic Functions	Enabling IS-IS	Required
	Configuring the IS Level and Circuit Level	
	Configuring the Network Type of an Interface as P2P	
Configuring IS-IS Routing Information Control	Configuring IS-IS Link Cost	Optional
	Specifying a Priority for IS-IS	Required
	Configuring the Maximum Number of Equal Cost Routes	Optional
	Configuring IS-IS Route Summarization	Optional
	Advertising a Default Route	Optional
	Configuring IS-IS Route Redistribution	Optional
	Configuring IS-IS Route Filtering	Optional
	Configuring IS-IS Route Leaking	Optional
Tuning and Optimizing IS-IS Networks	Specifying Intervals for Sending IS-IS Hello and CSNP Packets	Optional
	Specifying the IS-IS Hello Multiplier	Optional
	Configuring a DIS Priority for an Interface	Optional
	Disabling an Interface from Sending/Receiving IS-IS Packets	Optional
	Enabling an Interface to Send Small Hello Packets	Optional
	Configuring LSP Parameters	Optional
	Configuring SPF Parameters	Optional
	Setting the LSDB Overload Bit	Optional
Configuring IS-IS Authentication	Configuring Neighbor Relationship Authentication	Optional
	Configuring Area Authentication	Optional
	Configuring Routing Domain Authentication	Optional
Configuring System ID to Host Name Mappings	Configuring a Static System ID to Host Name Mapping	Optional
	Configuring Dynamic System ID to Host Name Mapping	Optional

Task	Remarks
Configuring IS-IS GR	Optional
Enabling the Logging of Neighbor State Changes	Optional
Enabling IS-IS SNMP Trap	Optional
Binding an IS-IS Process with MIBs	Optional

Configuring IS-IS Basic Functions

Configuration Prerequisites

Before the configuration, accomplish the following tasks:

- Configure the link layer protocol.
- Configure an IP address for each interface, and make sure all neighboring nodes are reachable to each other at the network layer.

Enabling IS-IS

Follow these steps to enable IS-IS:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the IS-IS routing process and enter its view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	Required Not enabled by default
Assign a network entity title (NET)	**network-entity** *net*	Required Not assigned by default
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable an IS-IS process on the interface	**isis enable** [*process-id*]	Required Disabled by default

Configuring the IS Level and Circuit Level

If only one area is available, it is recommended that:

- Configure the IS level of all routers as Level-1 or Level-2 and don't configure different levels in this case because there is no need for all routers to maintain two identical LSDBs;
- Configure the IS level as Level-2 on all routers in an IP network for scalability.

For an interface of a Level-1 (or Level-2) router, the circuit level can only be Level-1 (or Level-2). For an interface of a Level-1-2 router, the default circuit level is Level-1-2; if the router only needs to form Level-1 (or Level-2) neighbor relationships, you can configure the circuit level for its interfaces as Level-1 (or Level-2) to limit neighbor relationship establishment.

Follow these steps to configure the IS level and circuit level:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the IS level	**is-level** { **level-1** | **level-1-2** | **level-2** }	Optional The default is Level-1-2.
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the circuit level	**isis circuit-level** [**level-1** | **level-1-2** | **level-2**]	Optional The default is Level-1-2.

Configuring the Network Type of an Interface as P2P

Interfaces with different network types operate differently. For example, broadcast interfaces on a network need to elect the DIS and flood CSNP packets to synchronize the LSDBs, while P2P interfaces on a network need not elect the DIS and have a different LSDP synchronization mechanism.

If there are only two routers on a broadcast network, you can configure the network type of attached interfaces as P2P to avoid DIS election and CSNP flooding, saving network bandwidth and speeding up network convergence.

Follow these steps to configure the network type of an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the network type for the interface as P2P	**isis circuit-type p2p**	Optional By default, The network type of a VLAN interface is broadcast.

Note

You can only perform this configuration for a broadcast network with only two attached routers.

Configuring IS-IS Routing Information Control

Configuration Prerequisites

Before the configuration, accomplish the following tasks:

- Configure network layer addresses for interfaces, and make sure adjacent nodes are reachable to each other at the network layer.
- Enable IS-IS.

Configuring IS-IS Link Cost

The IS-IS cost of an interface is determined in the following order:

- ISIS cost specified in interface view.
- ISIS cost specified in system view. The cost is applied to the interfaces associated to the IS-IS process.
- Automatically calculated cost: When the cost style is **wide** or **wide-compatible**, IS-IS automatically calculates the cost using the formula: interface cost= (bandwidth reference value/interface bandwidth) ×10. When the cost style is some other type: if the interface bandwidth does not exceed 10 Mbps, the interface cost equals 60; if the interface bandwidth does not exceed 100 Mbps, the interface cost equals 50; if the interface bandwidth does not exceed 155 Mbps, the interface cost equals 40; if the interface bandwidth does not exceed 622 Mbps, the interface cost equals 30; if the interface bandwidth does not exceed 2500 Mbps, the interface cost equals 20; if the interface bandwidth exceeds 2500 Mbps, the interface cost equals 10.
- If none of the above costs is used, a default cost of 10 applies.

Configuring an IS-IS cost for an interface

Follow these steps to configure a cost for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify an IS-IS cost style	**cost-style** { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }	Optional **narrow** by default
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify a cost for the interface	**isis cost** *value* [**level-1** | **level-2**]	Optional No cost is specified for the interface by default.

Configuring a global IS-IS cost

Follow these steps to configure a global IS-IS cost:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify an IS-IS cost style	**cost-style** { **narrow** | **wide** | **wide-compatible** | { **compatible** | **narrow-compatible** } [**relax-spf-limit**] }	Optional **narrow** by default
Specify a global IS-IS cost	**circuit-cost** *value* [**level-1** | **level-2**]	Required No global cost is specified by default.

Enable automatic IS-IS cost calculation

Follow these steps to enable automatic IS-IS cost calculation:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify an IS-IS cost style	**cost-style** { **wide** | **wide-compatible** }	Required **narrow** by default
Enable automatic IS-IS cost calculation	**auto-cost enable**	Required Disabled by default
Configure a bandwidth reference value for automatic IS-IS cost calculation	**bandwidth-reference** *value*	Optional 100 Mbps by default

Specifying a Priority for IS-IS

A router may run multiple routing protocols. When routes to the same destination are found by multiple routing protocols, the route learned by the protocol with the highest priority wins. You can reference a routing policy to specify a priority for specific routes. For information about routing policy, refer to *Routing Policy Configuration* in the *IP Routing Volume*.

Follow these steps to configure the priority of IS-IS.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify a priority for IS-IS	**preference** { **route-policy** *route-policy-name* | *preference* } *	Required 15 by default

Configuring the Maximum Number of Equal Cost Routes

If there are multiple equal cost routes to the same destination, the traffic can be load balanced to enhance efficiency.

Follow these steps to configure the maximum number of equal cost routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the maximum number of equal cost routes for load balancing	**maximum load-balancing** *number*	Required 4 by default.

Configuring IS-IS Route Summarization

This task is to configure a summary route, so routes falling into the network range of the summary route are summarized into one route for advertisement. Doing so can reduce the size of routing tables, as well as the scale of LSP and LSDB. Both IS-IS routes and redistributed routes can be summarized.

Follow these steps to configure route summarization:

To do…	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure IS-IS route summarization	**summary** *ip-address* { *mask* | *mask-length* } [**avoid-feedback** | **generate_null0_route** | **tag** *tag* | [**level-1** | **level-1-2** | **level-2**]] *	Required No route summarization is configured by default.

Note

- The cost of the summary route is the lowest one among the costs of summarized routes.
- The router summarizes only the routes in the locally generated LSPs.

Advertising a Default Route

A router running IS-IS cannot redistribute any default route and thus cannot advertise a default route to other neighbors. You can use the following commands to advertise a default route of 0.0.0.0/0 to the same level neighbors.

Follow these steps to advertise a default route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Advertise a default route	**default-route-advertise** [**route-policy** *route-policy-name* | [**level-1** | **level-2** | **level-1-2**]] *	Required The function is disabled by default.

Note

The default route is only advertised to routers at the same level. You can use a routing policy to generate the default route only when a local routing entry is matched by the policy.

Configuring IS-IS Route Redistribution

Redistribution of large numbers of routes on a device may affect the performance of other devices in the network. In that case, you can configure a limit on the number of redistributed routes to limit the number of routes to be advertised.

Follow these steps to configure IS-IS route redistribution from other routing protocols:

To do…	Use the command…	Remarks
Enter system view	**system-view**	––
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	––
Redistribute routes from another routing protocol	**import-route** *protocol* [*process-id* | **all-processes** | **allow-ibgp**] [**cost** *cost* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	Required No route is redistributed by default. If no level is specified, routes are redistributed into the Level-2 routing table by default.
Configure the maximum number of redistributed Level 1/Level 2 IPv4 routes	**import-route limit** *number*	Optional By default the maximum number of redistributed Level 1/Level 2 IPv4 routes is 12288

Note

Only active routes can be redistributed. You can use the **display ip routing-table protocol** command to display route state information.

Configuring IS-IS Route Filtering

You can reference a configured ACL, IP prefix list or routing policy to filter routes calculated from the received LSPs and the routes redistributed from other routing protocols.

Filtering routes calculated from received LSPs

IS-IS saves the LSPs received from neighbors in the LSDB, uses the SPF algorithm to calculate the shortest path tree with itself as the root and installs the routes into the IS-IS routing table.

By reference a configured ACL, IP prefix list or routing policy, you can filter the calculated routes and only the routes matching the filter can be added into the IS-IS routing table.

Follow these steps to filter routes calculated from received LSPs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—

To do…	Use the command…	Remarks
Filter routes calculated from received LSPs	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **import**	Required No filtering is configured by default.

Filtering redistributed routes

IS-IS can redistribute routes from other routing protocols or other IS-IS processes, add them into the IS-IS routing table and advertise them in LSPs.

By reference a configured ACL, IP prefix list or routing policy, you can filter redistributed routes and only the routes matching the filter can be added into the IS-IS routing table and advertised to neighbors.

Follow these steps to configure the filtering of redistributed routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure the filtering of routes redistributed from another routing protocol or IS-IS process	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]	Required Not configured by default

Configuring IS-IS Route Leaking

With IS-IS route leaking enabled, the Level-1-2 router can advertise the routing information of other Level-1 areas and Level-2 area routing information to Level-1 routers.

Follow these steps to configure IS-IS route leaking:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable IS-IS route leaking	**import-route isis level-2 into level-1** [**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } | **tag** *tag*] *	Required Disabled by default

Note

- If a filter policy is specified, only routes passing it can be advertised into Level-1 area.
- You can specify a routing policy in the **import-route isis level-2 into level-1** command to filter routes from Level-2 to Level-1. Other routing policies specified for route reception and redistribution does not affect the route leaking.

Tuning and Optimizing IS-IS Networks

Configuration Prerequisites

Before the configuration, accomplish the following tasks:

- Configure IP addresses for interfaces, and make adjacent nodes reachable to each other at the network layer.
- Enable IS-IS.

Specifying Intervals for Sending IS-IS Hello and CSNP Packets

Follow these steps to configure intervals for sending IS-IS hello and CSNP packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the interval for sending hello packets	**isis timer hello** *seconds* [**level-1** | **level-2**]	Optional 10 seconds by default
Specify the interval for sending CSNP packets on the DIS of a broadcast network	**isis timer csnp** *seconds* [**level-1** | **level-2**]	Optional 10 seconds by default

Note

The interval between hello packets sent by the DIS is 1/3 the hello interval set with the **isis timer hello** command.

Specifying the IS-IS Hello Multiplier

If a neighbor receives no hello packets from the router within the advertised hold time, it considers the router down and recalculates the routes. The hold time is the hello multiplier times the hello interval.

Follow these steps to specify the IS-IS hello multiplier:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the number of hello packets a neighbor must miss before declaring the router is down	**isis timer holding-multiplier** *value* [**level-1** | **level-2**]	Optional 3 by default

Note

On a broadcast link, Level-1 and Level-2 hello packets are advertised separately and therefore you need to set a hello multiplier for each level. On a P2P link, Level-1 and Level-2 hello packets are advertised in P2P hello packets, and you need not specify Level-1 or Level-2.

Configuring a DIS Priority for an Interface

On an IS-IS broadcast network, a router should be elected as the DIS at a routing level. You can specify a DIS priority at a level for an interface. The greater the interface's priority is, the more likely it becomes the DIS. If multiple routers in the broadcast network have the same highest DIS priority, the router with the highest MAC address becomes the DIS.

Follow these steps to specify a DIS priority for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify a DIS priority for the interface	**isis dis-priority** *value* [**level-1** | **level-2**]	Optional 64 by default

Disabling an Interface from Sending/Receiving IS-IS Packets

After disabled from sending and receiving hello packets, an interface cannot form any neighbor relationship, but can advertise directly connected networks in LSPs through other interfaces. By doing so, you can save bandwidth and CPU resources while ensuring other routers know networks directly connected to the interfaces.

Follow these steps to disable an interface from sending and receiving IS-IS packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Disable the interface from sending and receiving IS-IS packets	**isis silent**	Required Not disabled by default

Enabling an Interface to Send Small Hello Packets

IS-IS messages cannot be fragmented at the IP layer because they are directly encapsulated into frames. Therefore, any two IS-IS neighboring routers need to negotiate a common MTU. To avoid sending big hellos for saving bandwidth, you can enable the interface to send small hello packets without CLVs.

Follow these steps to enable an interface to send small hello packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the interface to send small hello packets without CLVs	**isis small-hello**	Required Standard hello packets are sent by default.

Configuring LSP Parameters

Configuring LSP timers

1) Specify the maximum age of LSPs

Each LSP has an age that decreases in the LSDB. Any LSP with an age of 0 is deleted from the LSDB. You can adjust the age value based on the scale of a network.

Follow these steps to specify the maximum age of LSPs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the maximum LSP age	**timer lsp-max-age** *seconds*	Optional 1200 seconds by default

2) Specify the LSP refresh interval and generation interval

Each router needs to refresh LSPs generated by itself at a configurable interval and send them to other routers to prevent valid routes from being aged out. A smaller refresh interval speeds up network convergence but consumes more bandwidth.

When the network topology changes, for example, a neighbor is down/up, or the interface metric, system ID or area ID is changed, the router generates an LSP after a configurable interval. If such changes occur frequently, excessive LSPs are generated, consuming a large amount of router resources and bandwidth; in this case, you can adjust the LSP generation interval.

Follow these steps to specify the LSP refresh interval and generation interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the LSP refresh interval	**timer lsp-refresh** *seconds*	Optional 900 seconds by default
Specify the LSP generation interval	**timer lsp-generation** *maximum-interval* [*initial-interval* [*second-wait-interval*]] [**level-1** | **level-2**]	Optional 2 seconds by default

3) Specify LSP sending intervals

If a change occurs in the LSDB, IS-IS advertises the changed LSP to neighbors. You can specify the minimum interval for sending such LSPs.

On a P2P link, IS-IS requires an advertised LSP be acknowledged. If no acknowledgement is received within a configurable interval, IS-IS will retransmit the LSP.

Follow these steps to configure LSP sending intervals:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the minimum interval for sending LSPs and the maximum LSP number that can be sent at a time	**isis timer lsp** *time* [**count** *count*]	Optional By default, the minimum interval is 33 milliseconds, and the maximum LSP number that can be sent at a time is 5.
Specify the LSP retransmission interval on a P2P link	**isis timer retransmit** *seconds*	Optional 5 seconds by default

Note

Configure a proper LSP retransmission interval to avoid unnecessary retransmissions.

Specifying LSP lengths

IS-IS messages cannot be fragmented at the IP layer because they are directly encapsulated in frames. Therefore, IS-IS routers in an area need to send LSPs smaller than the smallest interface MTU in this area.

If the IS-IS routers have different interface MTUs, it is recommended to configure the maximum size of generated LSP packets to be smaller than the smallest interface MTU in this area. Otherwise, the routers have to dynamically adjust the LSP packet size to fit the smallest interface MTU, which takes time and affects other services.

Follow these steps to specify LSP lengths:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the maximum length of generated Level-1 LSPs or Level-2 LSPs	**lsp-length originate** *size* [**level-1** | **level-2**]	1497 bytes by default
Specify the maximum length of received LSPs	**lsp-length receive** *size*	1497 bytes by default

Enabling LSP flash flooding

Since changed LSPs may trigger SPF recalculation, you can enable LSP flash flooding to advertise the changed LSPs before the router recalculates routes. Doing so can speed up network convergence.

Follow these steps to enable LSP flash flooding:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable LSP flash flooding	**flash-flood** [**flood-count** *flooding-count* | **max-timer-interval** *flooding-interval* | [**level-1 | level-2**]] *	Required Not enabled by default

Enabling LSP fragment extension

Follow these steps to enable LSP fragment extension:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable LSP fragment extension and specify the working mode	**lsp-fragments-extend** [[**level-1** | **level-2** | **level-1-2**] | [**mode-1** | **mode-2**]] *	Required Not enabled by default
Configure a virtual system ID	**virtual-system** *virtual-system-id*	Required Not configured by default

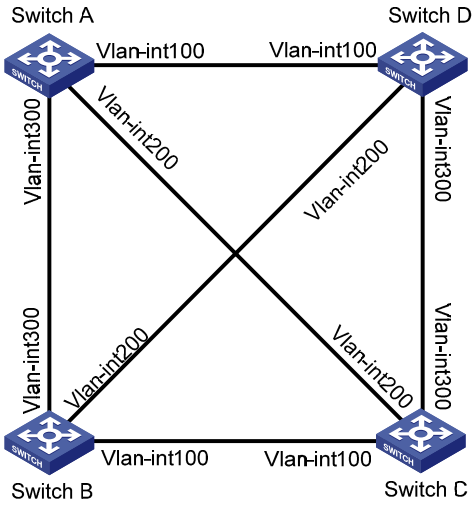
Note

- After LSP fragment extension is enabled for an IS-IS process, the MTUs of all the interfaces running the IS-IS process must not be less than 512; otherwise, LSP fragment extension will not take effect.
- At least one virtual system needs to be configured for the router to generate extended LSP fragments. An IS-IS process allows 50 virtual systems at most.

Limiting LSP flooding

In well connected NBMA networks, many P2P links exist. The following figure shows a fully meshed network, where Switchs A, B, C and D run IS-IS. When Switch A generates an LSP, it floods the LSP out VLAN-interface 100, VLAN-interface 200 and VLAN-interface 300. After receiving the LSP from VLAN-interface 100, Switch D floods it out VLAN-interface 200 and VLAN-interface 300 to Switch B and Switch C, which however has received the LSP from Router A. In this case, LSP flooding consumes extra bandwidth.

Figure 1-14 Network diagram of a fully meshed network



To avoid this, you can configure some interfaces as a mesh group or/and configure the blocked interfaces.

- After receiving an LSP, a member interface in a mesh group floods it out the interfaces that does not belong to the mesh group.
- If an interface is blocked, it does not send LSPs unless the neighbor sends LSP requests to it.

Before configuring this task, you need to consider redundancy for interfaces to avoid the fact that LSP packets cannot be flooded due to link failures.

Follow these steps to add an interface into a mesh group and block an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Add the interface to a mesh group	**isis mesh-group** *mesh-group-number*	Required to choose either. By default, the interface neither belongs to any mesh group nor is blocked.
Block the interface	**isis mesh-group mesh-blocked**	

Note

The mesh group feature takes effect only on P2P interfaces.

Configuring SPF Parameters

When the LSDB changes on a router, a route calculation starts. Frequent route calculations consume a lot of system resources, while route calculations at a proper interval improve efficiency. You can set an appropriate interval for SPF calculations as needed.

Follow these steps to configure the SPF parameters:

To do…	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure the SPF calculation interval	**timer spf** *maximum-interval* [*initial-interval* [*second-wait-interval*]]	Optional The default SPF calculation interval is 10 seconds.

Setting the LSDB Overload Bit

By setting the overload bit in sent LSPs, a router informs other routers of a failure that makes it incapable of routing and forwarding packets.

When an IS-IS router cannot record the complete LSDP due to running out of memory or some other reasons, it will calculate wrong routes. To make troubleshooting easier in this case, you can temporarily isolate the router from the IS-IS network by setting the overload bit.

Follow these steps to set the LSDB overload bit:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Set the overload bit	**set-overload** [**on-startup** [[**start-from-nbr** *system-id* [*timeout1* [*nbr-timeout*]]] | *timeout2*] [**allow** { **interlevel** | **external** } *]	Required Not set by default

Configuring IS-IS Authentication

Configuration Prerequisites

Complete the following tasks before this configuration:

- Configure network layer addresses for interfaces to make neighboring nodes accessible to each other at the network layer.
- Enable IS-IS.

Configuring Neighbor Relationship Authentication

With neighbor relationship authentication configured, an interface adds the password in the specified mode into hello packets to the peer and checks the password in the received hello packets. If the authentication succeeds, it forms the neighbor relationship with the peer.

The authentication mode and password at both ends must be identical.

Follow these steps to configure neighbor relationship authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify the authentication mode and password	**isis authentication-mode** { **simple** | **md5** } *password* [**level-1** | **level-2**] [**ip** | **osi**]	Required Not authentication is configured by default.

Note

The **level-1** and **level-2** keywords in the **isis authentication-mode** command are only supported on VLAN interfaces of switches, and the interfaces must be configured with the **isis enable** command first.

Configuring Area Authentication

Area authentication enables a router not to install routing information from untrusted routers into the Level-1 LSDB. The router encapsulates the authentication password in the specified mode into Level-1 packets (LSP, CSNP, PSNP) and check the password in received Level-1 packets.

Routers in a common area must have the same authentication mode and password.

Follow these steps to configure area authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify the area authentication mode and password	**area-authentication-mode** { **simple** | **md5** } *password* [**ip** | **osi**]	Required No area authentication is configured by default.

Configuring Routing Domain Authentication

Routing domain authentication prevents untrusted routing information from entering into a routing domain. A router with the authentication configured encapsulates the password in the specified mode into Level-2 packets (LSP, CSNP, PSNP) and check the password in received Level-2 packets.

All the routers in the backbone must have the same authentication mode and password.

Follow these steps to configure routing domain authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—

To do…	Use the command…	Remarks
Specify the routing domain authentication mode and password	**domain-authentication-mode** { **simple** | **md5** } *password* [**ip** | **osi**]	Required No routing domain authentication is configured by default.

Configuring System ID to Host Name Mappings

In IS-IS, a system ID identifies a router or host uniquely. A system ID has a fixed length of 6 bytes. When an administrator needs to view IS-IS neighbor information, routing table or LSDB information, using the system IDs in dotted decimal notation is not convenient. To solve it, you can configure the mappings between system IDs and host names since host names are easier to remember and use.

Such mappings can be configured manually or dynamically. Note that:

- Using the **display isis lsdb** command on a router configured with dynamic system ID to host name mapping displays router names rather than system IDs.
- If you configure both dynamic and static system ID to host name mappings on a router, the host name for dynamic system ID to host name mapping applies.

Configuring a Static System ID to Host Name Mapping

Follow these steps to configure a static system ID to host name mapping:

To do…	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Configure a system ID to host name mapping for a remote IS	**is-name map** *sys-id map-sys-name*	Required A system ID can only correspond to a host name.

Configuring Dynamic System ID to Host Name Mapping

You need to configure a static system ID to host name mapping for any other router in a network. When a new router is added into the network or a mapping needs to be modified, you need to perform configuration on all routers.

In this case, you can configure dynamic system ID to host name mapping. To do so, you need to configure a host name for each router in the network. Each router advertises the host name in dynamic host name CLVs to other routers. At last, all routers in the network have all the mappings to generate a mapping table.

In addition, you can configure a name for the DIS in a broadcast network to help check the origin of LSPs in the LSDB.

Follow these steps to configure dynamic system ID to host name mapping:

To do…	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Specify a host name for the router	**is-name** *sys-name*	Required No specified by default.
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a DIS name	**isis dis-name** *symbolic-name*	Optional Not configured by default. This command takes effect only on a router with dynamic system ID to host name mapping configured. This command is not supported on P2P interfaces.

Configuring IS-IS GR

Restarting ISIS on a router will cause network disconnections and route re-convergence.

With the Graceful Restart (GR) feature, the restarting router, known as the GR restarter, can notify the event to its GR capable neighbors, which, known as the GR helpers, will keep their adjacencies with the router within a configurable GR interval. After the restart, the router contacts its neighbors to retrieve its routing table.

During the whole process, the network keeps stable.

You can enable the GR Restarter to suppress the Suppress-Advertisement (SA) bit in the hello PDUs. In this way, its neighbors will still advertise the adjacencies within the specified period.

Follow these steps to configure GR on the GR Restarter and GR Helper respectively:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable IS-IS, and enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	Required Disabled by default
Enable the GR capability for IS-IS	**graceful-restart**	Required Disabled by default
Set the Graceful Restart interval	**graceful-restart interval** *timer*	Required 300 seconds by default
Suppress the SA bit during restart	**graceful-restart suppress-sa**	Optional By default, the SA bit is not suppressed.

Enabling the Logging of Neighbor State Changes

Follow these steps to enable the logging of neighbor state changes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable the logging of neighbor state changes	**log-peer-change**	Required Enabled by default

Note

With this feature enabled, the router delivers information about neighbor state changes to the terminal for display.

Enabling IS-IS SNMP Trap

Follow these steps to enable IS-IS SNMP trap:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Enable SNMP trap	**is-snmp-traps enable**	Required Enabled by default

Binding an IS-IS Process with MIBs

Follow these steps to bind an IS-IS process with MIBs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Bind the IS-IS process with MIBs	**isis mib-binding** *process-id*	Required By default, MIBs are bound with IS-IS process 1.

Displaying and Maintaining IS-IS

To do…	Use the command…	Remarks
Display brief IS-IS configuration information	**display isis brief** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display the status of IS-IS debug switches	**display isis debug-switches** { *process-id* | **vpn-instance** *vpn-instance-name* }	Available in any view
Display the IS-IS Graceful Restart state	**display isis graceful-restart status** [**level-1** | **level-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display information about IS-IS enabled interfaces	**display isis interface** [**statistics** | [*interface-type interface-number*] [**verbose**]] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS license information	**display isis license**	Available in any view
Display IS-IS LSDB information	**display isis lsdb** [[**l1** | **l2** | **level-1** | **level-2**] | [**lsp-id** *lspid* | **lsp-name** *lspname*] | **local** | **verbose**] * [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS mesh group information	**display isis mesh-group** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display the host-name-to-system-ID mapping table	**display isis name-table** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS neighbor information	**display isis peer** [**verbose | statistics**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS IPv4 routing information	**display isis route** [**ipv4**] [[**level-1** | **level-2**] | **verbose**] * [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS SPF calculation log information	**display isis spf-log** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS statistics	**display isis statistics** [**level-1** | **level-2** | **level-1-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Clear ISIS process data structure information	**reset isis all** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in user view
Clear the data structure information of an IS-IS neighbor	**reset isis peer** *system-id* [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in user view

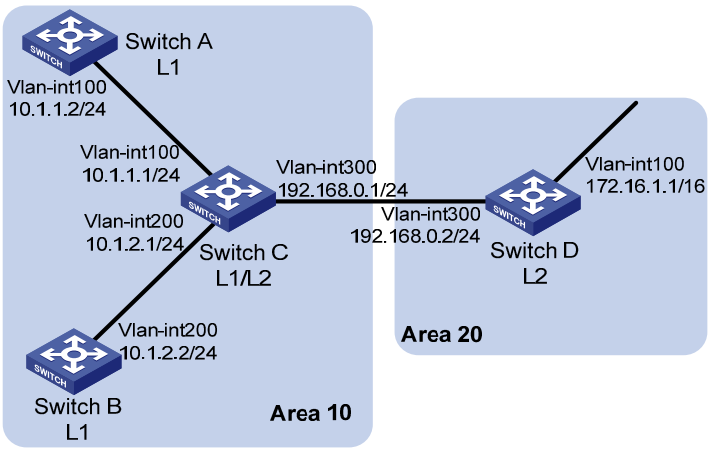
IS-IS Configuration Example

IS-IS Basic Configuration

Network requirements

As shown in Figure 1-15, Switch A, B, C and Switch D reside in an IS-IS AS. Switch A and B are Level-1 switches, Switch D is a Level-2 switch and Switch C is a Level-1-2 switch. Switch A, B and C are in Area 10, while Switch D is in Area 20.

Figure 1-15 Network diagram for IS-IS basic configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
```

```
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
```

3) Verify the configuration

Display the IS-IS LSDB of each switch to check the LSP integrity.

```
[SwitchA] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum   Holdtime   Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0001.00-00* 0x00000004   0xdf5e     1096       68      0/0/0
0000.0000.0002.00-00  0x00000004   0xee4d     1102       68      0/0/0
0000.0000.0002.01-00  0x00000001   0xdaaf     1102       55      0/0/0
0000.0000.0003.00-00  0x00000009   0xcaa3     1161       111     1/0/0
0000.0000.0003.01-00  0x00000001   0xadda     1112       55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[SwitchB] display isis lsdb

                        Database information for ISIS(1)
```

```
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0001.00-00  0x00000006   0xdb60     988         68      0/0/0
0000.0000.0002.00-00* 0x00000008   0xe651     1189        68      0/0/0
0000.0000.0002.01-00* 0x00000005   0xd2b3     1188        55      0/0/0
0000.0000.0003.00-00  0x00000014   0x194a     1190        111     1/0/0
0000.0000.0003.01-00  0x00000002   0xabdb     995         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[SwitchC] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------

                          Level-1 Link State Database

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0001.00-00  0x00000006   0xdb60     847         68      0/0/0
0000.0000.0002.00-00  0x00000008   0xe651     1053        68      0/0/0
0000.0000.0002.01-00  0x00000005   0xd2b3     1052        55      0/0/0
0000.0000.0003.00-00* 0x00000014   0x194a     1051        111     1/0/0
0000.0000.0003.01-00* 0x00000002   0xabdb     854         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload


                          Level-2 Link State Database

LSPID                 Seq Num      Checksum   Holdtime    Length  ATT/P/OL
-------------------------------------------------------------------------
0000.0000.0003.00-00* 0x00000012   0xc93c     842         100     0/0/0
0000.0000.0004.00-00  0x00000026   0x331      1173        84      0/0/0
0000.0000.0004.01-00  0x00000001   0xee95     668         55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[SwitchD] display isis lsdb

                        Database information for ISIS(1)
                        --------------------------------
```

```
                        Level-2 Link State Database

LSPID                 Seq Num      Checksum      Holdtime      Length  ATT/P/OL
-------------------------------------------------------------------------------
0000.0000.0003.00-00  0x00000013   0xc73d        1003          100     0/0/0
0000.0000.0004.00-00* 0x0000003c   0xd647        1194          84      0/0/0
0000.0000.0004.01-00* 0x00000002   0xec96        1007          55      0/0/0

    *-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload
```

Display the IS-IS routing information of each switch. Level-1 switches should have a default route with the next hop being the Level-1-2 switch. The Level-2 switch should have both routing information of Level-1 and Level-2.

```
[SwitchA] display isis route

                        Route information for ISIS(1)
                        ------------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop      Flags
-------------------------------------------------------------------------------
 10.1.1.0/24          10        NULL    Vlan100         Direct       D/L/-
 10.1.2.0/24          20        NULL    Vlan100         10.1.1.1     R/-/-
 192.168.0.0/24       20        NULL    Vlan100         10.1.1.1     R/-/-
 0.0.0.0/0            10        NULL    Vlan100         10.1.1.1     R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[SwitchC] display isis route

                        Route information for ISIS(1)
                        ------------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost    ExtCost ExitInterface  NextHop      Flags
-------------------------------------------------------------------------------
 192.168.0.0/24       10         NULL    Vlan300        Direct       D/L/-
 10.1.1.0/24          10         NULL    Vlan100        Direct       D/L/-
 10.1.2.0/24          10         NULL    Vlan200        Direct       D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

```
                    ISIS(1) IPv4 Level-2 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop      Flags
-------------------------------------------------------------------------
 192.168.0.0/24       10        NULL    Vlan300         Direct       D/L/-
 10.1.1.0/24          10        NULL    Vlan100         Direct       D/L/-
 10.1.2.0/24          10        NULL    Vlan200         Direct       D/L/-
 172.16.0.0/16        20        NULL    Vlan300         192.168.0.2  R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[SwitchD] display isis route

                        Route information for ISIS(1)
                        -----------------------------

                    ISIS(1) IPv4 Level-2 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop      Flags
-------------------------------------------------------------------------
 192.168.0.0/24       10        NULL    Vlan300         Direct       D/L/-
 10.1.1.0/24          20        NULL    Vlan300         192.168.0.1  R/-/-
 10.1.2.0/24          20        NULL    Vlan300         192.168.0.1  R/-/-
 172.16.0.0/16        10        NULL    Vlan100         Direct       D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

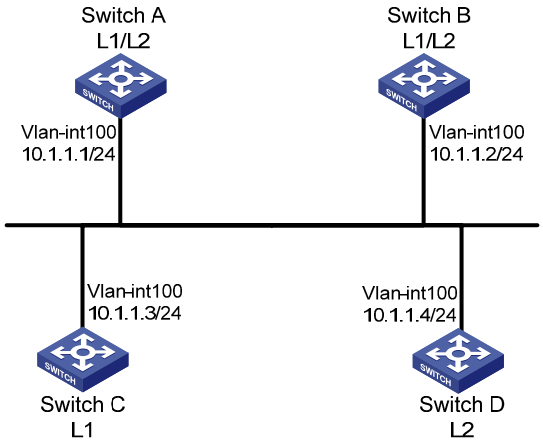
DIS Election Configuration

Network requirements

As shown in Figure 1-16, Switch A, B, C and Switch D reside in IS-IS area 10 on a broadcast network (Ethernet). Switch A and Switch B are Level-1-2 switches, Switch C is a Level-1 switch, and Switch D is a Level-2 switch.

Change the DIS priority of Switch A to make it elected as the Level-1-2 DIS router.

Figure 1-16 Network diagram for DIS selection



Configuration procedure

1) Configure an IP address for each interface (omitted)
2) Enable IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] isis enable 1
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] is-level level-1
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
```

```
[SwitchD-isis-1] network-entity 10.0000.0000.0004.00
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
```

Display information about IS-IS neighbors of Switch A.

```
[SwitchA] display isis peer

                        Peer information for ISIS(1)
                        ----------------------------

  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0003.01
  State: Up     HoldTime: 21s        Type: L1(L1L2)     PRI: 64

  System Id: 0000.0000.0003
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0003.01
  State: Up     HoldTime: 27s        Type: L1           PRI: 64

  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0004.01
  State: Up     HoldTime: 28s        Type: L2(L1L2)     PRI: 64

  System Id: 0000.0000.0004
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0004.01
  State: Up     HoldTime: 30s        Type: L2          PRI: 64
```

Display information about IS-IS interfaces of Switch A.

```
[SwitchA] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: Vlan-interface100
Id      IPV4.State      IPV6.State    MTU    Type    DIS
001        Up             Down       1497    L1/L2   No/No
```

Display information about IS-IS interfaces of Switch C.

```
[SwitchC] display isis interface

                     Interface information for ISIS(1)
                     ---------------------------------
Interface: Vlan-interface100
Id      IPV4.State      IPV6.State    MTU    Type    DIS
001         Up            Down        1497   L1/L2   Yes/No
```

Display information about IS-IS interfaces of Switch D.

```
[SwitchD] display isis interface
```

```
                    Interface information for ISIS(1)
                    ---------------------------------
Interface: Vlan-interface100
Id      IPV4.State       IPV6.State   MTU     Type     DIS
001         Up             Down       1497    L1/L2    No/Yes
```

Note

By using the default DIS priority, Switch C is the Level-1 DIS, and Switch D is the Level-2 DIS. The pseudonodes of Level-1 and Level-2 are 0000.0000.0003.01 and 0000.0000.0004.01 respectively.

---

3) Configure the DIS priority of Switch A.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis dis-priority 100
[SwitchA-Vlan-interface100] quit
```

Display IS-IS neighbors of Switch A.

```
[SwitchA] display isis peer

                        Peer information for ISIS(1)
                        ----------------------------
  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 21s        Type: L1(L1L2)     PRI: 64

  System Id: 0000.0000.0003
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 27s        Type: L1           PRI: 64


  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 28s        Type: L2(L1L2)     PRI: 64

  System Id: 0000.0000.0004
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 30s        Type: L2           PRI: 64
```

Display information about IS-IS interfaces of Switch A.

```
[SwitchA] display isis interface

                    Interface information for ISIS(1)
                    ---------------------------------
Interface: Vlan-interface100
Id      IPV4.State      IPV6.State      MTU    Type   DIS
```

```
001         Up             Down         1497   L1/L2  Yes/Yes
```

Note

After the DIS priority configuration, Switch A becomes the Level-1-2 DIS, and the pseudonode is 0000.0000.0001.01.

Display information about IS-IS neighbors and interfaces of Switch C.

```
[SwitchC] display isis peer

                         Peer information for ISIS(1)
                         ----------------------------
  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 25s        Type: L1           PRI: 64

  System Id: 0000.0000.0001
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 7s         Type: L1           PRI: 100

[SwitchC] display isis interface

                      Interface information for ISIS(1)
                      ---------------------------------
Interface: Vlan-interface100
Id      IPV4.State          IPV6.State    MTU    Type   DIS
001         Up                 Down       1497   L1/L2  No/No
```

Display information about IS-IS neighbors and interfaces of Switch D.

```
[SwitchD] display isis peer

                         Peer information for ISIS(1)
                         ----------------------------
  System Id: 0000.0000.0001
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 9s         Type: L2           PRI: 100

  System Id: 0000.0000.0002
  Interface: Vlan-interface100       Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 28s        Type: L2           PRI: 64

[SwitchD] display isis interface

                      Interface information for ISIS(1)
                      ---------------------------------
Interface: Vlan-interface100
```

```
Id      IPV4.State          IPV6.State    MTU   Type    DIS
001         Up                  Down      1497  L1/L2   No/No
```

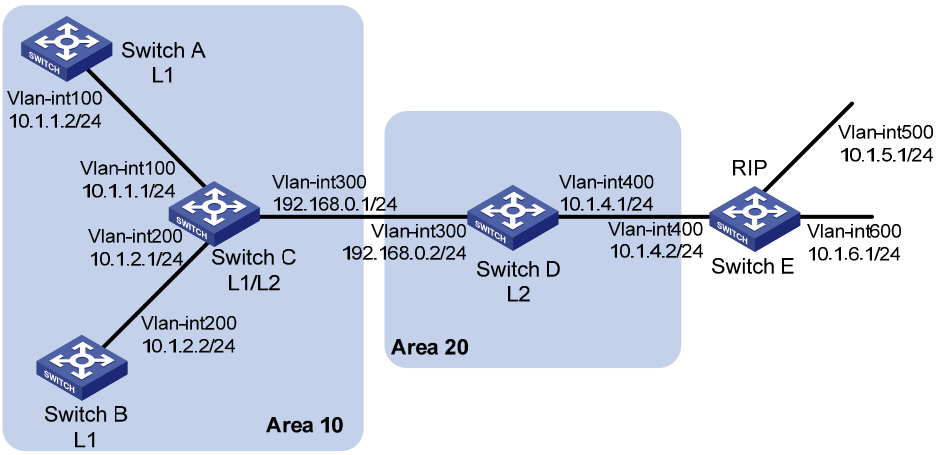
Configuring IS-IS Route Redistribution

Network requirements

As shown in the following figure, Switch A, Switch B, Switch C and Switch D reside in the same AS. They use IS-IS to interconnect. Switch A and Switch B are Level-1 routers, Switch D is a Level-2 router, and Switch C is a Level-1-2 router.

It is required to redistribute RIP routes into IS-IS on Switch D.

Figure 1-17 IS-IS route redistribution



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure IS-IS basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] quit
[SwitchD] interface interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
```

Display IS-IS routing information on each switch.

```
[SwitchA] display isis route

                        Route information for ISIS(1)
                        ----------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost  ExtCost ExitInterface  NextHop       Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10         NULL     VLAN100           Direct        D/L/-
 10.1.2.0/24          20         NULL     VLAN100           10.1.1.1      R/-/-
 192.168.0.0/24       20         NULL     VLAN100           10.1.1.1      R/-/-
 0.0.0.0/0            10          NULL     VLAN100           10.1.1.1      R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[SwitchC] display isis route

                        Route information for ISIS(1)
                        ----------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
```

```
                    ------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop     Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10           NULL     VLAN100            Direct      D/L/-
 10.1.2.0/24          10           NULL     VLAN200            Direct      D/L/-
 192.168.0.0/24       10           NULL     VLAN300            Direct      D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set


                    ISIS(1) IPv4 Level-2 Forwarding Table
                    ------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
-------------------------------------------------------------------------
 10.1.1.0/24          10          NULL     VLAN100           Direct       D/L/-
 10.1.2.0/24          10          NULL     VLAN200           Direct       D/L/-
 192.168.0.0/24       10          NULL     VLAN300           Direct       D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[SwitchD] display isis route

                        Route information for ISIS(1)
                        -----------------------------

                    ISIS(1) IPv4 Level-2 Forwarding Table
                    ------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
-------------------------------------------------------------------------
 192.168.0.0/24       10          NULL     VLAN300           Direct       D/L/-
 10.1.1.0/24           20          NULL     VLAN300           192.168.0.1  R/-/-
 10.1.2.0/24           20          NULL     VLAN300           192.168.0.1  R/-/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

3) Configure RIPv2 on Switch D and Switch E, and configure route redistribution from RIP to IS-IS on Switch D.

Configure RIPv2 on Switch D.

```
[SwitchD] rip 1
[SwitchD-rip-1] network 10.0.0.0
[SwitchD-rip-1] version 2
[SwitchD-rip-1] undo summary
```

Configure RIPv2 on Switch E.

```
[SwitchE] rip 1
[SwitchE-rip-1] network 10.0.0.0
```

```
[SwitchE-rip-1] version 2
[SwitchE-rip-1] undo summary
```

Configure route redistribution from RIP to IS-IS on Switch D.

```
[SwitchD-rip-1] quit
[SwitchD] isis 1
[SwitchD–isis] import-route rip level-2
```

Display IS-IS routing information on Switch C.

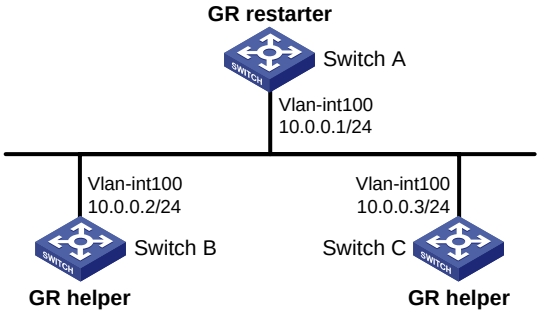
```
[SwitchC] display isis route

                        Route information for ISIS(1)
                        -----------------------------

                    ISIS(1) IPv4 Level-1 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface   NextHop     Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10         NULL    VLAN100            Direct      D/L/-
 10.1.2.0/24          10         NULL    VLAN200            Direct      D/L/-
 192.168.0.0/24       10         NULL    VLAN300            Direct      D/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set


                    ISIS(1) IPv4 Level-2 Forwarding Table
                    -------------------------------------

 IPV4 Destination     IntCost   ExtCost ExitInterface  NextHop      Flags
--------------------------------------------------------------------------------
 10.1.1.0/24          10          NULL    VLAN100           Direct       D/L/-
 10.1.2.0/24          10         NULL    VLAN200           Direct       D/L/-
 192.168.0.0/24       10         NULL    VLAN300           Direct       D/L/-
 10.1.4.0/24          10         NULL    VLAN300           192.168.0.2  R/L/-
 10.1.5.0/24          20         NULL    VLAN300           192.168.0.2  R/L/-
 10.1.6.0/24          20         NULL    VLAN300           192.168.0.2  R/L/-

      Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set
```

IS-IS-based Graceful Restart Configuration Example

Network requirements

Switch A, Switch B, and Switch C belong to the same IS-IS routing domain, as illustrated in Figure 1-18.

Figure 1-18 Network diagram for IS-IS-based GR configuration



Configuration procedure

1) Configure IP addresses of the interfaces on each switch and configure IS-IS.

Follow Figure 1-18 to configure the IP address and subnet mask of each interface. The configuration procedure is omitted.

Configure IS-IS on the switches, ensuring that Switch A, Switch B and Switch C can communicate with each other at layer 3 and dynamic route update can be implemented among them with IS-IS. The configuration procedure is omitted here.

2) Configure IS-IS Graceful Restart.

Enable IS-IS Graceful Restart on Switch A and configure the Graceful Restart Interval.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] graceful-restart
[SwitchA-isis-1] graceful-restart interval 150
[SwitchA-isis-1] return
```

Configurations for Switch B and Switch C are similar and therefore are omitted here.

3) Verify the configuration.

After Router A establishes adjacencies with Router B and Router C, they begin to exchange routing information. Restart IS-IS on Router A, which enters into the restart state and sends connection requests to its neighbors through the Graceful Restart mechanism to synchronize the LSDB. Using the **display isis graceful-restart status** command can display the IS-IS GR status on Router A.

Restart the IS-IS process on Switch A.

```
<SwitchA> reset isis all 1
Warning : Reset ISIS process? [Y/N]:y
```

Check the Graceful Restart status of IS-IS on Switch A.

```
<SwitchA> display isis graceful-restart status
                Restart information for IS-IS(1)
--------------------------------------------------------------------
IS-IS(1) Level-1 Restart Status
Restart Interval: 150
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTARTING
  Number of LSPs Awaited: 3
```

```
  T3 Timer Status:
    Remaining Time: 140
  T2 Timer Status:
    Remaining Time: 59

IS-IS(1) Level-2 Restart Status
Restart Interval: 150
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTARTING
  Number of LSPs Awaited: 3
  T3 Timer Status:
    Remaining Time: 140
  T2 Timer Status:
Remaining Time: 59
```

IS-IS Authentication Configuration Example

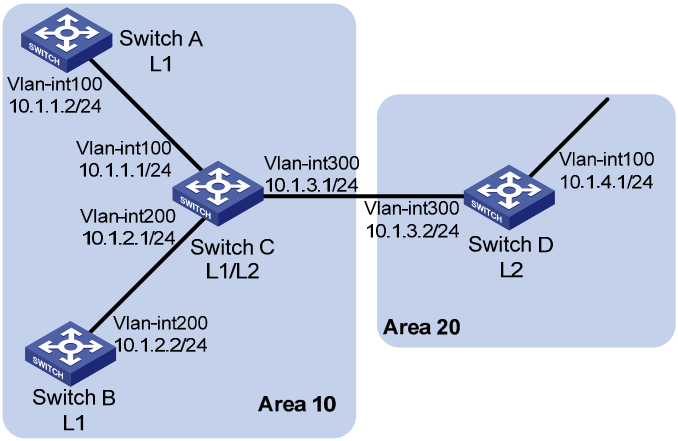
Network requirements

As shown in the following figure, Switch A, Switch B, Switch C and Switch D reside in the same IS-IS routing domain.

Switch A, Switch B, and Switch C belong to Area 10, and Switch D belongs to Area 20.

Configure neighbor relationship authentication between neighbors. Configure area authentication in Area 10 to prevent untrusted routes from entering into the area. Configure routing domain authentication on Switch C and Switch D to prevent untrusted routes from entering the routing domain.

Figure 1-19 IS-IS authentication configuration



Configuration procedure

1) Configure IP addresses for interfaces (Omitted).
2) Configure IS-IS basic functions.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
```

```
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[RouterB--Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] network-entity 20.0000.0000.0001.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit
```

3) Configure neighbor relationship authentication between neighbors.

Specify the MD5 authentication mode and password **eRq** on VLAN-interface 100 of Switch A and on VLAN-interface 100 of Switch C.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis authentication-mode md5 eRg
[SwitchA-Vlan-interface100] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis authentication-mode md5 eRg
[SwitchC-Vlan-interface100] quit
```

Specify the MD5 authentication mode and password **t5Hr** on VLAN-interface 200 of Switch B and on VLAN-interface 200 of Switch C.

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis authentication-mode md5 t5Hr
[SwitchB-Vlan-interface200] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis authentication-mode md5 t5Hr
[SwitchC-Vlan-interface200] quit
```

Specify the MD5 authentication mode and password **hSec** on VLAN-interface 300 of Switch D and on VLAN-interface 300 of Switch C.

```
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis authentication-mode md5 hSec
[SwitchC-Vlan-interface300] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis authentication-mode md5 hSec
[SwitchD-Vlan-interface300] quit
```

4) Configure area authentication. Specify the MD5 authentication mode and password **10Sec** on Switch A, Switch B and Switch C.

```
[SwitchA] isis 1
[SwitchA-isis-1] area-authentication-mode md5 10Sec
[SwitchA-isis-1] quit
[SwitchB] isis 1
[SwitchB-isis-1] area-authentication-mode md5 10Sec
[SwitchB-isis-1] quit
[SwitchC] isis 1
[SwitchC-isis-1] area-authentication-mode md5 10Sec
[SwitchC-isis-1] quit
```

5) Configure routing domain authentication. Specify the MD5 authentication mode and password **1020Sec** on Switch C and Switch D.

```
[SwitchC] isis 1
[SwitchC-isis-1] domain-authentication-mode md5 1020Sec
[SwitchC-isis-1] quit
[SwitchD] isis 1
[SwitchD-isis-1] domain-authentication-mode md5 1020Sec
```

Table of Contents

1 BGP Configuration

The Border Gateway Protocol (BGP) is a dynamic inter-AS Exterior Gateway Protocol.

When configuring BGP, go to these sections for information you are interested in:

Note

The term “router” refers to a router or a Layer 3 switch, and BGP refers to BGP-4 in this document.

BGP Overview

There are three early BGP versions, BGP-1 (RFC1105), BGP-2 (RFC1163) and BGP-3 (RFC1267). The current version in use is BGP-4 (RFC 4271), which is the defacto Internet exterior gateway protocol used between ISPs.

The characteristics of BGP are as follows:

- Focusing on the control of route propagation and the selection of optimal routes rather than the route discovery and calculation, which makes BGP, an exterior gateway protocol different from interior gateway protocols such as OSPF and RIP
- Using TCP to enhance reliability
- Supporting CIDR
- Reducing bandwidth consumption by advertising only incremental updates and therefore applicable to advertising a great amount of routing information on the Internet
- Eliminating routing loops completely by adding AS path information to BGP routes
- Providing abundant policies to implement flexible route filtering and selection
- Good scalability

A router advertising BGP messages is called a BGP speaker. It establishes peer relationships with other BGP speakers to exchange routing information. When a BGP speaker receives a new route or a route better than the current one from another AS, it will advertise the route to all the other BGP peers in the local AS.

To simplify configuration, multiple peers having an identical policy can be organized as a peer group.

BGP runs on a router in either of the following two modes:

- iBGP (internal BGP)
- eBGP (external BGP)

BGP is called iBGP when it runs within an AS and is called eBGP when it runs between ASs.

Formats of BGP Messages

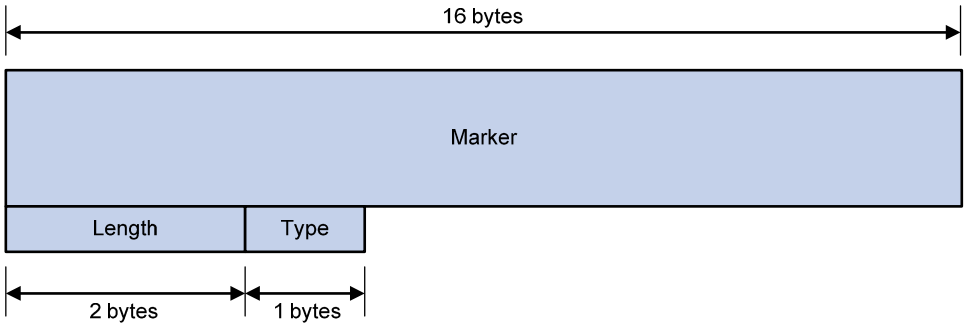
Header

BGP has five types of messages:

- Open
- Update
- Notification
- Keep-alive
- Route-refresh

They have the same header, as shown below:

Figure 1-1 BGP message header

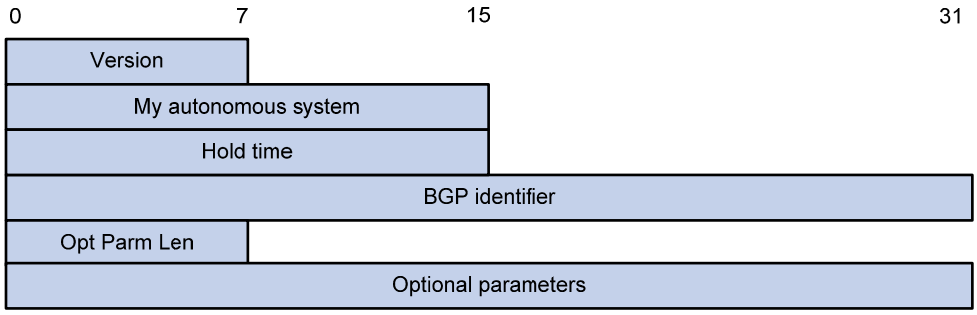


- Marker: The 16-byte field is used to delimit BGP messages. The Marker must be all ones.
- Length: The 2-byte unsigned integer indicates the total length of the message.
- Type: This 1-byte unsigned integer indicates the type code of the message. The following type codes are defined: 1–Open, 2-Update, 3-Notification, 4–Keepalive, and 5–Route-refresh. The former four are defined in RFC1771, and the last one is defined in RFC2918.

Open

After a TCP connection is established, the first message sent by each side is an Open message for peer relationship establishment. An Open message contains the following fields:

Figure 1-2 BGP open message format

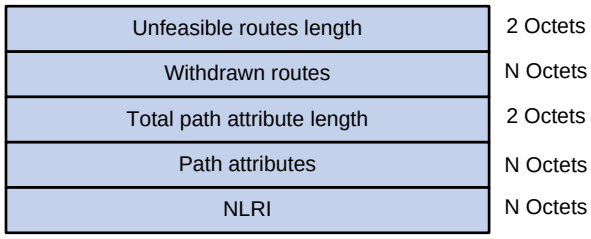


- Version: This 1-byte unsigned integer indicates the protocol version number. The current BGP version is 4.
- My autonomous system: This 2-byte unsigned integer indicates the Autonomous System number of the sender.
- Hold time: When establishing a peer relationship, two parties negotiate an identical hold time. If no Keepalive or Update is received from a peer within the hold time, the BGP connection is considered down.
- BGP identifier: An IP address that identifies the BGP router
- Opt Parm Len (Optional Parameters Length): Length of optional parameters, which is set to 0 if no optional parameter is available.
- Optional parameters: Used for multiprotocol extensions, and other functions.

Update

The Update messages are used to exchange routing information between peers. It can advertise a feasible route or remove multiple unfeasible routes. Its format is shown below:

Figure 1-3 BGP Update message format



Each update message can advertise a group of feasible routes with identical attributes, and the routes are contained in the network layer reachable information (NLRI) field. The Path Attributes field carries attributes of these routes. Each update message can also carry multiple withdrawn routes in the Withdrawn Routes field.

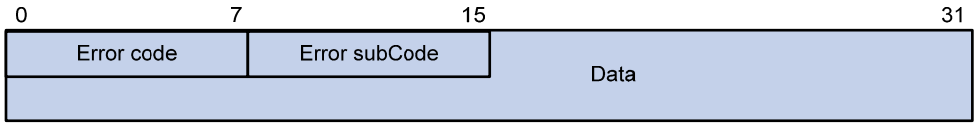
- Unfeasible routes length: The total length of the Withdrawn Routes field in bytes. A value of 0 indicates no route is withdrawn from service, and the Withdrawn Routes field is not present in this Update message.
- Withdrawn routes: This is a variable length field that contains a list of withdrawn IP prefixes.
- Total path attribute length: Total length of the Path Attributes field in bytes. A value of 0 indicates that no Network Layer Reachability Information field is present in this Update message.
- Path attributes: List of path attributes related to NLRI. Each path attribute is a triple <attribute type, attribute length, attribute value> of variable length. BGP uses these attributes to avoid routing loops, and perform routing and protocol extensions.

- NLRI (Network Layer Reachability Information): Each feasible route is represented as <length, prefix>.

Notification

A Notification message is sent when an error is detected. The BGP connection is closed immediately after sending it. The Notification message format is shown below:

Figure 1-4 BGP Notification message format



- Error code: Type of Notification.
- Error subcode: Specific information about the nature of the reported error.
- Data: Used to diagnose the reason for the Notification. The contents of the Data field depend upon the Error Code and Error Subcode. Erroneous part of data is recorded. The Data field length is variable.

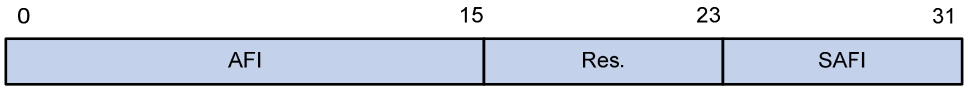
Keepalive

Keepalive messages are sent between peers to maintain connectivity. Its format contains only the message header.

Route-refresh

A route-refresh message is sent to a peer to request the resending of the specified address family routing information. Its format is shown below:

Figure 1-5 BGP Route-refresh message format



- AFI: Address family identifier.
- Res: Reserved. Set to 0.
- SAFI: Subsequent Address Family Identifier.

BGP Path Attributes

Classification of path attributes

Path attributes fall into four categories:

- Well-known mandatory: Must be recognized by all BGP routers and must be included in every update message. Routing information errors occur without this attribute.
- Well-known discretionary: Can be recognized by all BGP routers and optional to be included in every update message as needed.
- Optional transitive: Transitive attribute between ASs. A BGP router not supporting this attribute can still receive routes with this attribute and advertise them to other peers.

- Optional non-transitive: If a BGP router does not support this attribute, it will not advertise routes with this attribute.

The usage of each BGP path attribute is described in the following table.

Table 1-1 Usage of BGP path attributes

Name	Category
ORIGIN	Well-known mandatory
AS_PATH	Well-known mandatory
NEXT_HOP	Well-known mandatory
LOCAL_PREF	Well-known discretionary
ATOMIC_AGGREGATE	Well-known discretionary
AGGREGATOR	Optional transitive
COMMUNITY	Optional transitive
MULTI_EXIT_DISC (MED)	Optional non-transitive
ORIGINATOR_ID	Optional non-transitive
CLUSTER_LIST	Optional non-transitive

Usage of BGP path attributes

1) ORIGIN

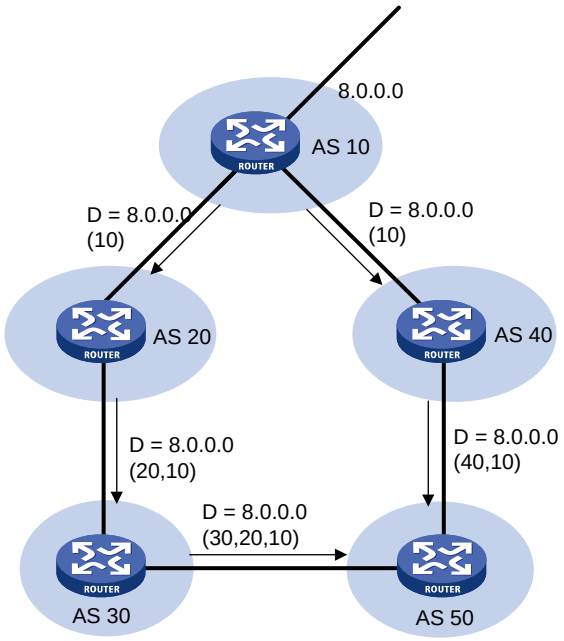
ORIGIN is a well-known mandatory attribute, which defines the origin of routing information, that is, how a route became a BGP route. It involves three types:

- IGP: Has the highest priority. Routes added to the BGP routing table using the **network** command have the IGP attribute.
- EGP: Has the second highest priority. Routes obtained via EGP have the EGP attribute.
- incomplete: Has the lowest priority. The source of routes with this attribute is unknown, which does not mean such routes are unreachable. The routes redistributed from other routing protocols have the incomplete attribute.

2) AS_PATH

AS_PATH is a well-known mandatory attribute. This attribute identifies the autonomous systems through which routing information carried in this Update message has passed. When a route is advertised from the local AS to another AS, each passed AS number is added into the AS_PATH attribute, thus the receiver can determine ASs to route the massage back. The number of the AS closest to the receiver's AS is leftmost, as shown below:

Figure 1-6 AS_PATH attribute



In general, a BGP router does not receive routes containing the local AS number to avoid routing loops.

Note

The current implementation supports using the **peer allow-as-loop** command to receive routes containing the local AS number to meet special requirements.

The AS_PATH attribute can be used for route selection and filtering. BGP gives priority to the route with the shortest AS_PATH length if other factors are the same. As shown in the above figure, the BGP router in AS50 gives priority to the route passing AS40 for sending data to the destination 8.0.0.0.

In some applications, you can apply a routing policy to control BGP route selection by modifying the AS_PATH length.

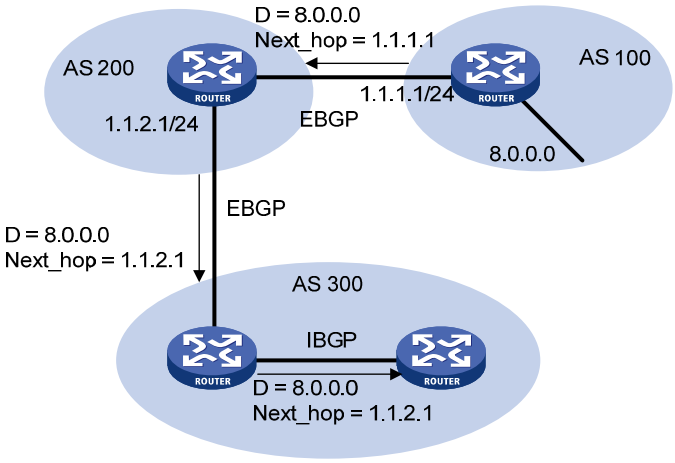
By configuring an AS path filtering list, you can filter routes based on AS numbers contained in the AS_PATH attribute.

3) NEXT_HOP

Different from IGP, the NEXT_HOP attribute may not be the IP address of a directly connected router. It involves three types of values, as shown in Figure 1-7.

- When advertising a self-originated route to an eBGP peer, a BGP speaker sets the NEXT_HOP for the route to the address of its sending interface.
- When sending a received route to an eBGP peer, a BGP speaker sets the NEXT_HOP for the route to the address of the sending interface.
- When sending a route received from an eBGP peer to an iBGP peer, a BGP speaker does not modify the NEXT_HOP attribute. If load-balancing is configured, the NEXT_HOP attribute will be modified. For load-balancing information, refer to BGP Route Selection.

Figure 1-7 NEXT_HOP attribute

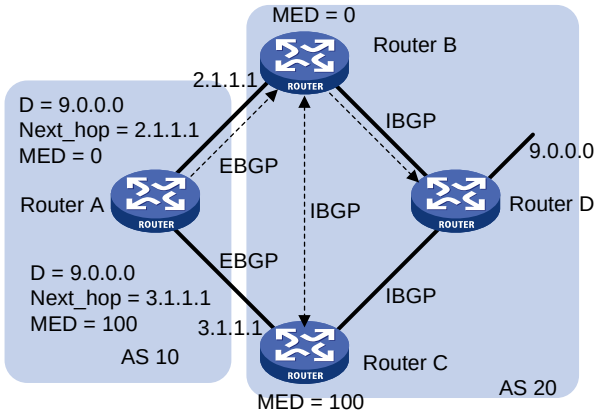


4) MED (MULTI_EXIT_DISC)

The MED attribute is exchanged between two neighboring ASs, each of which does not advertise the attribute to any other AS.

Similar with metrics used by IGP, MED is used to determine the best route for traffic going into an AS. When a BGP router obtains multiple routes to the same destination but with different next hops, it considers the route with the smallest MED value the best route if other conditions are the same. As shown below, traffic from AS10 to AS20 travels through Router B that is selected according to MED.

Figure 1-8 MED attribute



In general, BGP compares MEDs of routes received from the same AS only.

Note

The current implementation supports using the **compare-different-as-med** command to force BGP to compare MED values of routes received from different ASs.

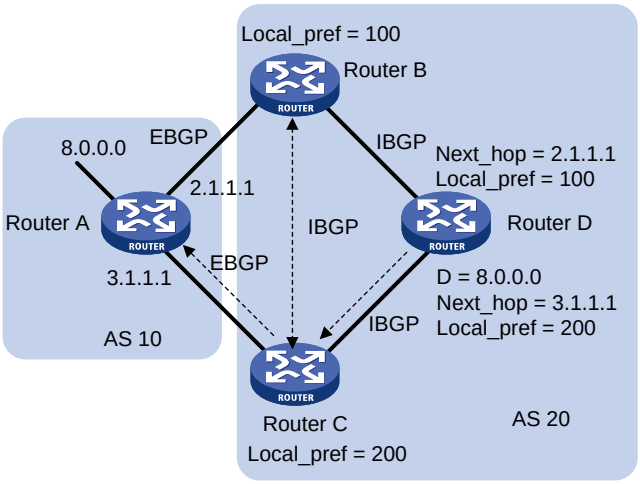
---

5) LOCAL_PREF

The LOCAL_PREF attribute is exchanged between iBGP peers only, and thus is not advertised to any other AS. It indicates the priority of a BGP router.

LOCAL_PREF is used to determine the best route for traffic leaving the local AS. When a BGP router obtains from several iBGP peers multiple routes to the same destination but with different next hops, it considers the route with the highest LOCAL_PREF value as the best route. As shown below, traffic from AS20 to AS10 travels through Router C that is selected according to LOCAL_PREF.

Figure 1-9 LOCAL_PREF attribute



6) COMMUNITY

The COMMUNITY attribute is used to simplify routing policy usage and ease management and maintenance. It identifies a collection of destination addresses having identical attributes, without physical boundaries in between, and having nothing to do with the local AS. Well known community attributes involve:

- Internet: By default, all routes belong to the Internet community. Routes with this attribute can be advertised to all BGP peers.
- No_Export: After received, routes with this attribute cannot be advertised out the local AS or out the local confederation but can be advertised to other sub-ASs in the confederation (for confederation information, refer to Settlements for Problems in Large Scale BGP Networks).
- No_Advertise: After received, routes with this attribute cannot be advertised to other BGP peers.
- No_Export_Subconfed: After received, routes with this attribute cannot be advertised out the local AS or other ASs in the local confederation.

BGP Route Selection

Route selection rules

The current BGP implementation supports the following route selection sequence:

- Discard routes with unreachable NEXT_HOPs first
- Select the route with the highest Preferred_value
- Select the route with the highest LOCAL_PREF
- Select the route originated by the local router
- Select the route with the shortest AS-PATH
- Select IGP, EGP, Incomplete routes in turn
- Select the route with the lowest MED value
- Select routes learned from eBGP, confederation, iBGP in turn

- Select the route with the smallest next hop cost
- Select the route with the shortest CLUSTER_LIST
- Select the route with the smallest ORIGINATOR_ID
- Select the route advertised by the router with the smallest Router ID
- Select the route with the lowest IP address

 **Note**

- CLUSTER_IDs of route reflectors form a CLUSTER_LIST. If a route reflector receives a route that contains its own CLUSTER ID in the CLUSTER_LIST, the router discards the route to avoid routing loops.
- If load balancing is configured, the system selects available routes to implement load balancing.

Route selection with BGP load balancing

The next hop of a BGP route may not be directly connected. One of the reasons is next hops in routing information exchanged between iBGPs are not modified. In this case, the BGP router needs to find the directly connected next hop via IGP. The matching route with the direct next hop is called the recursive route. The process of finding a recursive route is route recursion.

Currently, the system supports BGP load balancing based on route recursion, namely, if multiple recursive routes to the same destination are load balanced (suppose three direct next hop addresses), BGP generates the same number of next hops to forward packets. Note that BGP load balancing based on route recursion is always enabled by the system rather than configured using commands.

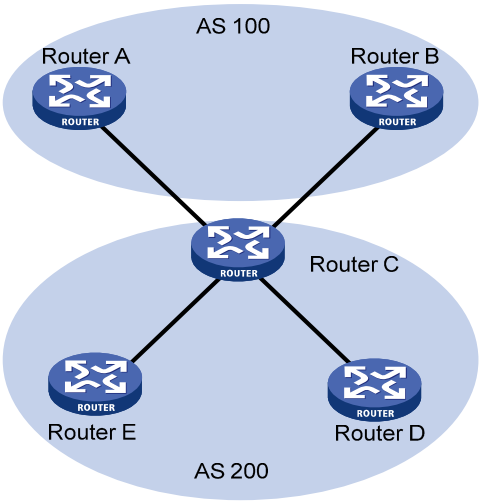
BGP differs from IGP in the implementation of load balancing in the following:

- IGP routing protocols such as RIP, OSPF compute metrics of routes, and then implement load balancing over routes with the same metric and to the same destination. The route selection criterion is metric.
- BGP has no route computation algorithm, so it cannot implement load balancing according to metrics of routes. However, BGP has abundant route selection rules, through which, it selects available routes for load balancing and adds load balancing to route selection rules.

 **Note**

- BGP implements load balancing only on routes that have the same AS_PATH, ORIGIN, LOCAL_PREF and MED.
- BGP load balancing is applicable between eBGP peers, between iBGP peers and between confederations.
- If multiple routes to the same destination are available, BGP selects a configurable number of routes for load balancing.

Figure 1-10 Network diagram for BGP load balancing



In the above figure, Router D and Router E are iBGP peers of Router C. Router A and Router B both advertise a route destined for the same destination to Router C. If load balancing is configured and the two routes have the same AS_PATH attribute, ORIGIN attribute, LOCAL_PREF and MED, Router C installs both the two routes to its route table for load balancing. After that, Router C forwards to Router D and Router E the route that has AS_PATH unchanged but has NEXT_HOP changed to Router C; other BGP transitive attributes are those of the best route.

BGP route advertisement rules

The current BGP implementation supports the following route advertisement rules:

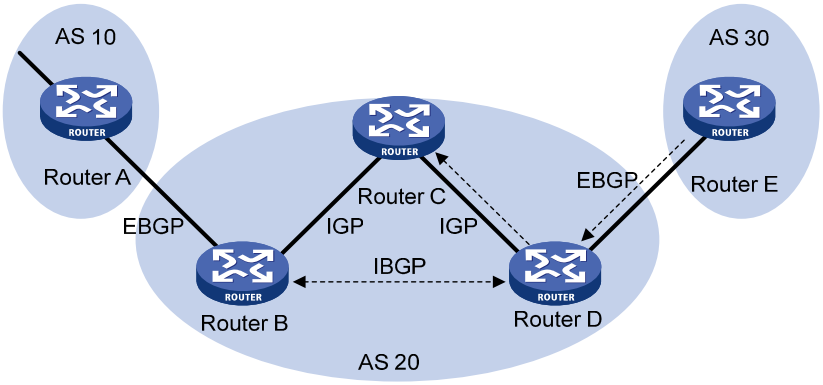
- When multiple feasible routes to a destination exist, the BGP speaker advertises only the best route to its peers.
- A BGP speaker advertises only routes used by itself.
- A BGP speaker advertises routes learned through eBGP to all BGP peers, including both eBGP and iBGP peers.
- A BGP speaker does not advertise routes from an iBGP peer to other iBGP peers.
- A BGP speaker advertises routes learned through iBGP to eBGP peers. Note that if BGP and IGP synchronization is disabled, those routes are advertised to eBGP peers directly. If the feature is enabled, only after IGP advertises those routes, can BGP advertise the routes to eBGP peers.
- A BGP speaker advertises all routes to a newly connected peer.

iBGP and IGP Synchronization

Routing information synchronization between iBGP and IGP avoids giving wrong directions to routers outside of the local AS.

If a non-BGP router works in an AS, it may discard a packet due to an unreachable destination. As shown in Figure 1-11, Router E has learned a route of 8.0.0.0/8 from Router D via BGP. Then Router E sends a packet to 8.0.0.0/8 through Router D, which finds from its routing table that Router B is the next hop (configured using the **peer next-hop-local** command). Because Router D has learned the route to Router B via IGP, it forwards the packet to Router C through route recursion. Router C has no idea about the route 8.0.0.0/8, so it discards the packet.

Figure 1-11 iBGP and IGP synchronization



If synchronization is enabled in this example, only when the route 8.0.0.0/24 received from Router B is available in its IGP routing table, can Router D add the route into its BGP routing table and advertise the route to the eBGP peer.

You can disable the synchronization feature in the following cases:

- The local AS is not a transitive AS (AS20 is a transitive AS in the above figure).
- Routers in the local AS are iBGP fully meshed.

Settlements for Problems in Large Scale BGP Networks

Route summarization

Route summarization can reduce the routing table size on a large network, and allow BGP routers to advertise only summary routes rather than more specific routes.

Currently, the system supports both manual and automatic route summarization. Manual route summarization allows you to determine the attribute of a summary route and whether to advertise the route.

Route dampening

BGP route dampening is used to solve the issue of route instability such as route flaps, that is, a route comes up and disappears in the routing table frequently.

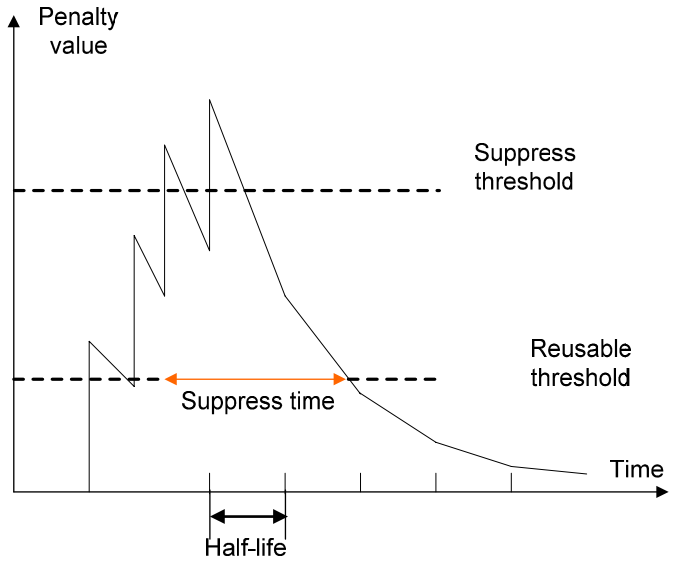
When a route flap occurs, the routing protocol sends an update to its neighbor, and then the neighbor needs to recalculate routes and modify the routing table. Therefore, frequent route flaps consume large bandwidth and CPU resources and even affect network normal operation.

In most cases, BGP is used in complex networks, where route changes are very frequent. To solve the problem caused by route flaps, BGP route dampening is used to suppress unstable routes.

BGP route dampening uses a penalty value to judge the stability of a route. The bigger the value, the less stable the route. Each time a route flap occurs, BGP adds a penalty value (1000, which is a fixed number and cannot be changed) to the route. When the penalty value of the route exceeds the suppress value, the route is suppressed from being added into the routing table or being advertised to other BGP peers.

The penalty value of the suppressed route will decrease to a half of the suppress value after a period of time. This period is called Half-life. When the value decreases to the reusable threshold value, the route is added into the routing table and advertised to other BGP peers.

Figure 1-12 BGP route dampening



Peer group

You can organize BGP peers with the same attributes into a group to simplify configurations on them.

When a peer joins the peer group, the peer obtains the same configuration as the peer group. If the configuration of the peer group is changed, the configuration of group members is changed accordingly.

When a peer is added into a peer group, the peer enjoys the same route update policy as the peer group to improve route distribution efficiency.

Caution

If an option is configured for both a peer and its peer group, the last configuration takes effect.

Community

A peer group makes peers in it enjoy the same policy, while a community makes a group of BGP routers in several ASs enjoy the same policy. Community is a path attribute and advertised between BGP peers, without being limited by AS.

A BGP router can modify the community attribute for a route before sending it to other peers.

Besides using well-known community attributes, you can define extended community attributes by using a community list to define a routing policy.

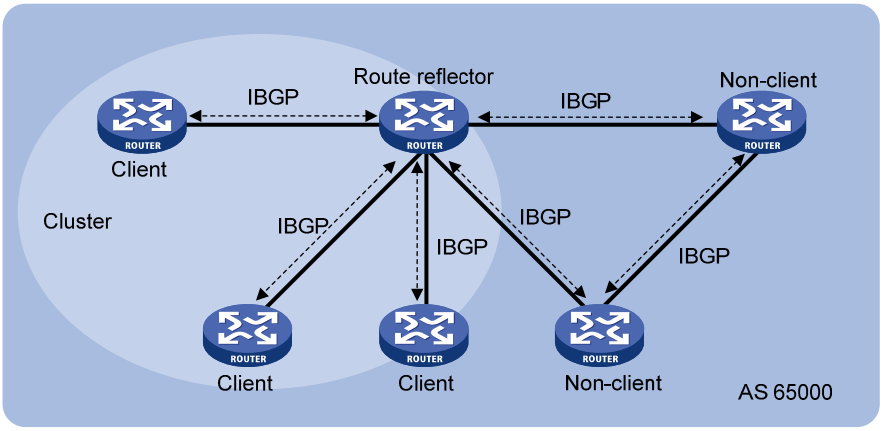
Route reflector

iBGP peers should be fully meshed to maintain connectivity. If there are n routers in an AS, the number of iBGP connections is n (n-1)/2, and therefore large amounts of network and CPU resources will be consumed.

Using route reflectors can solve this issue. In an AS, a router acts as a route reflector, and other routers act as clients connecting to the route reflector. The route reflector forwards routing information between clients, and thus BGP sessions between clients need not be established.

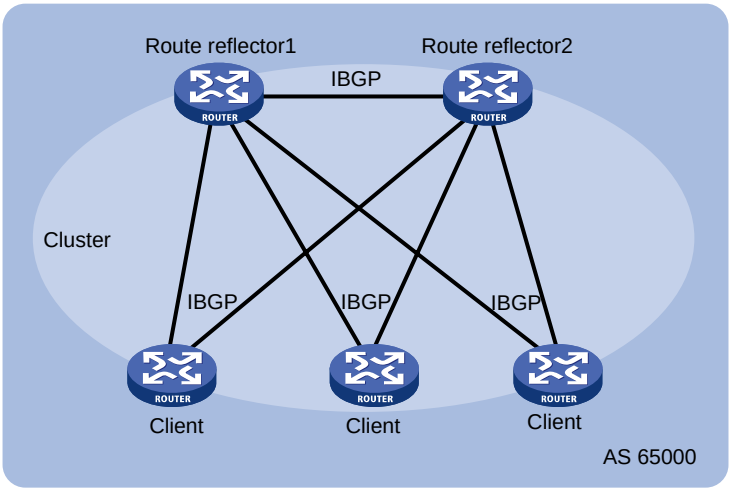
A router that is neither a route reflector nor a client is a non-client, which has to establish BGP sessions to the route reflector and other non-clients, as shown below.

Figure 1-13 Network diagram for route reflector



The route reflector and clients form a cluster. In some cases, you can configure more than one route reflector in a cluster to improve network reliability and prevent single point failure, as shown in the following figure. The configured route reflectors must have the same Cluster_ID to avoid routing loops.

Figure 1-14 Network diagram for route reflectors



When the BGP routers in an AS are fully meshed, route reflection is unnecessary because it consumes more bandwidth resources. You can use related commands to disable route reflection in this case.

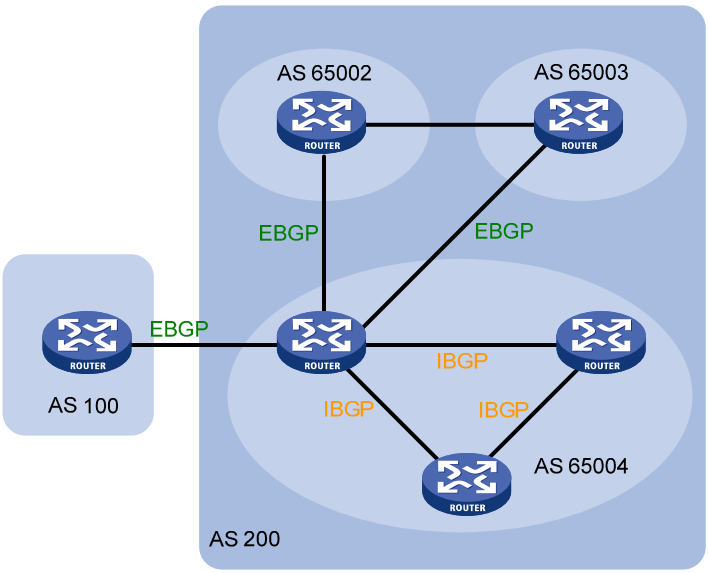
Note

After route reflection is disabled between clients, routes can still be reflected between a client and a non-client.

Confederation

Confederation is another method to deal with growing iBGP connections in ASs. It splits an AS into multiple sub-ASs. In each sub-AS, iBGP peers are fully meshed, and intra-confederation eBGP connections are established between sub-ASs, as shown below:

Figure 1-15 Confederation network diagram



From the perspective of a non-confederation BGP speaker, it needs not know sub-ASs in the confederation. The ID of the confederation is the number of the AS. In the above figure, AS 200 is the confederation ID.

The deficiency of confederation is: when changing an AS into a confederation, you need to reconfigure your routers, and the topology will be changed.

In large-scale BGP networks, both route reflector and confederation can be used.

BGP GR

Note

For GR (Graceful Restart) information, refer to *GR Configuration* in the *System Volume*.

---

1) To establish a BGP session with a peer, a BGP GR Restarter sends an OPEN message with GR capability to the peer.
2) Upon receipt of this message, the peer is aware that the sending router is capable of Graceful Restart, and sends an OPEN message with GR Capability to the GR Restarter to establish a GR session. If neither party has the GR capability, the session established between them will not be GR capable.
3) When a restart occurs on a device that acts as the GR Restarter, sessions on it will go down Then, GR capable peers will mark all routes associated with the GR Restarter as stale. However, during the configured GR Time, they still use these routes for packet forwarding.

4) After the restart is completed, the GR Restarter will reestablish GR sessions with its peers and send a new GR message notifying the completion of restart. Routing information is exchanged between them for the GR Restarter to create a new routing table and forwarding table and have stale routing information removed. Then the BGP routing convergence is complete.

MP-BGP

Overview

BGP-4 supports IPv4 unicasts, but does not support other network layer protocols like IPv6.

To support more network layer protocols, IETF extended BGP-4 by introducing Multiprotocol Extensions for BGP-4 (MP-BGP) in RFC 4760.

Routers supporting MP-BGP can communicate with routers not supporting MP-BGP.

MP-BGP extended attributes

In BGP-4, the three types of attributes for IPv4 address format, namely NLRI, NEXT_HOP and AGGREGATOR (AGGREGATOR contains the IP address of the speaker generating the summary route) are all carried in updates.

To support multiple network layer protocols, BGP-4 puts information about network layer into NLRI and NEXT_HOP. MP-BGP introduced two path attributes:

- MP_REACH_NLRI: Multiprotocol Reachable NLRI, for advertising feasible routes and next hops
- MP_UNREACH_NLRI: Multiprotocol Unreachable NLRI, for withdrawing unfeasible routes

The above two attributes are both optional non-transitive, so BGP speakers not supporting multi-protocol ignore the two attributes and do not forward them to its peers.

Address family

MP-BGP uses address families to differentiate network layer protocols. For address family values, refer to RFC 1700 (Assigned Numbers). Currently, the system supports multiple MP-BGP extensions, including VPN extension and IPv6 extension. Different extensions are configured in respective address family view.

Note

- For information about the VPN extension application, refer to MCE Configuration in the *IP Routing Volume*.
- For information about the IPv6 extension application, refer to IPv6 BGP Configuration in the *IP Routing Volume*.
- This chapter gives no detailed commands related to any specific extension application in MP-BGP address family view.

Protocols and Standards

- RFC1771: A Border Gateway Protocol 4 (BGP-4)
- RFC2858: Multiprotocol Extensions for BGP-4
- RFC3392: Capabilities Advertisement with BGP-4

- RFC2918: Route Refresh Capability for BGP-4
- RFC2439: BGP Route Flap Damping
- RFC1997: BGP Communities Attribute
- RFC2796: BGP Route Reflection
- RFC3065: Autonomous System Confederations for BGP
- draft-ietf-idr-restart-08: Graceful Restart Mechanism for BGP

BGP Configuration Task List

Complete the following tasks to configure BGP:

Task		Remarks
Configuring BGP Basic Functions	Creating a BGP Connection	Required
	Specifying the Source Interface for TCP Connections	Optional
	Allowing Establishment of eBGP Connection to a Non Directly Connected Peer/Peer Group	Optional
Controlling Route Generation	Injecting a Local Network	Required to choose either.
	Configuring BGP Route Redistribution	
	Enabling Default Route Redistribution into BGP	Optional
Controlling Route Distribution and Reception	Configuring BGP Route Summarization	Optional
	Advertising a Default Route to a Peer or Peer Group	
	Configuring BGP Route Distribution/Reception Filtering Policies	
	Enabling BGP and IGP Route Synchronization	
	Limiting Prefixes Received from a Peer/Peer Group	
	Configuring BGP Route Dampening	
	Configuring a Shortcut Route	
Configuring BGP Route Attributes	Specifying a Preferred Value for Routes Received	Optional
	Configuring Preferences for BGP Routes	Optional
	Configure the Default Local Preference	Optional
	Configuring the MED Attribute	Optional
	Configuring the Next Hop Attribute	Optional
	Configuring the AS-PATH Attribute	Optional

Task		Remarks
Tuning and Optimizing BGP Networks	Configuring BGP Keepalive Interval and Holdtime	Optional
	Configuring the Interval for Sending the Same Update	Optional
	Configuring BGP Soft-Reset	Optional
	Enabling Quick eBGP Session Reestablishment	Optional
	Enabling MD5 Authentication for TCP Connections	Optional
	Configuring BGP Load Balancing	Optional
	Forbiding Session Establishment with a Peer or Peer Group	Optional
Configuring a Large Scale BGP Network	Configuring BGP Peer Groups	Optional
	Configuring BGP Community	Optional
	Configuring a BGP Route Reflector	Optional
	Configuring a BGP Confederation	Optional
Configuring BGP GR		Optional
Enabling Trap		Optional
Enabling Logging of Peer State Changes		Optional

Configuring BGP Basic Functions

Note

This section does not differentiate between BGP and MP-BGP.

Prerequisites

The neighboring nodes are accessible to each other at the network layer.

Creating a BGP Connection

A router ID is the unique identifier of a BGP router in an AS.

- To ensure the uniqueness of a router ID and enhance network reliability, you can specify in BGP view the IP address of a local loopback interface as the router ID.
- If no router ID is specified in BGP view, the global router ID is used. For information about global router ID, refer to *IP Routing Overview* in the *IP Routing Volume*.
- If the global router ID is used and then it is removed, the system will select a new router ID.
- If the router ID is specified in BGP view, using the **undo router-id** command can make the system select a new router ID.

Follow these steps to create a BGP connection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable BGP and enter BGP view	**bgp** *as-number*	— Not enabled by default
Specify a Router ID	**router-id** *router-id*	Optional By default, the global router ID is used.
Specify a peer or a peer group and its AS number	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	Required Not specified by default
Enable the default use of IPv4 unicast address family for the peers that are established using the **peer as-number** command	**default ipv4-unicast**	Optional Enabled by default
Enable a peer	**peer** *ip-address* **enable**	Optional Enabled by default
Configure a description for a peer/peer group	**peer** { *group-name* | *ip-address* } **description** *description-text*	Not configured by default.

Caution

- Since a router can reside in only one AS, the router can run only one BGP process.
- You need to create a peer group before configuring it.

Specifying the Source Interface for TCP Connections

BGP uses TCP as the transport layer protocol. By default, BGP uses the output interface of the optimal router to a peer as the source interface for establishing TCP connections to the peer. If a BGP router has multiple links to a peer, when the source interface fails, BGP has to reestablish TCP connections, causing network oscillation. Therefore, it is recommended to use a loopback interface as the source interface to enhance stability of BGP connections.

Follow these steps to specify the source interface of TCP connections:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Specify the source interface for establishing TCP connections to a peer or peer group	**peer** { *group-name* | *ip-address* } **connect-interface** *interface-type interface-number*	Required By default, BGP uses the outbound interface of the best route to the BGP peer/peer group as the source interface for establishing a TCP connection to the peer/peer group.

Caution

To establish multiple BGP connections between two routers, you need to specify on the local router the source interface for establishing TCP connections to each peer; otherwise, the local BGP router may fail to establish TCP connections to a peer when using the outbound interface of the best route to the peer as the source interface.

Allowing Establishment of eBGP Connection to a Non Directly Connected Peer/Peer Group

In general, direct physical links should be available between eBGP peers. If not, you can use the **peer ebgp-max-hop** command to establish a TCP connection over multiple hops between two peers.

Follow these steps to allow establishment of eBGP connection to a non directly connected peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Allow the establishment of eBGP connection to a non directly connected peer/peer group	**peer** { *group-name* | *ip-address* } **ebgp-max-hop** [*hop-count*]	Optional Not allowed by default.

Note

The **peer ebgp-max-hop** command needs not be configured if the two eBGP peers are directly connected.

Controlling Route Generation

Different from IGP, BGP focuses on route generation and advertisement control and optimal route selection.

There are to ways to generate BGP routes:

- Configure BGP to advertise local networks
- Configure BGP to redistribute routes from other routing protocols, including the default route

Prerequisites

BGP connections have been created.

Injecting a Local Network

In BGP view, you can inject a local network to allow BGP to advertise it to BGP peers. The origin attribute of routes advertised in this way is IGP. You can also reference a route policy to flexibly control route advertisement. The network to be injected must be available in the local IP routing table.

Follow these steps to inject a local network:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Inject a network to the BGP routing table	**network** *ip-address* [*mask* | *mask-length*] **route-policy** *route-policy-name*	Optional Not injected by default

Configuring BGP Route Redistribution

BGP does not find routes by itself. Rather, it redistributes routing information in the local AS from other routing protocols. During route redistribution, you can configure BGP to filter routing information from specific routing protocols.

The origin attribute of routes redistributed using the **import-route** command is Incomplete.

Follow these steps to configure BGP route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable route redistribution from a routing protocol into BGP	**import-route** *protocol* [*process-id* | **all-processes**] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required Not enabled by default

Enabling Default Route Redistribution into BGP

Using the **import-route** command cannot redistribute a default route. To do so, complete the following configuration.

Follow these steps to enable default route redistribution into BGP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Enable route redistribution from a routing protocol into BGP	**import-route** *protocol* [*process-id* | **all-processes**] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required Not redistributed by default
Enable default route redistribution into BGP	**default-route imported**	Optional Not enabled by default

Controlling Route Distribution and Reception

Prerequisites

BGP connections have been created.

Configuring BGP Route Summarization

To reduce the routing table size on medium and large BGP networks, you need to configure route summarization on BGP routers. BGP supports two summarization modes: automatic and manual. Manual summary routes enjoy a higher priority than automatic ones.

Configure automatic route summarization

After automatic route summarization is configured, BGP summarizes redistributed IGP subnets to advertise only natural networks. Routes injected with the **network** command can not be summarized.

Follow these steps to configure automatic route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure automatic route summarization	**summary automatic**	Required Not configured by default.

Configure manual route summarization

By configuring manual route summarization, you can summarize both redistributed routes and routes injected using the **network** command and determine the mask length for a summary route as needed.

Follow these steps to configure BGP manual route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure manual route summarization	**aggregate** *ip-address* { *mask* | *mask-length* } [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*]*	Required Not configured by default.

Advertising a Default Route to a Peer or Peer Group

After this task is configured, the BGP router sends a default route with the next hop being itself to the specified peer/peer group, regardless of whether the default route is available in the routing table.

Follow these steps to advertise a default route to a peer or peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Advertise a default route to a peer or peer group	**peer** { *group-name* | *ip-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	Required Not advertised by default

Configuring BGP Route Distribution/Reception Filtering Policies

Prerequisites

You need to configure following filters as needed.

- ACL
- IP prefix list
- Route policy
- AS-path ACL

For how to configure an ACL, refer to *ACL Configuration* in the *Security Volume*.

For how to configure an IP prefix list, route policy and AS-path ACL, refer to *Route Policy Configuration* in the *Routing Volume*.

Configure BGP route distribution filtering policies

Follow these steps to configure BGP route distribution filtering policies:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Configure the filtering of redistributed routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* **| ospf** *process-id* **| rip** *process-id* **|** | **static**]	Required to choose any; Not configured by default. You can configure a filtering policy as needed; If several filtering policies are configured, they are applied in the following sequence: • **filter-policy export** • **peer filter-policy export** • **peer as-path-acl export** • **peer ip-prefix export** • **peer route-policy export** Only routes pass the first policy, can they go to the next, and only routes passing all the configured policies, can they be advertised.
Reference a routing policy to filter advertisements to a peer/peer group	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	
Reference an ACL to filer advertisements to a peer/peer group	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **export**	
Reference an AS path ACL to filter routing information sent to a peer/peer group	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **export**	
Reference an IP prefix list to filer routing information sent to a peer/peer group	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **export**	

Configure BGP route reception filtering policies

Only routes permitted by the configured filtering policies can be installed into the local BGP routing table.

Members of a peer group can have different route reception filtering policies from the peer group.

Follow these steps to configure BGP route reception filtering policies:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Filter incoming routes with an ACL or IP prefix list	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	Required to choose any; No route reception filtering is configured by default; You can configure a filtering policy as needed; If several filtering policies are configured, they are applied in the following sequence: • **filter-policy import** • **peer filter-policy import** • **peer as-path-acl import** • **peer ip-prefix import** • **peer route-policy import** Only routes passing the first policy, can they go to the next, and only routes passing all the configured policies, can they be received.
Reference a routing policy to filter routes from a peer/peer group	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **import**	
Reference an ACL to filter routing information from a peer/peer group	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **import**	
Reference an AS path ACL to filter routing information from a peer/peer group	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **import**	
Reference an IP prefix list to filter routing information from a peer/peer group	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **import**	

Enabling BGP and IGP Route Synchronization

By default, when a BGP router receives an iBGP route, it only checks the reachability of the route's next hop before advertisement. With BGP and IGP synchronization enabled, the BGP router cannot advertise the iBGP route to eBGP peers unless the route is also available in the IGP routing table.

Follow these steps to enable BGP and IGP synchronization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable synchronization between BGP and IGP	**synchronization**	Required Not enabled by default

Limiting Prefixes Received from a Peer/Peer Group

Follow these steps to configure the maximum number of prefixes allowed to be received from a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify the maximum number of prefixes that can be received from a peer/peer group. If the number is reached, the router breaks down the BGP connection to the peer.	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* [*percentage-value*]	Required to choose any; No limit is configured by default.
Specify the maximum number of prefixes that can be received from a peer/peer group. If the number is reached, the router outputs alert information but does not break down the BGP connection to the peer.	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* **alert-only** [*percentage-value*]	
Specify the maximum number of prefixes that can be received from a peer/peer group. If the number is reached, the router breaks down the BGP connection to the peer and then reestablishes a BGP connection to the peer.	**peer** { *group-name* | *ip-address* } **route-limit** *prefix-number* **reconnect** *reconnect-time* [*percentage-value*]	

Configuring BGP Route Dampening

By configuring BGP route dampening, you can suppress unstable routes from being added to the local routing table or being advertised to BGP peers.

Follow these steps to configure BGP route dampening:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Configure BGP route dampening	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	Required Not configured by default.

Configuring a Shortcut Route

An eBGP route received has a priority of 255, lower than a local route. This task allows you configure an eBGP route as a shortcut route that has the same priority as a local route and thus has greater likehood to become the optimal route.

Follow these steps to configure a shortcut route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure a shortcut route	**network** *ip-address* [*mask* | *mask-length*] **short-cut**	Optional By default, an eBGP route received has a priority of 255.

Configuring BGP Route Attributes

Prerequisites

BGP connections have been created.

Specifying a Preferred Value for Routes Received

By default, routes received from a peer have a preferred value of 0. Among multiple routes that have the same destination/mask and are learned from different peers, the one with the greatest preferred value is selected as the route to the destination.

Follow these steps to specify a preferred value for routes from a peer or peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify a preferred value for routes received from a peer or peer group	**peer** { *group-name* | *ip-address* } **preferred-value** *value*	Optional The preferred value is 0 by default.

Configuring Preferences for BGP Routes

A router may run multiple routing protocols, each of which has a preference specified. If they find the same route, the route found by the routing protocol with the highest preference is selected.

This task allows you configure preferences for external, internal, local BGP routes and reference a route policy to set preferences for matching routes as needed.

Follow these steps to configure preferences for BGP routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure preferences for external, internal, local BGP routes	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	Optional The default preferences of external, internal, and local BGP routes are 255, 255, and 130 respectively.

Configure the Default Local Preference

The local preference is used to determine the best route for traffic leaving the local AS. When a BGP router obtains from several iBGP peers multiple routes to the same destination but with different next hops, it considers the route with the highest local preference as the best route.

This task allows you to specify the default local preference for routes sent to iBGP peers.

Follow these steps to specify the default local preference:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the default local preference	**default local-preference** *value*	Optional 100 by default

Configuring the MED Attribute

MED is used to determine the best route for traffic going into an AS. When a BGP router obtains from eBGP peers multiple routes to the same destination but with different next hops, it considers the route with the smallest MED value as the best route if other conditions are the same.

Configure the default MED value

Follow these steps to configure the default MED value:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the default MED value	**default med** *med-value*	Optional 0 by default

Enable the comparison of MED of routes from different ASs

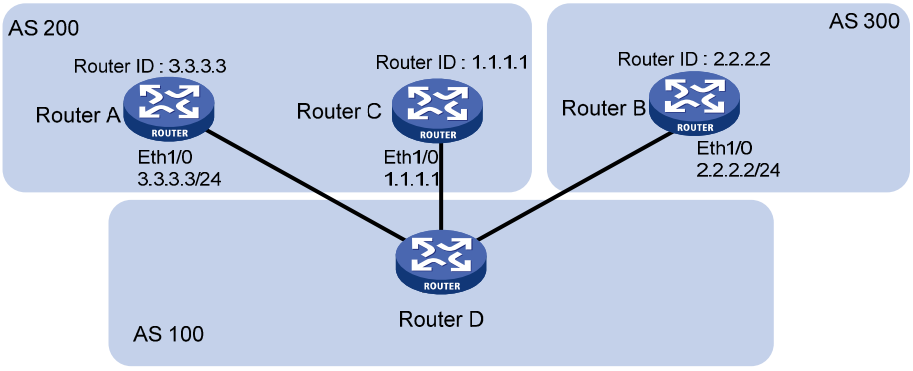
Follow these steps to enable the comparison of MED of routes from different ASs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable the comparison of MED of routes from different ASs	**compare-different-as-med**	Required Not enabled by default

Enable the comparison of MED of routes from each AS

Route learning sequence may affect optimal route selection.

Figure 1-16 Route selection based on MED



As shown in the figure above, Router D learns network 10.0.0.0 from both Router A and Router B. Because Router B has a smaller router ID, the route learned from it is optimal.

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    3.3.3.3         50                      0       200e
```

When Router D learns network 10.0.0.0 from Router C which has a smaller router ID than Router B, the route from Router C becomes optimal, as shown below.

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          1.1.1.1         60                      0       200e
 * i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    3.3.3.3         50                      0       200e
```

However, Router C and Router B reside in the same AS, and therefore BGP will compare the MEDs of them. Since Router C has a greater MED, network 10.0.0.0 learned from it is not optimal.

In this case, you can configure the **bestroute compare-med** command on Router D. After that, Router D will put routes received from the same AS into a group. For the same group, the route with the lowest MED is selected. Then, it compares routes from different groups. This mechanism avoids the above-mentioned problem. The following output is the BGP routing table on Router D after the comparison of MED of routes from each AS is enabled. Network 10.0.0.0 learned from Router C is the optimal route.

```
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn
 *>i  10.0.0.0          3.3.3.3         50                      0       200e
 * i  10.0.0.0          2.2.2.2         50                      0       300e
 * i                    1.1.1.1         60                      0       200e
```

Note that, in this case, BGP load balancing cannot be implemented because load balanced routes must have the same AS-path attribute.

Follow these steps to enable the comparison of MED of routes from each AS:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable the comparison of MED of routes from each AS	**bestroute compare-med**	Optional Not enabled by default

Enable the comparison of MED of routes from confederation peers

Follow these steps to enable the comparison of MED of routes from confederation peers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable the comparison of MED of routes from confederation peers	**bestroute med-confederation**	Optional Not enabled by default

Note

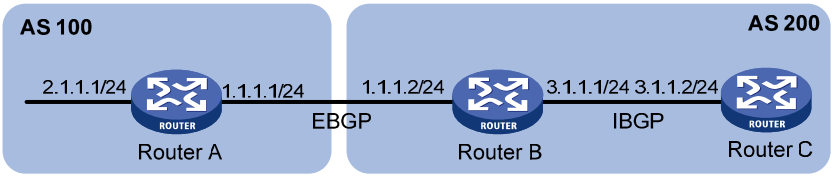
The MED attributes of routes from confederation peers are not compared if their AS-path attributes contain AS numbers that don't belong to the confederation. For example, there are three routes: AS-path attributes of them are 65006 65009, 65007 65009 and 65008 65009, and MED values of them are 2, 3, and 1. Because the third route contains an AS number that does not belong to the confederation, the first route becomes the optimal route.

Configuring the Next Hop Attribute

By default, when advertising routes to an iBGP peer/peer group, a BGP router does not set itself as the next hop. However, to ensure a BGP peer can find the correct next hop in some cases, you need to configure the router as the next hop for routes sent to the peer.

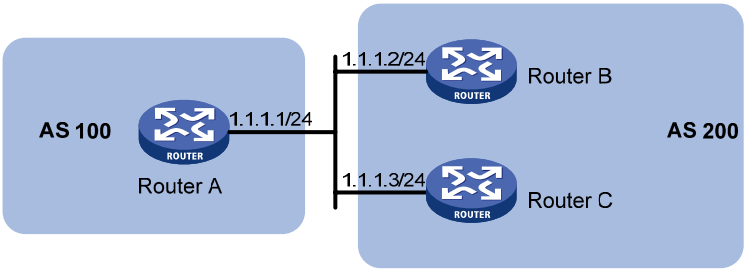
For example, as shown in the figure below, Router A and Router B establish an eBGP neighbor relationship, and Router B and Router C establish an iBGP neighbor relationship. When Router B advertises a network learned from Router A to Router C, if Router C has no route to IP address 1.1.1.1/24, you need to configure Router B to set itself as the next hop (3.1.1.1/24) for the route to be sent to Router C.

Figure 1-17 Next hop attribute configuration



If a BGP router has two peers on a common broadcast network, it does not set itself as the next hop for routes sent to an eBGP peer by default. As shown below, Router A and Router B establish an eBGP neighbor relationship, and Router B and Router C establish an iBGP neighbor relationship. They are on the same broadcast network 1.1.1.0/24. When Router B sends eBGP routes to Router A, it does not set itself as the next hop by default. However, you can configure Router B to set it as the nexthop (1.1.1.2/24) for routes sent to Router A by using the **peer next-hop-local** command as needed.

Figure 1-18 Next hop attribute configuration



Note that: if you have configured BGP load balancing on a BGP router, the router will set it as the next hop for routes sent to an iBGP peer/peer group regardless of whether the **peer next-hop-local** command is configured.

Follow these steps to configure the next hop attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify the router as the next hop of routes sent to a peer/peer group	**peer** { *group-name* | *ip-address* } **next-hop-local**	Optional By default, the router sets it as the next hop for routes sent to an eBGP peer/peer group, but does not set it as the next hop for routes sent to an iBGP peer/peer group.

Configuring the AS-PATH Attribute

Permit local AS number to appear in routes from a peer/peer group

In general, BGP checks whether the AS_PATH attribute of a route from a peer contains the local AS number. If so, it discards the route to avoid routing loops.

This task allows you to permit local AS number to appear in routes from a peer/peer group and specify the appearance times.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Permit local AS number to appear in routes from a peer/peer group and specify the appearance times	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	Optional By default, the local AS number is not allowed.

Disable BGP from considering AS_PATH during best route selection

Follow these steps to disable BGP from considering AS_PATH during best route selection:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Disable BGP from considering AS_PATH during best route selection	**bestroute as-path-neglect**	Optional By default, BGP considers AS_PATH during best route selection.

Specify a fake AS number for a peer/peer group

When Router A in AS 2 is moved to AS 3, you can configure Router A to specify a fake AS number of 2 for created connections to eBGP peers/peer groups. In this way, these eBGP peers still think Router A is in AS 2 and thus need not change their configurations. This feature ensures uninterrupted BGP services.

Follow these steps to specify a fake AS number for a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify a fake AS number for a peer/peer group	**peer** { *group-name* | *ip-address* } **fake-as** *as-number*	Optional Not specified by default

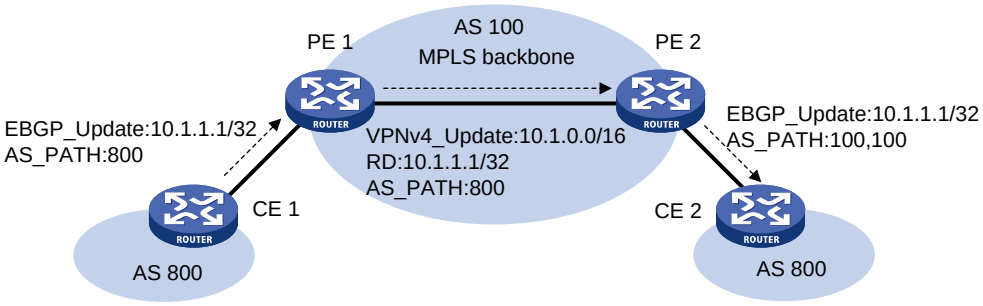
Note

This command is only applicable to an eBGP peer or peer group.

Configure AS number substitution

In MPLS L3VPN, if eBGP is used between PE and CE, sites in different geographical areas should have different AS numbers assigned to ensure correct route advertisement. If different CEs use the same AS number, you need to configure the corresponding PE to replace the AS number of the CE as its own AS number. This feature is used for route advertisement only.

Figure 1-19 AS number substitution configuration



As shown in the above figure, CE 1 and CE 2 use the same AS number of 800. If AS number substitution for CE 2 is configured on PE 2, when PE 2 receives a BGP update sent from CE 1, it replaces AS number 800 as its own AS number 100. Similar configuration should also be made on PE 1.

Follow these steps to configure AS number substitution for a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Replace the AS number of a peer/peer group in the AS_PATH attribute as the local AS number	**peer** { *group-name* | *ip-address* } **substitute-as**	Optional Not configured by default.

Caution

Improper AS number substitution configuration may cause route loops; use this command with caution.

Remove private AS numbers from updates to a peer/peer group

Follow these steps to remove private AS numbers from updates to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure BGP to remove private AS numbers from the AS_PATH attribute of updates to a peer/peer group	**peer** { *group-name* | *ip-address* } **public-as-only**	Optional By default, BGP updates carry private AS numbers.

Tuning and Optimizing BGP Networks

Prerequisites

BGP connections have been created.

Configuring BGP Keepalive Interval and Holdtime

After establishing a BGP connection, two routers send keepalive messages periodically to each other to keep the connection. If a router receives no keepalive or update message from the peer within the holdtime, it tears down the connection.

If two parties have the same timer assigned with different values, the smaller one is used by the two parties.

Follow these steps to configure BGP keepalive interval and holdtime:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the global keepalive interval and holdtime	**timer keepalive** *keepalive* **hold** *holdtime*	Optional By default, the keepalive interval is 60 seconds, and holdtime is 180 seconds.
Configure the keepalive interval and holdtime for a peer/peer group	**peer** { *group-name* | *ip-address* } **timer keepalive** *keepalive* **hold** *holdtime*	

Caution

- The maximum keepalive interval should be one third of the holdtime and no less than 1 second. The holdtime is no less than 3 seconds unless it is set to 0.
- The intervals set with the **peer timer** command are preferred to those set with the **timer** command.
- If the router has established a neighbor relationship with a peer, you need to reset the BGP connection to validate the new set timers.

Configuring the Interval for Sending the Same Update

Follow these steps to configure the interval for sending the same update to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the interval for sending the same update to a peer/peer group	**peer** { *group-name* | *ip-address* } **route-update-interval** *interval*	Optional The intervals for sending the same update to an iBGP peer and an eBGP peer default to 15 seconds and 30 seconds respectively.

Configuring BGP Soft-Reset

After modifying a route selection policy, you have to reset BGP connections to make the new one take effect, causing short time disconnection.

The current BGP implementation supports the route-refresh capability, with which, a router can dynamically refresh its BGP routing table when the route selection policy is modified, without tearing down BGP connections. If a BGP peer does not support route-refresh, you need to save updates from the peer on the local router. After that, when a route selection policy is modified, the router can refresh its BGP routing table by using such updates without tearing down BGP connections.

Configure automatic soft-reset

After route refresh is enabled for peers and then a policy is modified, the router advertises a route-refresh message to the peers, which then resend their routing information to the router. In this way, the router can perform dynamic route update and apply the new policy without tearing down BGP connections.

Follow these steps to enable BGP route refresh for a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable BGP route refresh for a peer/peer group	**peer** { *group-name* | *ip-address* } **capability-advertise route-refresh**	Optional Enabled by default

Configure manual soft-reset

If a BGP peer does not support route-refresh, you need to save updates from the peer on the local router by using the **peer keep-all-routes** command. When a route selection policy is modified, you can use the **refresh bgp** command to refresh the BGP routing table by applying the new policy.

Following these steps to save all route updates from a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Disable BGP route-refresh and multi-protocol extension capability for a peer/peer group	**peer** { *group-name* | *ip-address* } **capability-advertise conventional**	Optional Enabled by default
Save all routes from a peer/peer group	**peer** { *group-name* | *ip-address* } **keep-all-routes**	Optional Not saved by default
Return to user view	**return**	—
Perform manual soft reset on BGP connections	**refresh bgp** { **all** | *ip-address* | **group** *group-name* | **external** | **internal** } { **export** | **import** }	Required

Enabling Quick eBGP Session Reestablishment

If the router receives no keepalive messages from a BGP peer within the holdtime, it tears down the connection to the peer.

With quick eBGP connection reestablishment enabled, the router, when the link to a directly connected eBGP peer is down, will reestablish a session to the eBGP peer immediately.

Follow these steps to enable quick eBGP session reestablishment:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable quick eBGP session reestablishment	**ebgp-interface-sensitive**	Optional Not enabled by default

Enabling MD5 Authentication for TCP Connections

BGP employs TCP as the transport protocol. To enhance security, you can configure BGP to perform MD5 authentication when establishing a TCP connection. The two parties must have the same password configured to establish TCP connections. BGP MD5 authentication is not for BGP packets, but for TCP connections. If the authentication fails, no TCP connection can be established.

Follow these steps to enable MD5 authentication for TCP connections:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable MD5 authentication when establishing a TCP connection to the peer/peer group	**peer** { *group-name* | *ip-address* } **password** { **cipher** | **simple** } *password*	Optional Not enabled by default

Configuring BGP Load Balancing

If multiple paths to a destination exist, you can configure load balancing over such paths to improve link utilization.

Follow these steps to configure BGP load balancing:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the maximum number of BGP routes for load balancing	**balance** *number*	Optional Load balancing is not enabled by default.

Forbiding Session Establishment with a Peer or Peer Group

Follow these steps to forbid session establishment with a peer or peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Forbid session establishment with a peer or peer group	**peer** { *group-name* | *ip-address* } **ignore**	Optional Not forbidden by default

Configuring a Large Scale BGP Network

In a large-scale BGP network, configuration and maintenance become difficult due to large numbers of BGP peers. To facilitate configuration in this case, you can configure peer group, community, route reflector or confederation as needed.

Configuration Prerequisites

Peering nodes are accessible to each other at the network layer.

Configuring BGP Peer Groups

A peer group is a group of peers with the same route selection policy.

In a large scale network, many peers may use the same route selection policy. You can configure a peer group and add these peers into this group. In this way, peers can share the same policy as the peer group. When the policy of the group is modified, the modification also applies to peers in it, thus simplifying configuration.

A peer group is an iBGP peer group if peers in it belong to the same AS, and is an eBGP peer group if peers in it belong to different ASs.

Note that:

If a peer group has peers added, you cannot remove its AS number using the **undo** form of the command or change its AS number.

Configure an iBGP peer group

After you create an iBGP peer group and then add a peer into it, the system creates the peer in BGP view and specifies the local AS number for the peer.

Follow these steps to configure an iBGP peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Create an iBGP peer group	**group** *group-name* [**internal**]	Required
Add a peer into the iBGP peer group	**peer** *ip-address* **group** *group-name*	Required

Configure an eBGP peer group

If peers in an eBGP group belong to the same external AS, the eBGP peer group is a pure eBGP peer group; if not, it is a mixed eBGP peer group.

There are two approaches for configuring an eBGP peer group:

- Create the eBGP peer group, specify its AS number, and add peers into it.
- Create the eBGP peer group, and add a created peer into it or add a peer with the AS number specified

Follow these steps to configure an eBGP peer group using the first approach:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Create an eBGP peer group	**group** *group-name* **external**	Required
Specify the AS number for the group	**peer** *group-name* **as-number** *as-number*	Required
Add a peer into the group	**peer** *ip-address* **group** *group-name*	Required

Follow these steps to configure an eBGP peer group using the second approach:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Create an eBGP peer group	**group** *group-name* **external**	Required
Add a peer into the group	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	Required

Configuring BGP community can also help simplify routing policy management, and a community has a much larger management scope than a peer group by controlling routing policies of multiple BGP routers.

To guarantee the connectivity between iBGP peers, you need to make them fully meshed. But it becomes unpractical when there are large numbers of iBGP peers. Configuring route reflectors or confederation can solve it. In a large-scale AS, both of them can be used.

Configuring BGP Community

A BGP community is a group of destinations with the same characteristics. It has no geographical boundaries and is independent of ASs.

You can configure a route policy to define which destinations belong to a BGP community and then advertise the community attribute to a peer/peer group.

You can apply a route policy to filter routes advertised to/received from a peer/peer group according to the community attribute. This way helps simplify policy configuration and management.

For how to configure a route policy, refer to *Route Policy Configuration* in the *IP Routing Volume*.

Follow these steps to configure BGP community:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter BGP view		**bgp** *as-number*	—
Advertise the community attribute to a peer/peer group	Advertise the community attribute to a peer/peer group	**peer** { *group-name* | *ip-address* } **advertise-community**	Required Not configured by default.
	Advertise the extended community attribute to a peer/peer group	**peer** { *group-name* | *ip-address* } **advertise-ext-community**	
Apply a routing policy to routes advertised to a peer/peer group		**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	Required Not configured by default.

Configuring a BGP Route Reflector

If an AS has many BGP routers, you can configure them as a cluster and configure one of them as a route reflector and others as clients to reduce iBGP connections.

To enhance network reliability and prevent single point failures, you can specify multiple route reflectors for a cluster. The route reflectors in the cluster must have the same cluster ID specified to avoid routing loops.

Follow these steps to configure a BGP route reflector:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the router as a route reflector and specify a peer/peer group as its client	**peer** { *group-name* | *ip-address* } **reflect-client**	Required Not configured by default.
Enable route reflection between clients	**reflect between-clients**	Optional Enabled by default
Configure the cluster ID of the route reflector	**reflector cluster-id** *cluster-id*	Optional By default, a route reflector uses its router ID as the cluster ID.

Caution

- In general, it is not required to make clients of a route reflector fully meshed. The route reflector forwards routing information between clients. If clients are fully meshed, you can disable route reflection between clients to reduce routing costs.
- In general, a cluster has only one route reflector, and the router ID is used to identify the cluster. You can configure multiple route reflectors to improve network stability. In this case, you need to specify the same cluster ID for these route reflectors using the **reflector cluster-id** command to avoid routing loops.

Configuring a BGP Confederation

Configuring a BGP confederation is another way for reducing iBGP connections in an AS.

A confederation contains sub ASs. In each sub AS, iBGP peers are fully meshed. Between sub ASs, eBGP connections are established.

If routers not compliant with RFC 3065 exist in the confederation, you can use the **confederation nonstandard** command to make the local router compatible with these routers.

Configure a BGP confederation

After you split an AS into multiple sub ASs, you can configure a router in a sub AS in the following way:

1) Enable BGP and specify the AS number of the router.
2) Specify the confederation ID. From an outsider's perspective, the sub ASs of the confederation is a single AS, which is identified by the confederation ID.
3) If the router needs to establish eBGP connections to other sub ASs, you need to specify the peering sub ASs in the confederation.

A confederation contains 32 sub ASs at most. The AS number of a sub AS is effective only in the confederation.

Follow these steps to configure a BGP confederation:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure a confederation ID	**confederation id** *as-number*	Required Not configured by default.
Specify peering sub ASs in the confederation	**confederation peer-as** *as-number-list*	Required Not configured by default.

Configure confederation compatibility

If some other routers in the confederation do not comply with RFC 3065, you need to enable confederation compatibility to allow the router to work with those routers.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable compatibility with routers not compliant with RFC 3065 in the confederation	**confederation nonstandard**	Optional Not enabled by default

Configuring BGP GR

Perform the following configuration on the GR Restarter and GR Helper respectively.

Note

A device can act as a GR Restarter and GR Helper at the same time.

Follow these steps to configure BGP GR:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable BGP, and enter its view	**bgp** *as-number*	—
Enable GR Capability for BGP	**graceful-restart**	Required Disabled by default
Configure the maximum time allowed for the peer to reestablish a BGP session	**graceful-restart timer restart** *timer*	Optional 150 seconds by default
Configure the maximum time to wait for the End-of-RIB marker	**graceful-restart timer wait-for-rib** *timer*	Optional 180 seconds by default

Note

- In general, the maximum time allowed for the peer (the GR restarter) to reestablish a BGP session should be less than the Holdtime carried in the OPEN message.
- The End-Of-RIB (End of Routing-Information-Base) indicates the end of route updates.

Enabling Trap

After Trap is enabled for BGP, BGP generates Level-4 traps to report important events of it. The generated traps are sent to the Information Center of the device. The output rules of the traps, namely, whether to output the traps and the output direction, are determined according to the Information Center configuration. (For Information Center configuration, refer to "Information Center Configuration" in the *System Volume*.)

Follow these steps to enable Trap:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable Trap for BGP	**snmp-agent trap enable bgp**	Optional Enabled by default

Enabling Logging of Peer State Changes

Follow these steps to enable the logging of peer state changes:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter BGP view		**bgp** *as-number*	—
Enable the logging of peer state changes	globally	**log-peer-change**	Optional Enabled by default
	for a peer or peer group	**peer** { *group-name* | *ip-address* } **log-change**	Optional Enabled by default

Displaying and Maintaining BGP

Displaying BGP

To do…	Use the command…	Remarks
Display peer group information	**display bgp group** [*group-name*]	Available in any view
Display advertised BGP routing information	**display bgp network**	
Display AS path information	**display bgp paths** [*as-regular-expression*]	
Display BGP peer/peer group information	**display bgp peer** [*ip-address* { **log-info** | **verbose** } | *group-name* **log-info** | **verbose**]	
Display BGP routing information	**display bgp routing-table** [*ip-address* [{ *mask* | *mask-length* } [**longer-prefixes**]]]	
Display routing information matching the AS path ACL	**display bgp routing-table as-path-acl** *as-path-acl-number*	
Display BGP CIDR routing information	**display bgp routing-table cidr**	
Display BGP routing information matching the specified BGP community	**display bgp routing-table community** [*aa:nn*&<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**] * [**whole-match**]	
Display routing information matching a BGP community list	**display bgp routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	
Display BGP dampened routing information	**display bgp routing-table dampened**	
Display BGP dampening parameter information	**display bgp routing-table dampening parameter**	
Display BGP routing information originating from different ASs	**display bgp routing-table different-origin-as**	
Display BGP routing flap statistics	**display bgp routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *ip-address* [{ *mask* | *mask-length* } [**longer-match**]]]	
Display labeled BGP routing information	**display bgp routing-table label**	
Display routing information to or from a peer	**display bgp routing-table peer** *ip-address* { **advertised-routes** | **received-routes** } [*network-address* [*mask* | *mask-length*] | **statistic**]	
Display routing information matching a regular expression	**display bgp routing-table regular-expression** *as-regular-expression*	
Display BGP routing statistics	**display bgp routing-table statistic**	

Resetting BGP Connections

To do…	Use the command…	Remarks
Reset all BGP connections	**reset bgp all**	Available in user view
Reset the BGP connections to an AS	**reset bgp** *as-number*	
Reset the BGP connection to a peer	**reset bgp** *ip-address* [**flap-info**]	
Reset all eBGP connections	**reset bgp external**	
Reset the BGP connections to a peer group	**reset bgp group** *group-name*	
Reset all iBGP connections	**reset bgp internal**	
Reset all IPv4 unicast BGP connections	**reset bgp ipv4 all**	

Clearing BGP Information

To do…	Use the command…	Remarks
Clear dampened MBGP routing information and release suppressed routes	**reset bgp dampening** [*ip-address* [*mask* | *mask-length*]]	Available in user view
Clear route flap information	**reset bgp flap-info** [*ip-address* [*mask-length* | *mask*] | **as-path-acl** *as-path-acl-number* | **regexp** *as-path-regular-expression*]	

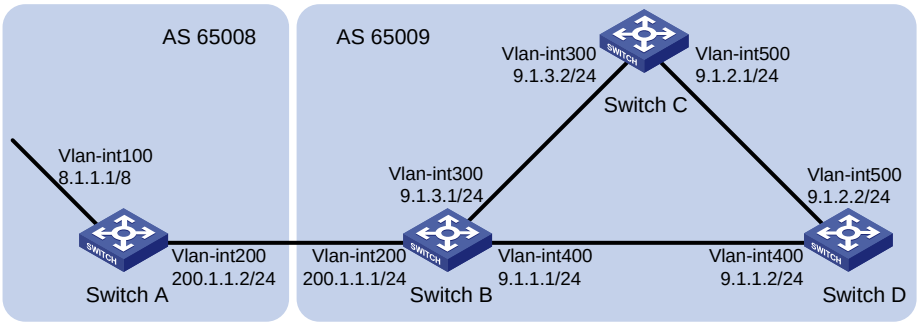
BGP Configuration Examples

BGP Basic Configuration

Network requirements

In the following figure are all BGP switches. Between Switch A and Switch B is an eBGP connection. iBGP speakers Switch B, Switch C and Switch D are fully meshed.

Figure 1-20 Network diagram for BGP basic configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure iBGP connections

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 9.1.1.2 as-number 65009
[SwitchB-bgp] peer 9.1.3.2 as-number 65009
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 9.1.3.1 as-number 65009
[SwitchC-bgp] peer 9.1.2.2 as-number 65009
[SwitchC-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 65009
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] peer 9.1.1.1 as-number 65009
[SwitchD-bgp] peer 9.1.2.1 as-number 65009
[SwitchD-bgp] quit
```

3) Configure the eBGP connection

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 200.1.1.1 as-number 65009
```

Inject network 8.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp] network 8.0.0.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] peer 200.1.1.2 as-number 65008
[SwitchB-bgp] quit
```

Display BGP peer information on Switch B.

```
[SwitchB] display bgp peer

 BGP local router ID : 2.2.2.2
 Local AS number : 65009
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  9.1.1.2     4 65009       56       56     0       0 00:40:54 Established
  9.1.3.2     4 65009       49       62     0       0 00:44:58 Established
```

```
  200.1.1.2   4 65008       49       65     0       1 00:44:03 Established
```

You can find Switch B has established BGP connections to other switches.

Display BGP routing table information on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0            0.0.0.0        0                    0        i
```

Display BGP routing table information on Switch B.

```
[SwitchB] display bgp routing-table
 Total Number of Routes: 1

 BGP Local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0          200.1.1.2     0                     0       65008i
```

Display the BGP routing table on Switch C.

```
[SwitchC] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED       LocPrf    PrefVal Path/Ogn

   i  8.0.0.0          200.1.1.2     0         100       0       65008i
```

Note

From the above outputs, you can find Switch A has learned no route to AS65009, and Switch C has learned network 8.0.0.0 but the next hop 200.1.1.2 is unreachable, so the route is invalid.

---

4) Redistribute direct routes

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] import-route direct
```

Display BGP routing table information on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 7

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0           0.0.0.0       0                  0       i
 *>   9.1.1.0/24        200.1.1.1     0                  0       65009?
 *>   9.1.1.2/32        200.1.1.1     0                  0       65009?
 *>   9.1.3.0/24        200.1.1.1     0                  0       65009?
 *>   9.1.3.2/32        200.1.1.1     0                  0       65009?
 *    200.1.1.0         200.1.1.1     0                  0       65009?
 *    200.1.1.2/32      200.1.1.1     0                  0       65009?
```

Display BGP routing table information on Switch C.

```
[SwitchC] display bgp routing-table

 Total Number of Routes: 7

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop       MED        LocPrf    PrefVal Path/Ogn

 *>i  8.0.0.0           200.1.1.2     0          100       0       65008i
 *>i  9.1.1.0/24        9.1.3.1       0          100       0       ?
 *>i  9.1.1.2/32        9.1.3.1       0          100       0       ?
 * i  9.1.3.0/24        9.1.3.1       0          100       0       ?
 * i  9.1.3.2/32        9.1.3.1       0          100       0       ?
 *>i  200.1.1.0         9.1.3.1       0          100       0       ?
 *>i  200.1.1.2/32      9.1.3.1       0          100       0       ?
```

You can find the route 8.0.0.0 becomes valid with the next hop being Switch A.

Ping 8.1.1.1 on Switch C.

```
[SwitchC] ping 8.1.1.1
  PING 8.1.1.1: 56  data bytes, press CTRL_C to break
    Reply from 8.1.1.1: bytes=56 Sequence=1 ttl=254 time=31 ms
    Reply from 8.1.1.1: bytes=56 Sequence=2 ttl=254 time=47 ms
    Reply from 8.1.1.1: bytes=56 Sequence=3 ttl=254 time=31 ms
```

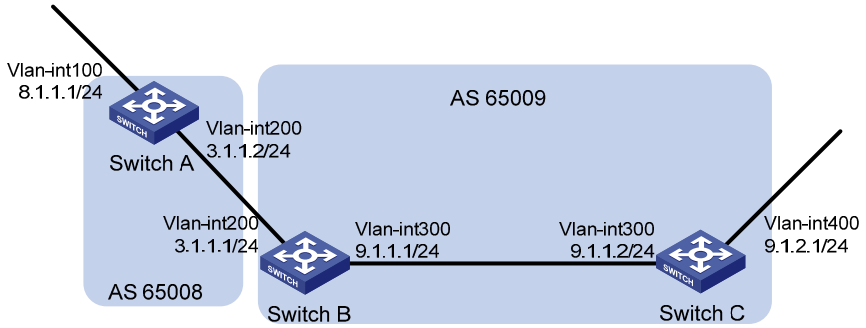
```
  Reply from 8.1.1.1: bytes=56 Sequence=4 ttl=254 time=16 ms
  Reply from 8.1.1.1: bytes=56 Sequence=5 ttl=254 time=31 ms

--- 8.1.1.1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 16/31/47 ms
```

BGP and IGP Synchronization Configuration

Network requirements

As shown below, OSPF is used as the IGP protocol in AS65009, where Switch C is a non-BGP switch. Between Switch A and Switch B is an eBGP connection.

Figure 1-21 Network diagram for BGP and IGP synchronization



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure OSPF (omitted)
3) Configure the eBGP connection

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 3.1.1.1 as-number 65009
```

Inject network 8.1.1.0/24 to the BGP routing table.

```
[SwitchA-bgp] network 8.1.1.0 24
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] peer 3.1.1.2 as-number 65008
[SwitchB-bgp] quit
```

4) Configure BGP and IGP synchronization

Configure BGP to redistribute routes from OSPF on Switch B.

```
[SwitchB] bgp 65009
```

```
[SwitchB-bgp] import-route ospf 1
[SwitchB-bgp] quit
```

Display routing table information on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.1.1.0/24        0.0.0.0         0                     0       i
 *>   9.1.1.0/24        3.1.1.1         0                     0       65009?
```

Configure OSPF to redistribute routes from BGP on Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

Display routing table information on Switch C.

```
<SwitchC> display ip routing-table
Routing Tables: Public
         Destinations : 7        Routes : 7

Destination/Mask    Proto  Pre  Cost       NextHop         Interface

8.1.1.0/24          O_ASE  150  1          9.1.1.1         Vlan300
9.1.1.0/24          Direct 0    0          9.1.1.2         Vlan300
9.1.1.2/32          Direct 0    0          127.0.0.1       InLoop0
9.1.2.0/24          Direct 0    0          9.1.2.1         Vlan400
9.1.2.1/32          Direct 0    0          127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0          127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0          127.0.0.1       InLoop0
```

5) Configure route automatic summarization

Configure route automatic summarization on Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] summary automatic
```

Display BGP routing table information on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
```

```
                Origin : i - IGP, e - EGP, ? - incomplete
     Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>  8.1.1.0/24        0.0.0.0        0                    0       i
 *>  9.0.0.0           3.1.1.1        0                    0       65009?
```

Use ping for verification.

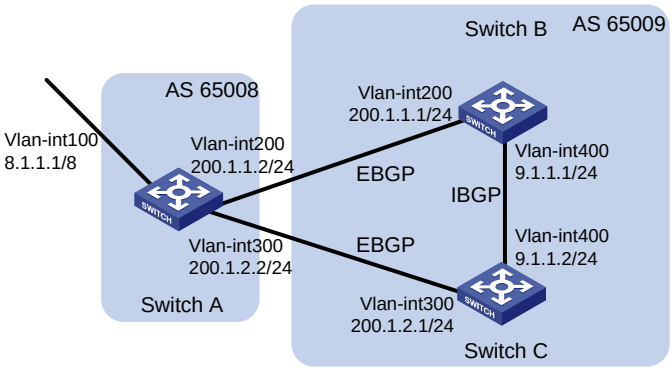
```
[SwitchA] ping -a 8.1.1.1 9.1.2.1
  PING 9.1.2.1: 56  data bytes, press CTRL_C to break
    Reply from 9.1.2.1: bytes=56 Sequence=1 ttl=254 time=15 ms
    Reply from 9.1.2.1: bytes=56 Sequence=2 ttl=254 time=31 ms
    Reply from 9.1.2.1: bytes=56 Sequence=3 ttl=254 time=47 ms
    Reply from 9.1.2.1: bytes=56 Sequence=4 ttl=254 time=46 ms
    Reply from 9.1.2.1: bytes=56 Sequence=5 ttl=254 time=47 ms

  --- 9.1.2.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 15/37/47 ms
```

BGP Load Balancing Configuration

Network requirements

As shown in the following figure, all the switches run BGP. Switch A resides in AS 65008, Switch B and Switch C in AS 65009. Between Switch A and Switch B, Switch A and Switch C are eBGP connections, and between Switch B and Switch C is an iBGP connection. Two routes are configured on Switch A for load balancing.

Figure 1-22 Network diagram for BGP load balancing configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure BGP connections

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
```

```
[SwitchA-bgp] peer 200.1.1.1 as-number 65009
[SwitchA-bgp] peer 200.1.2.1 as-number 65009
```

Inject route 8.0.0.0/8 to BGP routing table.

```
[SwitchA-bgp] network 8.0.0.0 255.0.0.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 200.1.1.2 as-number 65008
[SwitchB-bgp] peer 9.1.1.2 as-number 65009
[SwitchB-bgp] network 9.1.1.0 255.255.255.0
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 200.1.2.2 as-number 65008
[SwitchC-bgp] peer 9.1.1.1 as-number 65009
[SwitchC-bgp] network 9.1.1.0 255.255.255.0
[SwitchC-bgp] quit
```

Display the routing table on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0           0.0.0.0       0                   0       i
 *>   9.1.1.0/24        200.1.1.1     0                   0       65009i
 *                      200.1.2.1     0                   0       65009i
```

From the above output, you can find two valid routes to destination 9.1.1.0/24 are available: the route with next hop 200.1.1.1 is marked with a greater-than sign (>), indicating it is the best route; the route with next hop 200.1.2.1 is marked with an asterisk (*), indicating it is a valid route, but not the best.

3) Configure loading balancing

Configure Switch A.

```
[SwitchA] bgp 65008
[SwitchA-bgp] balance 2
[SwitchA-bgp] quit
```

Display the routing table on Switch A.

```
[SwitchA] display bgp routing-table

 Total Number of Routes: 3

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   8.0.0.0            0.0.0.0        0                    0       i
 *>   9.1.1.0/24         200.1.1.1      0                    0       65009i
 *>                      200.1.2.1      0                    0       65009i
```

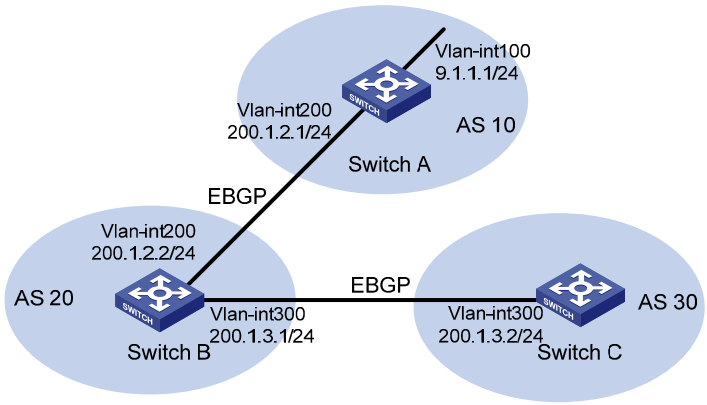
From the above output, you can find two valid routes to the destination 9.1.1.0/24 are available, and both of them are marked with a greater-than sign (>), indicating they are the best routes.

BGP Community Configuration

Network requirements

As shown in the following figure, Switch B establishes eBGP connections with Switch A and C. Configure No_Export community attribute on Switch A to make routes from AS 10 not advertised by AS 20 to any other AS.

Figure 1-23 Network diagram for BGP community configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure eBGP

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 10
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 200.1.2.2 as-number 20
[SwitchA-bgp] network 9.1.1.0 255.255.255.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 20
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 200.1.2.1 as-number 10
[SwitchB-bgp] peer 200.1.3.2 as-number 30
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 30
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 200.1.3.1 as-number 20
[SwitchC-bgp] quit
```

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table 9.1.1.0

 BGP local router ID : 2.2.2.2
 Local AS number : 20
 Paths:   1 available, 1 best

BGP routing table entry information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Original nexthop: 200.1.2.1
 AS-path         : 10
 Origin          : igp
 Attribute value : MED 0, pref-val 0, pre 255
 State           : valid, external, best,
 Advertised to such 1 peers:
    200.1.3.2
```

Switch B advertised routes to Switch C in AS30.

Display the routing table on Switch C.

```
[SwitchC] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 3.3.3.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   9.1.1.0/24        200.1.3.1        0                    0       20 10i
```

Switch C learned route 9.1.1.0/24 from Switch B.

3) Configure BGP community

Configure a routing policy.

```
[SwitchA] route-policy comm_policy permit node 0
```

```
[SwitchA-route-policy] apply community no-export
[SwitchA-route-policy] quit
```

Apply the routing policy.

```
[SwitchA] bgp 10
[SwitchA-bgp] peer 200.1.2.2 route-policy comm_policy export
[SwitchA-bgp] peer 200.1.2.2 advertise-community
```

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table 9.1.1.0
 BGP local router ID : 2.2.2.2
 Local AS number : 20
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 200.1.2.1 (1.1.1.1)
 Original nexthop: 200.1.2.1
 Community       : No-Export
 AS-path         : 10
 Origin          : igp
 Attribute value : MED 0, pref-val 0, pre 255
 State           : valid, external, best,
 Not advertised to any peers yet
```

The route 9.1.1.0/24 is not available in the routing table of Switch C.

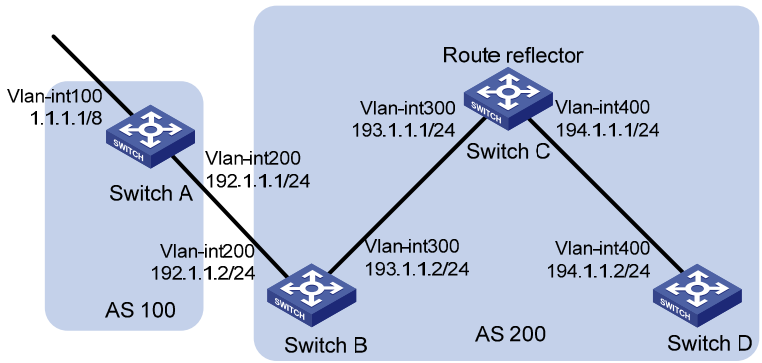
BGP Route Reflector Configuration

Network requirements

In the following figure, all switches run BGP.

- Between Switch A and Switch B is an eBGP connection, between Switch C and Switch B, and between Switch C and Switch D are iBGP connections.
- Switch C is a route reflector with clients Switch B and D.
- Switch D can learn route 1.0.0.0/8 from Switch C.

Figure 1-24 Network diagram for BGP route reflector configuration



Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure BGP connections

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 192.1.1.2 as-number 200
```

Inject network 1.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp] network 1.0.0.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 192.1.1.1 as-number 100
[SwitchB-bgp] peer 193.1.1.1 as-number 200
[SwitchB-bgp] peer 193.1.1.1 next-hop-local
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 193.1.1.2 as-number 200
[SwitchC-bgp] peer 194.1.1.2 as-number 200
[SwitchC-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] peer 194.1.1.1 as-number 200
[SwitchD-bgp] quit
```

3) Configure the route reflector

Configure Switch C.

```
[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.2 reflect-client
[SwitchC-bgp] peer 194.1.1.2 reflect-client
[SwitchC-bgp] quit
```

4) Verify the above configuration

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table

 Total Number of Routes: 1
```

```
BGP Local router ID is 200.1.2.2
Status codes: * - valid, > - best, d - damped,
              h - history,  i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
    Network         NextHop         MED        LocPrf     PrefVal Path/Ogn

*>  1.0.0.0         192.1.1.1       0                     0       100i
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 200.1.2.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network         NextHop         MED        LocPrf     PrefVal Path/Ogn

   i 1.0.0.0         193.1.1.2       0          100        0       100i
```

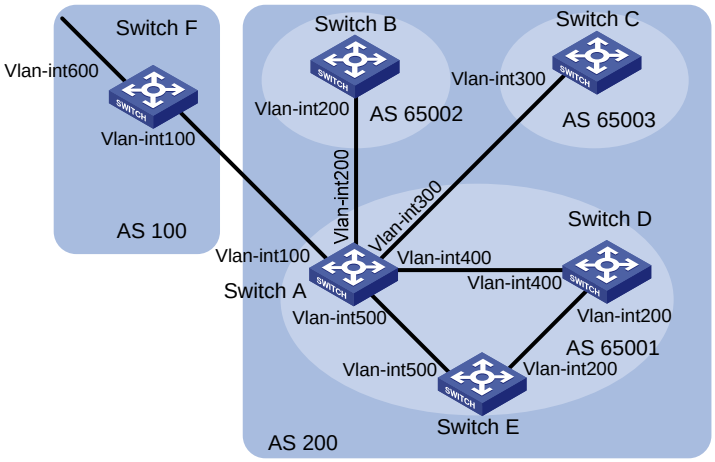
Switch D learned route 1.0.0.0/8 from Switch C.

BGP Confederation Configuration

Network requirements

As shown in the following figure, to reduce iBGP connections in AS 200, split it into three sub-ASs, AS65001, AS65002 and AS65003. Switches in AS65001 are fully meshed.

Figure 1-25 Network diagram for BGP confederation configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	200.1.1.1/24	Switch D	Vlan-int200	10.1.5.1/24
	Vlan-int200	10.1.1.1/24		Vlan-int400	10.1.3.2/24
	Vlan-int300	10.1.2.1/24	Switch E	Vlan-int200	10.1.5.2/24
	Vlan-int400	10.1.3.1/24		Vlan-int500	10.1.4.2/24
	Vlan-int500	10.1.4.1/24	Switch F	Vlan-int100	200.1.1.2/24
Switch B	Vlan-int200	10.1.1.2/24		Vlan-int600	9.1.1.1/24
Switch C	Vlan-int300	10.1.2.2/24			

Configuration procedure

1) Configure IP addresses for interfaces (omitted)
2) Configure BGP confederation

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65001
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] confederation id 200
[SwitchA-bgp] confederation peer-as 65002 65003
[SwitchA-bgp] peer 10.1.1.2 as-number 65002
[SwitchA-bgp] peer 10.1.1.2 next-hop-local
[SwitchA-bgp] peer 10.1.2.2 as-number 65003
[SwitchA-bgp] peer 10.1.2.2 next-hop-local
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65002
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] confederation id 200
[SwitchB-bgp] confederation peer-as 65001 65003
[SwitchB-bgp] peer 10.1.1.1 as-number 65001
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65003
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] confederation id 200
[SwitchC-bgp] confederation peer-as 65001 65002
[SwitchC-bgp] peer 10.1.2.1 as-number 65001
[SwitchC-bgp] quit
```

3) Configure iBGP connections in AS65001.

Configure Switch A.

```
[SwitchA] bgp 65001
[SwitchA-bgp] peer 10.1.3.2 as-number 65001
[SwitchA-bgp] peer 10.1.3.2 next-hop-local
[SwitchA-bgp] peer 10.1.4.2 as-number 65001
[SwitchA-bgp] peer 10.1.4.2 next-hop-local
[SwitchA-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 65001
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] confederation id 200
[SwitchD-bgp] peer 10.1.3.1 as-number 65001
[SwitchD-bgp] peer 10.1.5.2 as-number 65001
```

```
[SwitchD-bgp] quit
```

Configure Switch E.

```
<SwitchE> system-view
[SwitchE] bgp 65001
[SwitchE-bgp] router-id 5.5.5.5
[SwitchE-bgp] confederation id 200
[SwitchE-bgp] peer 10.1.4.1 as-number 65001
[SwitchE-bgp] peer 10.1.5.1 as-number 65001
[SwitchE-bgp] quit
```

4) Configure the eBGP connection between AS100 and AS200.

Configure Switch A.

```
[SwitchA] bgp 65001
[SwitchA-bgp] peer 200.1.1.2 as-number 100
[SwitchA-bgp] quit
```

Configure Switch F.

```
<SwitchF> system-view
[SwitchF] bgp 100
[SwitchF-bgp] router-id 6.6.6.6
[SwitchF-bgp] peer 200.1.1.1 as-number 200
[SwitchF-bgp] network 9.1.1.0 255.255.255.0
[SwitchF-bgp] quit
```

5) Verify above configuration

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop       MED      LocPrf   PrefVal Path/Ogn

 *>i  9.1.1.0/24        10.1.1.1       0         100       0       (65001) 100i
[SwitchB] display bgp routing-table 9.1.1.0

 BGP local router ID : 2.2.2.2
 Local AS number : 65002
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 10.1.1.1 (1.1.1.1)
 Relay Nexthop   : 0.0.0.0
 Original nexthop: 10.1.1.1
 AS-path         : (65001) 100
 Origin          : igp
```

```
 Attribute value : MED 0, localpref 100, pref-val 0, pre 255
 State           : valid, external-confed, best,
 Not advertised to any peers yet
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table

 Total Number of Routes: 1

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>i  9.1.1.0/24        10.1.3.1      0          100      0       100i
[SwitchD] display bgp routing-table 9.1.1.0

 BGP local router ID : 4.4.4.4
 Local AS number : 65001
 Paths:   1 available, 1 best

 BGP routing table entry information of 9.1.1.0/24:
 From            : 10.1.3.1 (1.1.1.1)
 Relay Nexthop   : 0.0.0.0
 Original nexthop: 10.1.3.1
 AS-path         : 100
 Origin          : igp
 Attribute value : MED 0, localpref 100, pref-val 0, pre 255
 State           : valid, internal, best,
 Not advertised to any peers yet
```

The output information shows that:

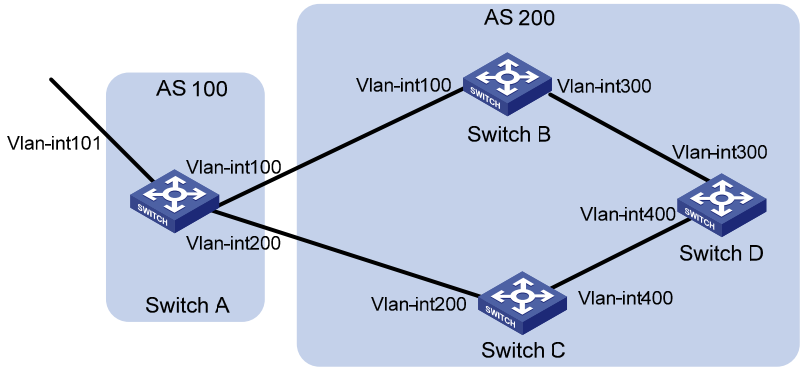
- Switch F can send route information to Switch B and Switch C through the confederation by establishing only an eBGP connection with Switch A.
- Switch B and Switch D are in the same confederation, but belong to different sub ASs. They obtain external route information from Switch A and generate the same BGP route entries; it seems like that they reside in the same AS although they have no direct connection in between.

BGP Path Selection Configuration

Network requirements

- In the figure below, all switches run BGP. Between Switch A and Switch B, and between Switch A and Switch C are eBGP connections. Between Switch B and Switch D, and between Switch D and Switch C are iBGP connections.
- OSPF is the IGP protocol in AS 200.
- Configure routing policies, making Switch D use the route 1.0.0.0/8 from Switch C as the optimal.

Figure 1-26 Network diagram for BGP path selection configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int101	1.0.0.0/8	Switch D	Vlan-int400	195.1.1.1/24
	Vlan-int100	192.1.1.1/24		Vlan-int300	194.1.1.1/24
	Vlan-int200	193.1.1.1/24	Switch C	Vlan-int400	195.1.1.2/24
Switch B	Vlan-int100	192.1.1.2/24		Vlan-int200	193.1.1.2/24
	Vlan-int300	194.1.1.2/24			

Configuration procedure

1) Configure IP addresses for interfaces (omitted).
2) Configure OSPF on Switch B, C, and D.

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 193.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

3) Configure BGP connections

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp] peer 192.1.1.2 as-number 200
[SwitchA-bgp] peer 193.1.1.2 as-number 200
```

Inject network 1.0.0.0/8 to the BGP routing table on Switch A.

```
[SwitchA-bgp] network 1.0.0.0 8
[SwitchA-bgp] quit
```

Configure Switch B.

```
[SwitchB] bgp 200
[SwitchB-bgp] peer 192.1.1.1 as-number 100
[SwitchB-bgp] peer 194.1.1.1 as-number 200
[SwitchB-bgp] quit
```

Configure Switch C.

```
[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.1 as-number 100
[SwitchC-bgp] peer 195.1.1.1 as-number 200
[SwitchC-bgp] quit
```

Configure Switch D.

```
[SwitchD] bgp 200
[SwitchD-bgp] peer 194.1.1.2 as-number 200
[SwitchD-bgp] peer 195.1.1.2 as-number 200
[SwitchD-bgp] quit
```

4) Configure attributes for route 1.0.0.0/8, making Switch D give priority to the route learned from Switch C.

- Configure a higher MED value for the route 1.0.0.0/8 advertised from Switch A to peer 192.1.1.2.

Define an ACL numbered 2000 to permit route 1.0.0.0/8.

```
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[SwitchA-acl-basic-2000] quit
```

Define two routing policies, apply_med_50, which sets the MED for route 1.0.0.0/8 to 50, and apply_med_100, which sets the MED for route 1.0.0.0/8 to 100.

```
[SwitchA] route-policy apply_med_50 permit node 10
[SwitchA-route-policy] if-match acl 2000
[SwitchA-route-policy] apply cost 50
[SwitchA-route-policy] quit
[SwitchA] route-policy apply_med_100 permit node 10
[SwitchA-route-policy] if-match acl 2000
[SwitchA-route-policy] apply cost 100
[SwitchA-route-policy] quit
```

Apply routing policy apply_med_50 to the route advertised to peer 193.1.1.2 (Switch C), and apply_med_100 to the route advertised to peer 192.1.1.2 (Switch B).

```
[SwitchA] bgp 100
[SwitchA-bgp] peer 193.1.1.2 route-policy apply_med_50 export
[SwitchA-bgp] peer 192.1.1.2 route-policy apply_med_100 export
```

```
[SwitchA-bgp] quit
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 194.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network         NextHop        MED        LocPrf     PrefVal Path/Ogn

 *>i 1.0.0.0         193.1.1.1      50         100        0       100i
 * i                 192.1.1.1      100        100        0       100i
```

You can find route 1.0.0.0/8 is the optimal.

- Configure different local preferences on Switch B and C for route 1.0.0.0/8, making Switch D give priority to the route from Switch C.

Define an ACL numbered 2000 on Router C, permitting route 1.0.0.0/8.

```
[SwitchC] acl number 2000
[SwitchC-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[SwitchC-acl-basic-2000] quit
```

Configure a routing policy named localpref on Switch C, setting the local preference of route 1.0.0.0/8 to 200 (the default is 100).

```
[SwitchC] route-policy localpref permit node 10
[SwitchC-route-policy] if-match acl 2000
[SwitchC-route-policy] apply local-preference 200
[SwitchC-route-policy] quit
```

Apply routing policy localpref to routes from peer 193.1.1.1.

```
[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.1 route-policy localpref import
[SwitchC-bgp] quit
```

Display the routing table on Switch D.

```
[SwitchD] display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 194.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network          NextHop       MED        LocPrf     PrefVal Path/Ogn

 *>i 1.0.0.0          193.1.1.1     0          200        0       100i
 * i                  192.1.1.1     0          100        0       100i
```

You can find route 1.0.0.0/8 from Switch D to Switch C is the optimal.

Troubleshooting BGP

No BGP Peer Relationship Established

Symptom

Display BGP peer information using the **display bgp peer** command. The state of the connection to a peer cannot become established.

Analysis

To become BGP peers, any two routers need to establish a TCP session using port 179 and exchange open messages successfully.

Solution

1) Use the **display current-configuration** command to verify the peer's AS number.
2) Use the **display bgp peer** command to verify the peer's IP address.
3) If the loopback interface is used, check whether the **peer connect-interface** command is configured.
4) If the peer is a non-direct eBGP peer, check whether the **peer ebgp-max-hop** command is configured.
5) Check whether a route to the peer is available in the routing table.
6) Use the **ping** command to check connectivity.
7) Use the **display tcp status** command to check the TCP connection.
8) Check whether an ACL disabling TCP port 179 is configured.

Table of Contents

1 IPv6 Static Routing Configuration

When configuring IPv6 Static Routing, go to these sections for information you are interested in:

Note

The term “router” in this document refers to either a router in a generic sense or a Layer 3 switch running routing protocols.

Introduction to IPv6 Static Routing

Static routes are special routes that are manually configured by network administrators. They work well in simple networks. Configuring and using them properly can improve the performance of networks and guarantee enough bandwidth for important applications.

However, static routes also have shortcomings: any topology changes could result in unavailable routes, requiring the network administrator to manually configure and modify the static routes.

Features of IPv6 Static Routes

Similar to IPv4 static routes, IPv6 static routes work well in simple IPv6 network environments.

Their major difference lies in the destination and next hop addresses. IPv6 static routes use IPv6 addresses whereas IPv4 static routes use IPv4 addresses. Currently, IPv6 static routes do not support VPN instance.

Default IPv6 Route

The IPv6 static route that has the destination address configured as **::/0** (indicating a prefix length of 0) is the default IPv6 route. If the destination address of an IPv6 packet does not match any entry in the routing table, this default route will be used to forward the packet.

Configuring an IPv6 Static Route

In small IPv6 networks, IPv6 static routes can be used to forward packets. In comparison to dynamic routes, it helps to save network bandwidth.

Configuration prerequisites

- Configuring parameters for the related interfaces
- Configuring link layer attributes for the related interfaces
- Enabling IPv6 packet forwarding
- Ensuring that the neighboring nodes are IPv6 reachable

Configuring an IPv6 Static Route

Follow these steps to configure an IPv6 static route:

To do…	Use the commands…	Remarks
Enter system view	**system-view**	—
Configure an IPv6 static route	**ipv6 route-static** *ipv6-address prefix-length* [*interface-type interface-number*] *nexthop-address* [**preference** *preference-value*]	Required The default preference of IPv6 static routes is 60.

Displaying and Maintaining IPv6 Static Routes

To do…	Use the command…	Remarks
Display IPv6 static route information	**display ipv6 routing-table protocol static** [**inactive** | **verbose**]	Available in any view
Remove all IPv6 static routes	**delete ipv6 static-routes all**	Available in system view

Note

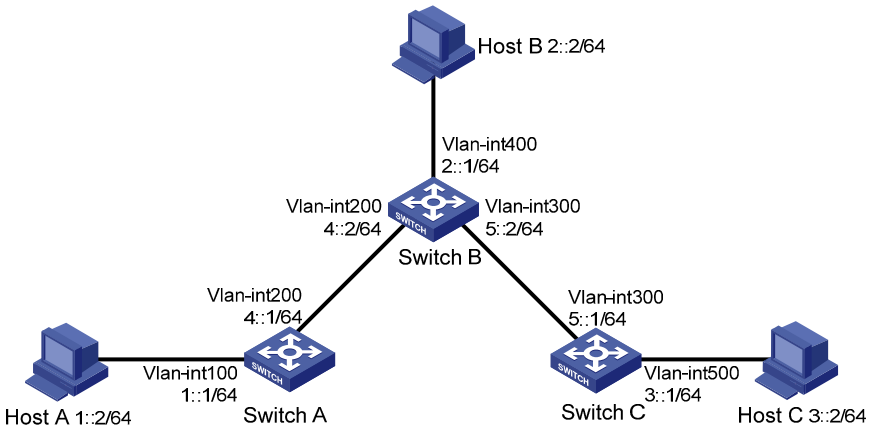
Using the **undo ipv6 route-static** command can delete a single IPv6 static route, while using the **delete ipv6 static-routes all** command deletes all IPv6 static routes including the default route.

IPv6 Static Routing Configuration Example

Network requirements

With IPv6 static routes configured, all hosts and switches can interact with each other.

Figure 1-1 Network diagram for static routes



Configuration procedure

1) Configure the IPv6 addresses of all VLAN interfaces (Omitted)
2) Configure IPv6 static routes.

Configure the default IPv6 static route on SwitchA.

```
<SwitchA> system-view
[SwitchA] ipv6 route-static :: 0 4::2
```

Configure two IPv6 static routes on SwitchB.

```
<SwitchB> system-view
[SwitchB] ipv6 route-static 1:: 64 4::1
[SwitchB] ipv6 route-static 3:: 64 5::1
```

Configure the default IPv6 static route on SwitchC.

```
<SwitchC> system-view
[SwitchC] ipv6 route-static :: 0 5::2
```

3) Configure the IPv6 addresses of hosts and gateways.

Configure the IPv6 addresses of all the hosts based upon the network diagram, configure the default gateway of Host A as 1::1, that of Host B as 2::1, and that of Host C as 3::1.

4) Display configuration information

Display the IPv6 routing table of SwitchA.

```
[SwitchA] display ipv6 routing-table
Routing Table :
         Destinations : 5        Routes : 5

 Destination  : :: /128                          Protocol     : Static
 NextHop      : FE80::510A:0:8D7:1               Preference   : 60
 Interface    : Vlan-interface200                Cost         : 0

 Destination  : ::1/128                          Protocol     : Direct
 NextHop      : ::1                              Preference   : 0
 Interface    : InLoop0                          Cost         : 0
```

```
 Destination  : 1:: /64                                Protocol     : Direct
 NextHop      : 1::1                                   Preference   : 0
 Interface    : Vlan-interface100                      Cost         : 0

 Destination  : 1::1/128                               Protocol     : Direct
 NextHop      : ::1                                    Preference   : 0
 Interface    : InLoop0                                Cost         : 0

 Destination  : FE80::/10                              Protocol     : Direct
 NextHop      : ::                                     Preference   : 0
 Interface    : NULL0                                  Cost         : 0
```

Verify the connectivity with the **ping** command.

```
[SwitchA] ping ipv6 3::1
  PING 3::1 : 56  data bytes, press CTRL_C to break
    Reply from 3::1
    bytes=56 Sequence=1 hop limit=254  time = 63 ms
    Reply from 3::1
    bytes=56 Sequence=2 hop limit=254  time = 62 ms
    Reply from 3::1
    bytes=56 Sequence=3 hop limit=254  time = 62 ms
    Reply from 3::1
    bytes=56 Sequence=4 hop limit=254  time = 63 ms
    Reply from 3::1
    bytes=56 Sequence=5 hop limit=254  time = 63 ms

  --- 3::1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 62/62/63 ms
```

Table of Contents

1 RIPng Configuration

When configuring RIPng, go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch.

Introduction to RIPng

RIP next generation (RIPng) is an extension of RIP-2 for IPv4. Most RIP concepts are applicable in RIPng.

RIPng for IPv6 has the following basic differences from RIP:

- UDP port number: RIPng uses UDP port 521 for sending and receiving routing information.
- Multicast address: RIPng uses FF02:9 as the link-local-router multicast address.
- Destination Prefix: 128-bit destination address prefix.
- Next hop: 128-bit IPv6 address.
- Source address: RIPng uses FE80::/10 as the link-local source address

RIPng Working Mechanism

RIPng is a routing protocol based on the distance vector (D-V) algorithm. RIPng uses UDP packets to exchange routing information through port 521.

RIPng uses a hop count to measure the distance to a destination. The hop count is referred to as metric or cost. The hop count from a router to a directly connected network is 0. The hop count between two directly connected routers is 1. When the hop count is greater than or equal to 16, the destination network or host is unreachable.

By default, the routing update is sent every 30 seconds. If the router receives no routing updates from a neighbor within 180 seconds, the routes learned from the neighbor are considered as unreachable. Within another 240 seconds, if no routing update is received, the router will remove these routes from the routing table.

RIPng supports split horizon and poison reverse to prevent routing loops and route redistribution.

Each RIPng router maintains a routing database, including route entries of all reachable destinations. A route entry contains the following information:

- Destination address: IPv6 address of a host or a network.
- Next hop address: IPv6 address of a neighbor along the path to the destination.
- Egress interface: Outbound interface that forwards IPv6 packets.
- Metric: Cost from the local router to the destination.
- Route time: Time that elapsed since a route entry is last changed. Each time a route entry is modified, the routing time is set to 0.
- Route tag: Identifies the route, used in a routing policy to control routing information. For information about routing policy, refer to *Routing Policy Configuration* in the *IP Routing Volume*.

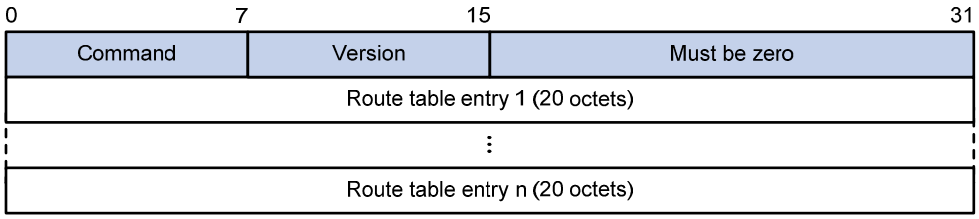
RIPng Packet Format

Basic format

A RIPng packet consists of a header and multiple route table entries (RTEs). The maximum number of RTEs in a packet depends on the IPv6 MTU of the sending interface.

Figure 1-1 shows the packet format of RIPng.

Figure 1-1 RIPng basic packet format



- Command: Type of message. 0x01 indicates Request, 0x02 indicates Response.
- Version: Version of RIPng. It can only be 0x01 currently.
- RTE: Route table entry, 20 bytes for each entry.

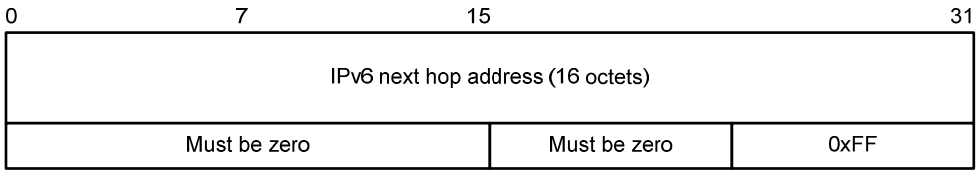
RTE format

There are two types of RTEs in RIPng.

- Next hop RTE: Defines the IPv6 address of a next hop
- IPv6 prefix RTE: Describes the destination IPv6 address, route tag, prefix length and metric in the RIPng routing table.

Figure 1-2 shows the format of the next hop RTE:

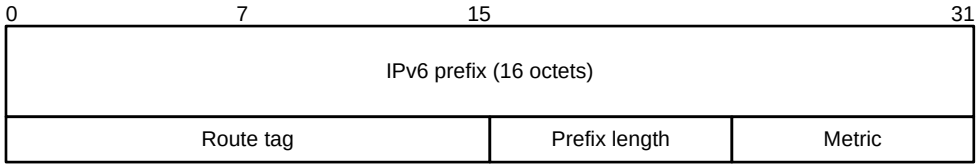
Figure 1-2 Next hop RTE format



IPv6 next hop address is the IPv6 address of the next hop.

Figure 1-3 shows the format of the IPv6 prefix RTE.

Figure 1-3 IPv6 prefix RTE format



- IPv6 prefix: Destination IPv6 address prefix.
- Route tag: Route tag.
- Prefix len: Length of the IPv6 address prefix.
- Metric: Cost of a route.

RIPng Packet Processing Procedure

Request packet

When a RIPng router first starts or needs to update some entries in its routing table, generally a multicast request packet is sent to ask for needed routes from neighbors.

The receiving RIPng router processes RTEs in the request. If there is only one RTE with the IPv6 prefix and prefix length both being 0, and with a metric value of 16, the RIPng router will respond with the entire routing table information in response messages. If there are multiple RTEs in the request message, the RIPng router will examine each RTE, update its metric, and send the requested routing information to the requesting router in the response packet.

Response packet

The response packet containing the local routing table information is generated as:

- A response to a request
- An update periodically
- A trigged update caused by route change

After receiving a response, a router checks the validity of the response before adding the route to its routing table, such as whether the source IPv6 address is the link-local address and whether the port number is correct. The response packet that failed the check will be discarded.

Protocols and Standards

- RFC 2080: RIPng for IPv6
- RFC 2081: RIPng Protocol Applicability Statement

Configuring RIPng Basic Functions

This section presents the information to configure the basic RIPng features.

You need to enable RIPng first before configuring other tasks, but it is not necessary for RIPng related interface configurations, such as assigning an IPv6 address.

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Enable IPv6 packet forwarding.

- Configure an IP address for each interface, and make sure all nodes are reachable to one another.

Configuration Procedure

Follow these steps to configure the basic RIPng functions:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RIPng process and enter RIPng view	**ripng** [*process-id*]	Required Not created by default
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable RIPng on the interface	**ripng** *process-id* **enable**	Required Disabled by default

Note

If RIPng is not enabled on an interface, the interface will not send or receive any RIPng route.

Configuring RIPng Route Control

This section covers the following topics:

- Configuring RIPng Route Summarization
- Advertising a Default Route
- Configuring a RIPng Route Filtering Policy
- Configuring a Priority for RIPng
- Configuring RIPng Route Redistribution

Before the configuration, accomplish the following tasks first:

- Configure an IPv6 address on each interface, and make sure all nodes are reachable to one another.
- Configure RIPng basic functions
- Define an IPv6 ACL before using it for route filtering. Refer to *ACL Configuration* in the *Security Volume* for related information.
- Define an IPv6 address prefix list before using it for route filtering. Refer to *Routing Policy Configuration* in the *IP Routing Volume* for related information.

Configuring an Additional Routing Metric

An additional routing metric can be added to the metric of an inbound or outbound RIP route, namely, the inbound and outbound additional metric.

The outbound additional metric is added to the metric of a sent route. The route's metric in the routing table is not changed.

The inbound additional metric is added to the metric of a received route before the route is added into the routing table, so the route's metric is changed.

Follow these steps to configure an inbound/outbound additional routing metric:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify an inbound routing additional metric	**ripng metricin** *value*	Optional 0 by default
Specify an outbound routing additional metric	**ripng metricout** *value*	Optional 1 by default

Configuring RIPng Route Summarization

Follow these steps to configure RIPng route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Advertise a summary IPv6 prefix	**ripng summary-address** *ipv6-address prefix-length*	Required

Advertising a Default Route

Follow these steps to advertise a default route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Advertise a default route	**ripng default-route** { **only** | **originate** } [**cost** *cost*]	Required Not advertised by default

Note

With this feature enabled, a default route is advertised through the specified interface regardless of whether the default route is available in the local IPv6 routing table.

Configuring a RIPng Route Filtering Policy

You can reference a configured IPv6 ACL or prefix list to filter received/advertised routing information as needed. For filtering outbound routes, you can also specify a routing protocol from which to filter routing information redistributed.

Follow these steps to configure a RIPng route filtering policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Configure a filter policy to filter incoming routes	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	Required By default, RIPng does not filter incoming routing information.
Configure a filter policy to filter outgoing routes	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol* [*process-id*]]	Required By default, RIPng does not filter outgoing routing information.

Configuring a Priority for RIPng

Any routing protocol has its own protocol priority used for optimal route selection. You can set a priority for RIPng manually. The smaller the value is, the higher the priority is.

Follow these steps to configure a RIPng priority:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Configure a RIPng priority	**preference** [**route-policy** *route-policy-name*] *preference*	Optional By default, the RIPng priority is 100.

Configuring RIPng Route Redistribution

Follow these steps to configure RIPng route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Configure a default routing metric for redistributed routes	**default cost** *cost*	Optional The default metric of redistributed routes is 0.
Redistribute routes from another routing protocol	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name*] *	Required No route redistribution is configured by default.

Tuning and Optimizing the RIPng Network

This section describes how to tune and optimize the performance of the RIPng network as well as applications under special network environments. Before tuning and optimizing the RIPng network, complete the following tasks:

- Configure a network layer address for each interface
- Configure the basic RIPng functions

This section covers the following topics:

Configuring RIPng Timers

You can adjust RIPng timers to optimize the performance of the RIPng network.

Follow these steps to configure RIPng timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Configure RIPng timers	**timers** { **garbage-collect** *garbage-collect-value* | **suppress** *suppress-value* | **timeout** *timeout-value* | **update** *update-value* } *	Optional. The RIPng timers have the following defaults: • 30 seconds for the update timer • 180 seconds for the timeout timer • 120 seconds for the suppress timer • 120 seconds for the garbage-collect timer

Note

When adjusting RIPng timers, you should consider the network performance and perform unified configurations on routers running RIPng to avoid unnecessary network traffic increase or route oscillation.

Configuring Split Horizon and Poison Reverse

Note

If both split horizon and poison reverse are configured, only the poison reverse function takes effect.

Configure split horizon

The split horizon function disables a route learned from an interface from being advertised through the same interface to prevent routing loops between neighbors.

Follow these steps to configure split horizon:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the split horizon function	**ripng split-horizon**	Optional Enabled by default

Note

Generally, you are recommended to enable split horizon to prevent routing loops.

Configuring the poison reverse function

The poison reverse function enables a route learned from an interface to be advertised through the interface. However, the metric of the route is set to 16. That is to say, the route is unreachable.

Follow these steps to configure poison reverse:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the poison reverse function	**ripng poison-reverse**	Required Disabled by default

Configuring Zero Field Check on RIPng Packets

Some fields in the RIPng packet must be zero. These fields are called zero fields. With zero field check on RIPng packets enabled, if such a field contains a non-zero value, the entire RIPng packet will be discarded. If you are sure that all packets are trusty, you can disable the zero field check to reduce the CPU processing time.

Follow these steps to configure RIPng zero field check:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Enable the zero field check	**checkzero**	Optional Enabled by default

Configuring the Maximum Number of Equal Cost Routes for Load Balancing

Follow these steps to configure the maximum number of equal cost RIPng routes for load balancing:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter RIPng view	**ripng** [*process-id*]	—
Configure the maximum number of equal cost RIPng routes for load balancing	**maximum load-balancing** *number*	Optional 4 by default.

Displaying and Maintaining RIPng

To do…	Use the command…	Remarks
Display configuration information of a RIPng process	**display ripng** [*process-id*]	Available in any view
Display routes in the RIPng database	**display ripng** *process-id* **database**	Available in any view
Display the routing information of a specified RIPng process	**display ripng** *process-id* **route**	Available in any view
Display RIPng interface information	**display ripng** *process-id* **interface** [*interface-type interface-number*]	Available in any view

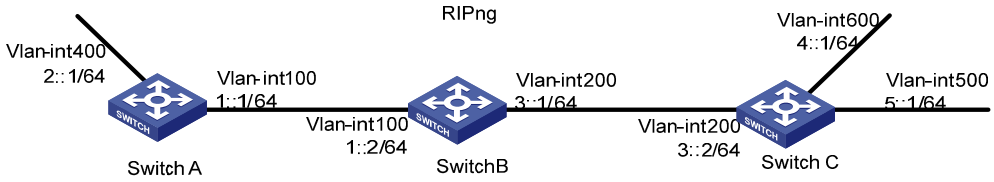
RIPng Configuration Example

Configure RIPng Basic Functions

Network requirements

As shown in Figure 1-4, all switches run RIPng. Configure Switch B to filter the route (3::/64) learnt from Switch C, which means the route will not be added to the routing table of Switch B, and Switch B will not forward it to Switch A.

Figure 1-4 Network diagram for RIPng configuration



Configuration procedure

1) Configure the IPv6 address for each interface (omitted)
2) Configure basic RIPng functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ripng 1
[SwitchA-ripng-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 400
[SwitchA-Vlan-interface400] ripng 1 enable
[SwitchA-Vlan-interface400] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ripng 1
[SwitchB-ripng-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ripng 1 enable
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ripng 1
[SwitchC-ripng-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ripng 1 enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 500
[SwitchC-Vlan-interface500] ripng 1 enable
[SwitchC-Vlan-interface500] quit
[SwitchC] interface vlan-interface 600
[SwitchC-Vlan-interface600] ripng 1 enable
[SwitchC-Vlan-interface600] quit
```

Display the routing table of Switch B.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::20F:E2FF:FE23:82F5  on Vlan-interface100
 Dest 1::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 6 Sec
 Dest 2::/64,
     via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 6 Sec

 Peer FE80::20F:E2FF:FE00:100  on Vlan-interface200
 Dest 3::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
 Dest 4::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
 Dest 5::/64,
     via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 11 Sec
```

Display the routing table of Switch A.

```
[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::200:2FF:FE64:8904  on Vlan-interface100
 Dest 1::/64,
     via FE80::200:2FF:FE64:8904, cost  1, tag 0, A, 31 Sec
 Dest 4::/64,
     via FE80::200:2FF:FE64:8904, cost  2, tag 0, A, 31 Sec
 Dest 5::/64,
     via FE80::200:2FF:FE64:8904, cost  2, tag 0, A, 31 Sec
 Dest 3::/64,
     via FE80::200:2FF:FE64:8904, cost  1, tag 0, A, 31 Sec
```

3) Configure Switch B to filter incoming and outgoing routes.

```
[SwitchB] acl ipv6 number 2000
[SwitchB-acl6-basic-2000] rule deny source 3::/64
[SwitchB-acl6-basic-2000] rule permit
[SwitchB-acl6-basic-2000] quit
[SwitchB] ripng 1
[SwitchB-ripng-1] filter-policy 2000 import
[SwitchB-ripng-1] filter-policy 2000 export
```

Display routing tables of Switch B and Switch A.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::20F:E2FF:FE23:82F5  on Vlan-interface100
 Dest 1::/64,
```

```
    via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 2 Sec
 Dest 2::/64,
    via FE80::20F:E2FF:FE23:82F5, cost  1, tag 0, A, 2 Sec

 Peer FE80::20F:E2FF:FE00:100  on Vlan-interface200
 Dest 4::/64,
    via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 5 Sec
 Dest 5::/64,
    via FE80::20F:E2FF:FE00:100, cost  1, tag 0, A, 5 Sec
[SwitchA] display ripng 1 route
  Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::20F:E2FF:FE00:1235  on Vlan-interface100
 Dest 1::/64,
    via FE80::20F:E2FF:FE00:1235, cost  1, tag 0, A, 2 Sec
 Dest 4::/64,
    via FE80::20F:E2FF:FE00:1235, cost  2, tag 0, A, 2 Sec
 Dest 5::/64,
    via FE80::20F:E2FF:FE00:1235, cost  2, tag 0, A, 2 Sec
```

Table of Contents

1 OSPFv3 Configuration

When configuring OSPF, go to these sections for information you are interested in:

Introduction to OSPFv3

OSPFv3 Overview

Open Shortest Path First version 3 (OSPFv3) supports IPv6 and complies with RFC2740 (OSPF for IPv6).

The same between OSPFv3 and OSPFv2:

- 32 bits router ID and area ID
- Packets: Hello, DD (Data Description), LSR (Link State Request), LSU (Link State Update), LSAck (Link State Acknowledgment)
- Mechanism for finding neighbors and establishing adjacencies
- Mechanism for LSA flooding and aging

Differences between OSPFv3 and OSPFv2:

- OSPFv3 runs on a per-link basis, instead of on a per-IP-subnet basis.
- OSPFv3 supports multiple instances per link.
- OSPFv3 identifies neighbors by Router ID, while OSPFv2 by IP address.

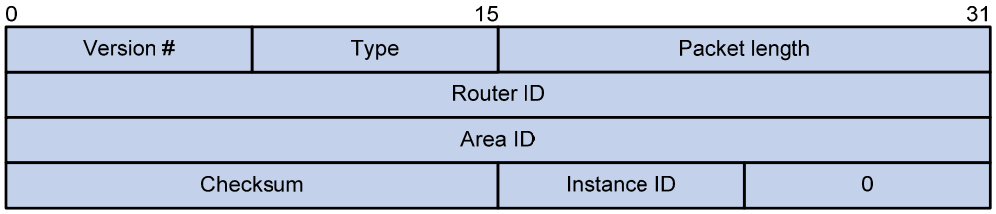
OSPFv3 Packets

OSPFv3 has also five types of packets: hello, DD, LSR, LSU, and LSAck.

The five packets have the same packet header, which different from the OSPFv2 packet header is only 16 bytes in length, has no authentication field, but is added with an Instance ID field to support multi-instance per link.

Figure 1-1 gives the OSPFv3 packet header.

Figure 1-1 OSPFv3 packet header



Major fields:

- Version #: Version of OSPF, which is 3 for OSPFv3.
- Type: Type of OSPF packet; Types 1 to 5 are hello, DD, LSR, LSU, and LSAck respectively.
- Packet Length: Packet length in bytes, including header.
- Instance ID: Instance ID for a link.
- 0: Reserved. It must be 0.

OSPFv3 LSA Types

OSPFv3 sends routing information in LSAs, which as defined in RFC2740 have the following types:

- Router-LSA: Originated by all routers. This LSA describes the collected states of the router's interfaces to an area. Flooded throughout a single area only.
- Network-LSA: Originated for broadcast and NBMA networks by the Designated Router. This LSA contains the list of routers connected to the network. Flooded throughout a single area only.
- Inter-Area-Prefix-LSA: Similar to Type 3 LSA of OSPFv2, originated by ABRs (Area Border Routers), and flooded throughout the LSA's associated area. Each Inter-Area-Prefix-LSA describes a route with IPv6 address prefix to a destination outside the area, yet still inside the AS (an inter-area route).
- Inter-Area-Router-LSA: Similar to Type 4 LSA of OSPFv2, originated by ABRs and flooded throughout the LSA's associated area. Each Inter-Area-Router-LSA describes a route to ASBR (Autonomous System Boundary Router).
- AS-external-LSA: Originated by ASBRs, and flooded throughout the AS (except Stub and NSSA areas). Each AS-external-LSA describes a route to another Autonomous System. A default route can be described by an AS external LSA.
- Link-LSA: A router originates a separate Link-LSA for each attached link. Link-LSAs have link-local flooding scope. Each Link-LSA describes the IPv6 address prefix of the link and Link-local address of the router.
- Intra-Area-Prefix-LSA: Each Intra-Area-Prefix-LSA contains IPv6 prefix information on a router, stub area or transit area information, and has area flooding scope. It was introduced because Router-LSAs and Network-LSAs contain no address information now.
- RFC 5187 defines the Type 11 LSA, Grace-LSA. A Grace-LSA is generated by a GR (Graceful Restart) Restarter at reboot and transmitted on the local link. The restarter describes the cause and interval of the reboot in the Grace-LSA to tell its neighbors that it performs a GR operation.

Timers of OSPFv3

Timers in OSPFv3 include:

- OSPFv3 packet timer
- LSA delay timer

- SPF timer
- GR timer

OSPFv3 packet timer

Hello packets are sent periodically between neighboring routers for finding and maintaining neighbor relationships, or for DR/BDR election. The hello interval must be identical on neighboring interfaces. The smaller the hello interval, the faster the network convergence speed and the bigger the network load.

If a router receives no hello packet from a neighbor within a period, it will declare the peer is down. The period is called the dead interval.

After sending an LSA to its adjacency, a router waits for an acknowledgment from the adjacency. If no response is received after the retransmission interval elapses, the router will send again the LSA. The retransmission interval must be longer than the round-trip time of the LSA.

LSA delay time

Each LSA has an age in the local LSDB (incremented by 1 per second), but an LSA does not age on transmission. You need to add an LSA delay time into the age time before transmission, which is important for low-speed networks.

SPF timer

Whenever the LSDB changes, an SPF calculation happens. If recalculations become so frequent, a large amount of resources will be occupied. You can adjust the SPF calculation interval and delay time to protect networks from being overloaded due to frequent changes.

GR timer

If a failure to establish adjacencies occurs during a GR, the device will be in the GR process for a long time. To avoid such cases, you can configure the GR timer for the device to exit the GR process when the timer expires.

OSPFv3 Features Supported

- Basic features defined in RFC2740
- OSPFv3 stub area
- OSPFv3 GR

Protocols and Standards

- RFC2740: OSPF for IPv6
- RFC2328: OSPF Version 2
- RFC 5187: OSPFv3 Graceful Restart

IPv6 OSPFv3 Configuration Task List

Complete the following tasks to configure OSPFv3:

Task		Remarks
Enabling OSPFv3		Required
Configuring OSPFv3 Area Parameters	Configuring an OSPFv3 Stub Area	Optional
	Configuring an OSPFv3 Virtual Link	Optional
Configuring OSPFv3 Network Types	Configuring the OSPFv3 Network Type for an Interface	Optional
	Configuring an NBMA or P2MP Neighbor	Optional
Configuring OSPFv3 Routing Information Control	Configuring OSPFv3 Route Summarization	Optional
	Configuring OSPFv3 Inbound Route Filtering	Optional
	Configuring an OSPFv3 Cost for an Interface	Optional
	Configuring the Maximum Number of OSPFv3 Load-balanced Routes	Optional
	Configuring a Priority for OSPFv3	Optional
	Configuring OSPFv3 Route Redistribution	Optional
Tuning and Optimizing OSPFv3 Networks	Configuring OSPFv3 Timers	Optional
	Configuring a DR Priority for an Interface	Optional
	Ignoring MTU Check for DD Packets	Optional
	Disable Interfaces from Sending OSPFv3 Packets	Optional
	Enable the Logging of Neighbor State Changes	Optional
Configuring OSPFv3 GR	Configuring GR Restarter	Optional
	Configuring GR Helper	Optional

Enabling OSPFv3

Prerequisites

- Make neighboring nodes accessible with each other at the network layer.
- Enable IPv6 packet forwarding

Enabling OSPFv3

To enable an OSPFv3 process on a router, you need to enable the OSPFv3 process globally, assign the OSPFv3 process a router ID, and enable the OSPFv3 process on related interfaces.

A router ID uniquely identifies a router within an AS. Therefore, you need to specify a unique router ID for each OSPFv3 router within the AS to ensure normal operation.

Follow these steps to enable OSPFv3:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable an OSPFv3 process and enter its view	**ospfv3** [*process-id*]	Required By default, no OSPFv3 process is enabled.
Specify a router ID	**router-id** *router-id*	Required
Enter interface view	**interface** *interface-type interface-number*	—
Enable an OSPFv3 process on the interface	**ospfv3** *process-id* **area** *area-id* [**instance** *instance-id*]	Required Not enabled by default

Configuring OSPFv3 Area Parameters

The stub area and virtual link features of OSPFv3 are the same as OSPFv2.

Splitting an OSPFv3 AS into multiple areas reduces the number of LSAs and extends OSPFv3 applications. For those non-backbone areas residing on the AS boundary, you can configure them as stub areas to further reduce the size of routing tables and the number of LSAs.

Non-backbone areas exchange routing information via the backbone area. Therefore, the backbone and non-backbone areas, including the backbone itself must be contiguous. In practice, necessary physical links may not be available for such connectivity. You can configure virtual links to address the problem.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring an OSPFv3 Stub Area

Follow these steps to configure an OSPFv3 stub area:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enter OSPFv3 area view	**area** *area-id*	—
Configure the area as a stub area	**stub** [**no-summary**]	Required Not configured by default
Specify a cost for the default route advertised to the stub area	**default-cost** *value*	Optional Defaults to 1

Note

- You cannot remove an OSPFv3 area directly. Only when you remove all configurations in area view and all interfaces attached to the area become down, can the area be removed.
- All the routers attached to a stub area must be configured with the **stub** command. The keyword **no-summary** is only available on the ABR of the stub area.
- If you use the **stub** command with the keyword **no-summary** on an ABR, the ABR advertises a default route in an Inter-Area-Prefix-LSA into the stub area. No AS-external-LSA, Inter-Area-Prefix-LSA or Inter-Area-Router-LSA is advertised in the area. The stub area of this kind is also known as a totally stub area.

Configuring an OSPFv3 Virtual Link

You can configure a virtual link to maintain connectivity between a non-backbone area and the backbone, or in the backbone itself.

Follow these steps to configure a virtual link:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enter OSPFv3 area view	**area** *area-id*	—
Configure a virtual link	**vlink-peer** *router-id* [**hello** *seconds* | **retransmit** *seconds* | **trans-delay** *seconds* | **dead** *seconds* | **instance** *instance-id*] *	Required

Note

- Both ends of a virtual link are ABRs that must be configured with the **vlink-peer** command.
- Do not configure virtual links in the areas of a GR-capable process.

Configuring OSPFv3 Network Types

OSPFv3 classifies networks into four types upon the link layer protocol:

You can change the network type of an OSPFv3 interface as needed. For example:

- An NBMA network must be fully connected. That is, any two routers in the network must be directly reachable to each other through a virtual circuit. In the event no such direct link is available, you need to change the network type through a command.
- If direct connections are not available between some routers in an NBMA network, the type of interfaces associated should be configured as P2MP, or as P2P for interfaces with only one neighbor.

Prerequisites

Before configuring OSPFv3 network types, you have configured:

- IPv6 functions
- OSPFv3 basic functions

Configuring the OSPFv3 Network Type for an Interface

Follow these steps to configure the OSPFv3 network type for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a network type for the OSPFv3 interface	**ospfv3 network-type** { **broadcast** | **nbma** | **p2mp** [**non-broadcast**] | **p2p** } [**instance** *instance-id*]	Optional By default, the network type is broadcast.

Configuring an NBMA or P2MP Neighbor

For NBMA and P2MP interfaces (only when in unicast mode), you need to specify the link-local IP addresses of their neighbors because such interfaces cannot find neighbors via broadcasting Hello packets. You can also specify DR priorities for neighbors.

Follow these steps to configure an NBMA or P2MP (unicast) neighbor and its DR priority:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify an NBMA or P2MP (unicast) neighbor and its DR priority	**ospfv3 peer** *ipv6-address* [**dr-priority** *dr-priority*] [**instance** *instance-id*]	Required

Configuring OSPFv3 Routing Information Control

This section is to configure the control of OSPF routing information advertisement and reception, and redistribution from other protocols.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring OSPFv3 Route Summarization

If contiguous network segments exist in an area, you can use the **abr-summary** command to summarize them into one network segment on the ABR. The ABR will advertise only the summary route. Any LSA falling into the specified network segment will not be advertised, reducing the LSDB size in other areas.

Follow these steps to configure route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enter OSPFv3 area view	**area** *area-id*	—
Configure a summary route	**abr-summary** *ipv6-address prefix-length* [**not-advertise**]	Required Not configured by default

Note

The **abr-summary** command takes effect on ABRs only.

Configuring OSPFv3 Inbound Route Filtering

You can configure OSPFv3 to filter routes that are computed from received LSAs according to some rules.

Follow these steps to configure OSPFv3 inbound route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Configure inbound route filtering	**filter-policy** { *acl-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	Required Not configured by default

Note

Use of the **filter-policy import** command can only filter routes computed by OSPFv3. Only routes not filtered out can be added into the local routing table.

Configuring an OSPFv3 Cost for an Interface

You can configure an OSPFv3 cost for an interface with one of the following two methods:

- Configure the cost value in interface view.
- Configure a bandwidth reference value for the interface, and OSPFv3 computes the cost automatically based on the bandwidth reference value: Interface OSPFv3 cost = Bandwidth reference value/Interface bandwidth. If the calculated cost is greater than 65535, the value of 65535 is used.

If the cost value is not configured for an interface, OSPFv3 computes the interface cost value automatically

Follow these steps to configure an OSPFv3 cost for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an OSPFv3 cost for the interface	**ospfv3 cost** *value* [**instance** *instance-id*]	Optional By default, OSPFv3 computes an interface's cost according to its bandwidth. The cost value defaults to 1 for VLAN interfaces of switches and defaults to 0 for loopback interfaces.

Follow these steps to configure a bandwidth reference value:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Configure a bandwidth reference value	**bandwidth-reference** *value*	Optional 100 Mbps by default

Configuring the Maximum Number of OSPFv3 Load-balanced Routes

If multiple equal-cost routes to a destination are available, enabling load balancing among these routes can improve link utilization.

Follow these steps to configure the maximum number of load-balanced routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Specify the maximum number of load-balanced routes	**maximum load-balancing** *maximum*	Optional 4 by default.

Configuring a Priority for OSPFv3

A router may run multiple routing protocols. The system assigns a priority for each protocol. When these routing protocols find the same route, the route found by the protocol with the highest priority is selected.

Follow these steps to configure a priority for OSPFv3:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—

To do…	Use the command…	Remarks
Configure a priority for OSPFv3	**preference** [**ase**] [**route-policy** *route-policy-name*] *preference*	Optional By default, the priority of OSPFv3 internal routes is 10, and priority of OSPFv3 external routes is 150.

Configuring OSPFv3 Route Redistribution

Follow these steps to configure OSPFv3 route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Specify a default cost for redistributed routes	**default cost** *value*	Optional Defaults to 1
Redistribute routes from another protocol, or another OSPFv3 process	**import-route** { **isisv6** *process-id* | **ospfv3** *process-id* | **ripng** *process-id* | **bgp4+** [**allow-ibgp**] | **direct** | **static** } [**cost** *value* | **type** *type* | **route-policy** *route-policy-name*] *	Required Not configured by default
Inject a default route	**default-route-advertise** [**always** | **cost** *cost* | **type** *type* | **route-policy** *route-policy-name*] *	Optional Not injected by default
Filter redistributed routes	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [**isisv6** *process-id* | **ospfv3** *process-id* | **ripng** *process-id* | **bgp4+** | **direct** | **static**]	Optional Not configured by default

Note

- Executing the **import-route** or **default-route-advertise** command on a router makes it become an ASBR,.
- You can only inject and advertise a default route using the **default-route-advertise** command.
- Since OSPFv3 is a link state routing protocol, it cannot directly filter LSAs to be advertised. Therefore, you need to filter redistributed routes first, and thus only routes that are not filtered out can be advertised in LSAs into the routing domain.
- Use of the **filter-policy export** command filters routes redistributed with the **import-route** command. If the **import-route** command is not configured, executing the **filter-policy export** command does not take effect.

Tuning and Optimizing OSPFv3 Networks

This section describes configurations of OSPFv3 timers, interface DR priority, MTU check ignorance for DD packets, and disabling interfaces from sending OSPFv3 packets.

OSPFv3 timers:

- Packet timer: Specified to adjust topology convergence speed and network load
- LSA delay timer: Specified especially for low-speed links
- SPF timer: Specified to protect networks from being over-loaded due to frequent network changes.

For a broadcast network, you can configure DR priorities for interfaces to affect DR/BDR election.

By disabling an interface from sending OSPFv3 packets, you can make other routers on the network obtain no information from the interface.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring OSPFv3 Timers

Follow these steps to configure OSPFv3 timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the hello interval	**ospfv3 timer hello** *seconds* [**instance** *instance-id*]	Optional Defaults to 10 seconds on P2P, broadcast interfaces.
Specify the poll interval	**ospfv3 timer poll** *seconds* [**instance** *instance-id*]	Optional The poll interval defaults to 120 seconds.
Configure the dead interval	**ospfv3 timer dead** *seconds* [**instance** *instance-id*]	Optional Defaults to 40 seconds on P2P, broadcast interfaces.
Configure the LSA retransmission interval	**ospfv3 timer retransmit** *interval* [**instance** *instance-id*]	Optional Defaults to 5 seconds.
Configure the LSA transmission delay	**ospfv3 trans-delay** *seconds* [**instance** *instance-id*]	Optional Defaults to 1 second.
Return to system view	**quit**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Configure the SPF timers	**spf timers** *delay-interval hold-interval*	Optional By default, *delay-interval* is 5 seconds, and *hold-interval* is 10 seconds.

Note

- The dead interval set on neighboring interfaces cannot be too short. Otherwise, a neighbor is easily considered down.
- The LSA retransmission interval cannot be too short; otherwise, unnecessary retransmissions occur.

Configuring a DR Priority for an Interface

Follow these steps to configure a DR priority for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a DR priority	**ospfv3 dr-priority** *priority* [**instance** *instance-id*]	Optional Defaults to 1

Note

The DR priority of an interface determines the interface's qualification in DR election. Interfaces having the priority 0 cannot become a DR or BDR.

Ignoring MTU Check for DD Packets

When LSAs are few in DD packets, it is unnecessary to check the MTU in DD packets in order to improve efficiency.

Follow these steps to ignore MTU check for DD packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Ignore MTU check for DD packets	**ospfv3 mtu-ignore** [**instance** *instance-id*]	Required Not ignored by default

Disable Interfaces from Sending OSPFv3 Packets

Follow these steps to disable interfaces from sending OSPFv3 packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Disable interfaces from sending OSPFv3 packets	**silent-interface** { *interface-type interface-number* | **all** }	Required Not disabled by default

Note

After an OSPF interface is set to **silent**, direct routes of the interface can still be advertised in Intra-Area-Prefix-LSAs via other interfaces, but other OSPFv3 packets cannot be advertised. Therefore, no neighboring relationship can be established on the interface. This feature can enhance the adaptability of OSPFv3 networking.

Enable the Logging of Neighbor State Changes

Follow these steps to enable the logging of neighbor state changes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enable the logging of neighbor state changes	**log-peer-change**	Required Enabled by default

Configuring OSPFv3 GR

Note

You cannot configure OSPFv3 GR after configuring OSPFv3 virtual links, becase they are not supported at the same time.

To prevent service interruption after a master/backup switchover, a GR Restarter running OSPFv3 must complete the following tasks:

- Keep the GR Restarter forwarding entries stable during reboot.
- Establish all adjacencieis and obtain complete topology information after reboot.

After reboot, the GR Restarter sends a Grace-LSA to tell its neighbors that it performs a GR. Upon receiving the Grace-LSA, the neighbors with the GR Helper capability enter the helper mode (and are

thus called GR Helpers). Then, the GR Restarter retrieves its adjacencies and LSDB with the help of the GR Helpers. Thus, the normal data forwarding is ensured.

Configuring GR Restarter

You can configure the GR Restarter capability on a GR Restarter.

Follow these steps to configure GR Restarter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enable the GR capability	**graceful-restart enable**	Required Disabled by default.
Configure the GR interval	**graceful-restart interval** *interval-value*	Optional 120 seconds by default.

Configuring GR Helper

You can configure the GR Helper capability on a GR Helper.

Follow these steps to configure GR Helper

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter OSPFv3 view	**ospfv3** [*process-id*]	—
Enable the GR Helper capability	**graceful-restart helper enable**	Optional Enabled by default.
Enable strict LSA checking	**graceful-restart helper strict-lsa-checking**	Optional Disabled by default.

Displaying and Maintaining OSPFv3

To do…	Use the command…	Remarks
Display OSPFv3 debugging state information	**display debugging ospfv3**	Available in any view
Display OSPFv3 process brief information	**display ospfv3** [*process-id*]	
Display OSPFv3 interface information	**display ospfv3 interface** [*interface-type interface-number* | **statistic**]	
Display OSPFv3 LSDB information	**display ospfv3** [*process-id*] **lsdb** [[**external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace**] [*link-state-id*] [**originate-router** *router-id*] | **total**]	
Display OSPFv3 LSDB statistics	**display ospfv3 lsdb statistic**	
Display OSPFv3 neighbor information	**display ospfv3** [*process-id*] [**area** *area-id*] **peer** [[*interface-type interface-number*] [**verbose**] | *peer-router-id*]	
Display OSPFv3 neighbor statistics	**display ospfv3 peer statistic**	
Display OSPFv3 routing table information	**display ospfv3** [*process-id*] **routing** [*ipv6-address prefix-length* **|** *ipv6-address/prefix-length* | **abr-routes** | **asbr-routes** | **all | statistics**]	
Display OSPFv3 area topology information	**display ospfv3** [*process-id*] **topology** [**area** *area-id*]	
Display OSPFv3 virtual link information	**display ospfv3** [*process-id*] **vlink**	
Display OSPFv3 next hop information	**display ospfv3** [*process-id*] **next-hop**	
Display OSPFv3 link state request list information	**display ospfv3** [*process-id*] **request-list** [{ **external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace** } [*link-state-id*] [**originate-router** *ip-address*] | **statistics**]	
Display OSPFv3 link state retransmission list information	**display ospfv3** [*process-id*] **retrans-list** [{ **external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router** | **grace** } [*link-state-id*] [**originate-router** *ip-address*] | **statistics**]	
Display OSPFv3 statistics	**display ospfv3 statistics**	
Display the GR status of the specified OSPFv3 process	**display ospfv3** [*process-id*] **graceful-restart status**	

OSPFv3 Configuration Examples

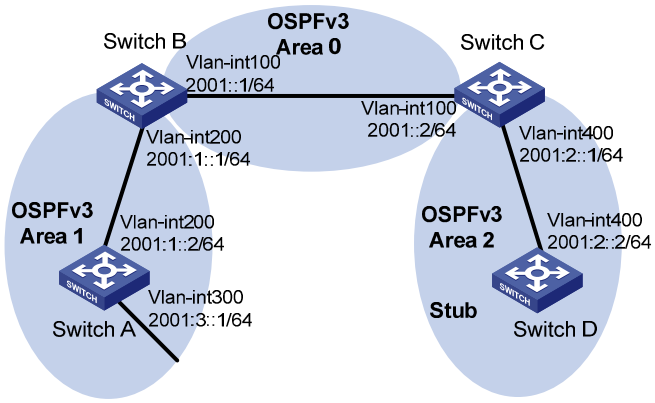
Configuring OSPFv3 Areas

Network requirements

In the following figure, all switches run OSPFv3. The AS is split into three areas, in which, Switch B and Switch C act as ABRs to forward routing information between areas.

It is required to configure Area 2 as a stub area to reduce LSAs in the area without affecting route reachability.

Figure 1-2 Network diagram for OSPFv3 area configuration



Configuration procedure

1) Configure IPv6 addresses for interfaces (omitted)
2) Configure OSPFv3 basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 300
[SwitchA-Vlan-interface300] ospfv3 1 area 1
[SwitchA-Vlan-interface300] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ospfv3 1 area 1
[SwitchA-Vlan-interface200] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ospfv3
[SwitchB-ospf-1] router-id 2.2.2.2
[SwitchB-ospf-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 0
[SwitchB-Vlan-interface100] quit
```

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ospfv3 1 area 2
[SwitchC-Vlan-interface400] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface Vlan-interface 400
[SwitchD-Vlan-interface400] ospfv3 1 area 2
[SwitchD-Vlan-interface400] quit
```

Display OSPFv3 neighbor information on Switch B.

```
[SwitchB] display ospfv3 peer

            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID    Pri   State          Dead Time   Interface      Instance ID
3.3.3.3        1     Full/DR        00:00:39    Vlan100        0

            OSPFv3 Area ID 0.0.0.1 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State         Dead Time   Interface      Instance ID
1.1.1.1         1     Full/Backup   00:00:38    Vlan200        0
```

Display OSPFv3 neighbor information on Switch C.

```
[SwitchC] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID    Pri   State          Dead Time   Interface      Instance ID
2.2.2.2        1     Full/Backup    00:00:39    Vlan100        0

            OSPFv3 Area ID 0.0.0.2 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State         Dead Time   Interface      Instance ID
```

```
4.4.4.4         1     Full/DR       00:00:38    Vlan400        0
```

Display OSPFv3 routing table information on Switch D.

```
[SwitchD] display ospfv3 routing

E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route

            OSPFv3 Router with ID (4.4.4.4) (Process 1)
------------------------------------------------------------------------
 *Destination: 2001::/64
  Type       : IA                                    Cost     : 2
  NextHop    : FE80::F40D:0:93D0:1                   Interface: Vlan400

 *Destination: 2001:1::/64
  Type       : IA                                    Cost     : 3
  NextHop    : FE80::F40D:0:93D0:1                   Interface: Vlan400

 *Destination: 2001:2::/64
  Type       : I                                     Cost     : 1
  NextHop    : directly-connected                    Interface: Vlan400

 *Destination: 2001:3::/64
  Type       : IA                                    Cost     : 4
  NextHop    : FE80::F40D:0:93D0:1                   Interface: Vlan400
```

3) Configure Area 2 as a stub area

Configure Switch D

```
[SwitchD] ospfv3
[SwitchD-ospfv3-1] area 2
[SwitchD-ospfv3-1-area-0.0.0.2] stub
```

Configure Switch C, and specify the cost of the default route sent to the stub area as 10.

```
[SwitchC] ospfv3
[SwitchC-ospfv3-1] area 2
[SwitchC-ospfv3-1-area-0.0.0.2] stub
[SwitchC-ospfv3-1-area-0.0.0.2] default-cost 10
```

Display OSPFv3 routing table information on Switch D. You can find a default route is added, and its cost is the cost of a direct route plus the configured cost.

```
[SwitchD] display ospfv3 routing
E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route

            OSPFv3 Router with ID (4.4.4.4) (Process 1)
 ------------------------------------------------------------------------
 *Destination: ::/0
  Type       : IA                                    Cost     : 11
  NextHop    : FE80::F40D:0:93D0:1                   Interface: Vlan400
```

```
 *Destination: 2001::/64
  Type       : IA                                  Cost     : 2
  NextHop    : FE80::F40D:0:93D0:1                 Interface: Vlan400

 *Destination: 2001:1::/64
  Type       : IA                                  Cost     : 3
  NextHop    : FE80::F40D:0:93D0:1                 Interface: Vlan400

 *Destination: 2001:2::/64
  Type       : I                                   Cost     : 1
  NextHop    : directly-connected                  Interface: Vlan400

 *Destination: 2001:3::/64
  Type       : IA                                  Cost     : 4
  NextHop    : FE80::F40D:0:93D0:1                 Interface: Vlan400
```

4) Configure Area 2 as a totally stub area

Configure Area 2 as a totally stub area on Switch C.

```
[SwitchC-ospfv3-1-area-0.0.0.2] stub no-summary
```

Display OSPFv3 routing table information on Switch D. You can find route entries are reduced. All non-direct routes are removed except the default route.

```
[SwitchD] display ospfv3 routing
E1 - Type 1 external route,    IA - Inter area route,    I  - Intra area route
E2 - Type 2 external route,    *  - Seleted route

            OSPFv3 Router with ID (4.4.4.4) (Process 1)
 ------------------------------------------------------------------------
 *Destination: ::/0
  Type       : IA                                  Cost     : 11
  NextHop    : FE80::F40D:0:93D0:1                 Interface: Vlan400

 *Destination: 2001:2::/64
  Type       : I                                   Cost     : 1
  NextHop    : directly-connected                  Interface: Vlan400
```

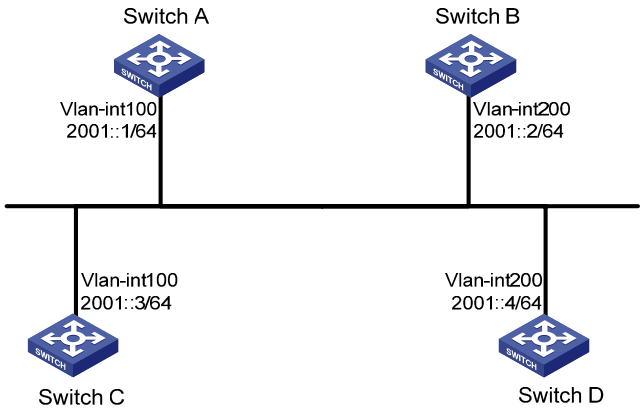
Configuring OSPFv3 DR Election

Network requirements

In the following figure:

- The priority of Switch A is 100, the highest priority on the network, so it will be the DR.
- The priority of Switch C is 2, the second highest priority on the network, so it will be the BDR.
- The priority of Switch B is 0, so it cannot become the DR.
- Router D has the default priority 1.

Figure 1-3 Network diagram for OSPFv3 DR election configuration



Configuration procedure

1) Configure IPv6 addresses for interfaces (omitted)
2) Configure OSPFv3 basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 0
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 0
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] ospfv3 1 area 0
[SwitchD-Vlan-interface200] quit
```

Display neighbor information on Switch A. You can find the switches have the same default DR priority 1. In this case, the switch with the highest Router ID is elected as the DR. Therefore, Switch D is the DR, and Switch C is the BDR.

```
 [SwitchA] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         1     2-Way/DROther    00:00:36    Vlan200        0
3.3.3.3         1     Full/Backup      00:00:35    Vlan100        0
4.4.4.4         1     Full/DR          00:00:33    Vlan200        0
```

Display neighbor information on Switch D. You can find the neighbor states are all full.

```
 [SwitchD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         1     Full/DROther     00:00:30    Vlan100        0
2.2.2.2         1     Full/DROther     00:00:37    Vlan200        0
3.3.3.3         1     Full/Backup      00:00:31    Vlan100        0
```

3) Configure DR priorities for interfaces.

Configure the DR priority of VLAN-interface 100 as 100 on Switch A.

```
[SwitchA] interface Vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 dr-priority 100
[SwitchA-Vlan-interface100] quit
```

Configure the DR priority of VLAN-interface 200 as 0 on Switch B.

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 dr-priority 0
[SwitchB-Vlan-interface200] quit
```

Configure the DR priority of VLAN-interface 100 of Switch C as 2.

```
[SwitchC] interface Vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 dr-priority 2
[SwitchC-Vlan-interface100] quit
```

Display neighbor information on Switch A. You can find DR priorities have been updated, but the DR and BDR are not changed.

```
[SwitchA] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ------------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
```

```
2.2.2.2         0     2-Way/DROther   00:00:38    Vlan200         0
3.3.3.3         2     Full/Backup     00:00:32    Vlan100         0
4.4.4.4         1     Full/DR         00:00:36    Vlan200         0
```

Display neighbor information on Switch D. You can find Switch D is still the DR.

```
[SwitchD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ----------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         100   Full/DROther     00:00:33    Vlan100         0
2.2.2.2         0     Full/DROther     00:00:36    Vlan200         0
3.3.3.3         2     Full/Backup      00:00:40    Vlan100         0
```

4) Restart DR/BDR election

Use the **shutdown** and **undo shutdown** commands on interfaces to restart DR/BDR election (omitted).

Display neighbor information on Switch A. You can find Switch C becomes the BDR.

```
[SwitchA] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ----------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
2.2.2.2         0     Full/DROther     00:00:31    Vlan200         0
3.3.3.3         2     Full/Backup      00:00:39    Vlan100         0
4.4.4.4         1     Full/DROther     00:00:37    Vlan200         0
```

Display neighbor information on Switch D. You can find Switch A becomes the DR.

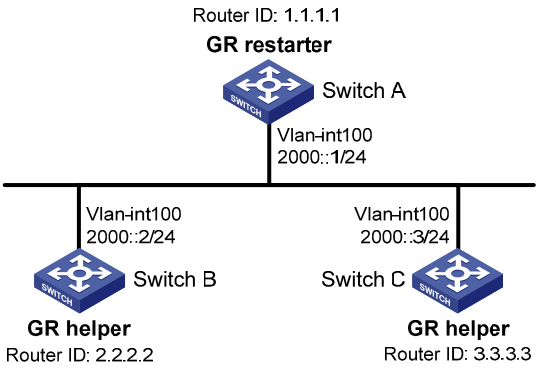
```
[SwitchD] display ospfv3 peer
            OSPFv3 Area ID 0.0.0.0 (Process 1)
 ----------------------------------------------------------------------
Neighbor ID     Pri   State            Dead Time   Interface      Instance ID
1.1.1.1         100   Full/DR          00:00:34    Vlan100         0
2.2.2.2         0     2-Way/DROther    00:00:34    Vlan200         0
3.3.3.3         2     Full/Backup      00:00:32    Vlan100         0
```

Configuring OSPFv3 GR

Network requirements

- As shown in Figure 1-4, Switch A, Switch B and Switch C that belong to the same AS and the same OSPFv3 routing domain are GR capable.
- Switch A acts as the GR Restarter. Switch B and Switch C are the GR Helpers and synchronize their LSDBs with Switch A through out-of-band (OOB) communication of GR.

Figure 1-4 Network diagram for OSPFv3 GR configuration



Configuration procedure

1) Configure IPv6 addresses for interfaces (omitted).
2) Configure OSPFv3 basic functions

On Switch A, enable OSPFv3 process 1, enable GR and set the router ID to 1.1.1.1.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ospfv3 1
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] graceful-restart enable
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 1
[SwitchA-Vlan-interface100] quit
```

Enable OSPFv3 on Switch B and set the router ID to 2.2.2.2. (By default, GR helpler is enabled on Switch B).

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ospfv3 1
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ospfv3 1 area 1
[SwitchB-Vlan-interface100] quit
```

Enable OSPFv3 on Switch C and set the router ID to 3.3.3.3. (By default, GR helpler is enabled on Switch C).

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ospfv3 1
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 1
[SwitchC-Vlan-interface100] quit
```

3) Verify the configuration

After all switches function properly, perform a master/backup switchover on Switch A to trigger a OSPFv3 GR operation.

Troubleshooting OSPFv3 Configuration

No OSPFv3 Neighbor Relationship Established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower protocol work well, check OSPF parameters configured on interfaces. The two neighboring interfaces must have the same parameters, such as the area ID, network segment and mask and network type. If the network type is broadcast, at least one interface must have a DR priority higher than 0.

Process steps

1) Display neighbor information using the **display ospfv3 peer** command.
2) Display OSPFv3 interface information using the **display ospfv3 interface** command.
3) Ping the neighbor router's IP address to check connectivity.
4) Check OSPF timers. The dead interval on an interface must be at least four times the hello interval.
5) On a broadcast network, at least one interface must have a DR priority higher than 0.

Incorrect Routing Information

Symptom

OSPFv3 cannot find routes to other areas.

Analysis

The backbone area must maintain connectivity to all other areas. If a router connects to more than one area, at least one area must be connected to the backbone. The backbone cannot be configured as a Stub area.

In a Stub area, all routers cannot receive external routes, and all interfaces connected to the Stub area must be associated with the Stub area.

Solution

1) Use the **display ospfv3 peer** command to display OSPFv3 neighbors.
2) Use the **display ospfv3 interface** command to display OSPFv3 interface information.
3) Use the **display ospfv3 lsdb** command to display Link State Database information to check integrity.
4) Display information about area configuration using the **display current-configuration configuration** command. If more than two areas are configured, at least one area is connected to the backbone.
5) In a Stub area, all routers are configured with the **stub** command.
6) If a virtual link is configured, use the **display ospf vlink** command to check the neighbor state.

Table of Contents

1 IPv6 IS-IS Configuration

Note

IPv6 IS-IS supports all the features of IPv4 IS-IS except that it advertises IPv6 routing information instead. This document describes only IPv6 IS-IS exclusive configuration tasks. For other configuration tasks, refer to *IS-IS Configuration* in the *IP Routing Volume*.

When configuring IPv6 IS-IS, go to these sections for information you are interested in:

Introduction to IPv6 IS-IS

The IS-IS routing protocol (Intermediate System-to-Intermediate System intra-domain routing information exchange protocol) supports multiple network protocols, including IPv6. IS-IS with IPv6 support is called IPv6 IS-IS dynamic routing protocol. The international engineer task force (IETF) defines two type-length-values (TLVs) and a new network layer protocol identifier (NLPID) to enable IPv6 support for IS-IS.

TLV is a variable-length field in the link state PDU or link state packet (LSP). The two TLVs are:

- IPv6 Reachability: Defines the prefix, metric of routing information to indicate network reachability, and has a type value of 236 (0xEC).
- IPv6 Interface Address: Same as the “IP Interface Address” TLV in IPv4 ISIS, except that the 32-bit IPv4 address is translated to the 128-bit IPv6 address.

The NLPID is an 8-bit field that identifies which network layer protocol is supported. For IPv6, the NLPID is 142 (0x8E), which must be carried in hello packets sent by a router that supports IPv6 IS-IS.

For information about IS-IS, refer to *IS-IS Configuration* in the *IP Routing Volume*.

Configuring IPv6 IS-IS Basic Functions

Note

You can implement IPv6 inter-networking through configuring IPv6 IS-IS in IPv6 network environment.

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Enable IPv6 globally
- Configure IP addresses for interfaces, and make sure all neighboring nodes are reachable.
- Enable IS-IS

Configuration Procedure

Follow these steps to configure the basic functions of IPv6 IS-IS:

To do…	Use command to…	Remarks
Enter system view	**system-view**	—
Enable an IS-IS process and enter IS-IS view	**isis** [*process-id*]	Required Not enabled by default
Configure the network entity title for the IS-IS process	**network-entity** *net*	Required Not configured by default
Enable IPv6 for the IS-IS process	**ipv6 enable**	Required Disabled by default
Return to system view	**quit**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable IPv6 for an IS-IS process on the interface	**isis ipv6 enable** [*process-id*]	Required Disabled by default

Configuring IPv6 IS-IS Routing Information Control

Configuration Prerequisites

You need to complete the IPv6 IS-IS basic function configuration before configuring this task.

Configuration Procedure

Follow these steps to configure IPv6 IS-IS routing information control:

To do…	Use command to…	Remarks
Enter system view	**system-view**	—
Enter IS-IS view	**isis** [*process-id*]	—
Define the priority for IPv6 IS-IS routes	**ipv6 preference** { **route-policy** *route-policy-name* | *preference* } *	Optional 15 by default
Configure an IPv6 IS-IS summary route	**ipv6 summary** *ipv6-prefix prefix-length* [**avoid-feedback** | **generate_null0_route** | [**level-1** | **level-1-2** | **level-2**] | **tag** *tag*] *	Optional Not configured by default
Generate an IPv6 IS-IS default route	**ipv6 default-route-advertise** [[**level-1** | **level-2** | **level-1-2**] | **route-policy** *route-policy-name*] *	Optional No IPv6 default route is defined by default.
Configure IPv6 IS-IS to filter incoming routes	**ipv6 filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** *route-policy-name* } **import**	Optional No filtering policy is defined by default
Configure IPv6 IS-IS to redistribute routes from another routing protocol	**ipv6 import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | [**level-1** | **level-2** | **level-1-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	Optional Not configured by default
Configure the maximum number of redistributed Level 1/Level 2 IPv6 routes	**ipv6 import-route limit** *number*	Optional 6144 by default
Configure the filtering of outgoing redistributed routes	**ipv6 filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]	Optional Not configured by default
Enable route leaking	**ipv6 import-route isisv6 level-2 into level-1** [**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* | **route-policy** route-policy-name } | **tag** *tag*] *	Optional Not enabled by default
Specify the maximum number of equal-cost load balanced routes	**ipv6 maximum load-balancing** *number*	Optional 4 by default

Note

- The **ipv6 filter-policy export** command is usually used in combination with the **ipv6 import-route** command. If no *protocol* is specified for the **ipv6 filter-policy export** command, routes redistributed from all routing protocols are filtered before advertisement. If a protocol is specified, only routes redistributed from the routing protocol are filtered for advertisement.
- For information about ACL, refer to *ACL Configuration* in the *Security Volume*.
- For information about routing policy and IPv6 prefix list, refer to *Routing Policy Configuration* in the *IP Routing Volume*.

Displaying and Maintaining IPv6 IS-IS

To do…	Use the command…	Remarks
Display brief IPv6 IS-IS information	**display isis brief**	Available in any view
Display the status of the debug switches	**display isis debug-switches** { *process-id* | **vpn-instance** *vpn-instance-name* }	Available in any view
Display IS-IS enabled interface information	**display isis interface** [**statistics** | [*interface-type interface-number*] [**verbose**]] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS license information	**display isis license**	Available in any view
Display LSDB information	**display isis lsdb** [[**l1** | **l2** | **level-1** | **level-2**] | [[**lsp-id** *lsp-id* | **lsp-name** *lspname* | **local**] | **verbose**] *] * [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS mesh group information	**display isis mesh-group** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display the mapping table between the host name and system ID	**display isis name-table** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IS-IS neighbor information	**display isis peer** [**verbose | statistics**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display IPv6 IS-IS routing information	**display isis route ipv6** [[**level-1** | **level-2**] | **verbose**] * [*process-id*]	Available in any view
Display SPF log information	**display isis spf-log** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Display the statistics of the IS-IS process	**display isis statistics** [**level-1** | **level-2** | **level-1-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view
Clear all IS-IS data structure information	**reset isis all** [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in user view

To do…	Use the command…	Remarks
Clear the IS-IS data information of a neighbor	**reset isis peer** *system-id* [*process-id* | **vpn** *vpn-instance-name*]	Available in user view

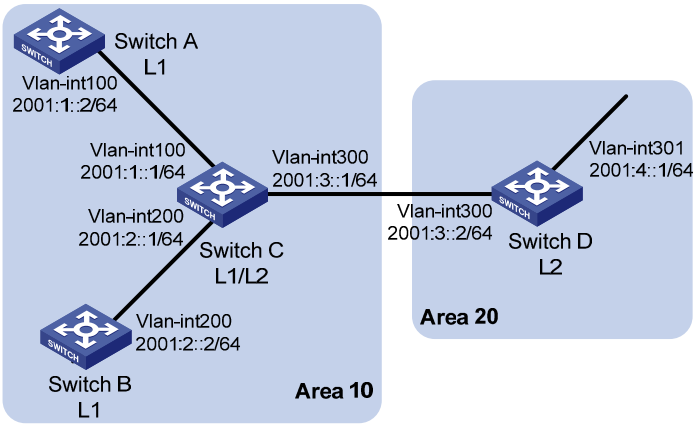
IPv6 IS-IS Configuration Example

Network requirements

As shown in Figure 1-1, Switch A, Switch B, Switch C and Switch D reside in the same autonomous system, and all are enabled with IPv6.

Switch A and Switch B are Level-1 switches, Switch D is a Level-2 switch, and Switch C is a Level-1-2 switch. Switch A, Switch B, and Switch C are in area 10, while Switch D is in area 20.

Figure 1-1 Network diagram for IPv6 IS-IS basic configuration



Configuration procedure

1) Configure IPv6 addresses for interfaces (omitted)
2) Configure IPv6 IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] ipv6 enable
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis ipv6 enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] ipv6 enable
```

```
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis ipv6 enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] ipv6 enable
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis ipv6 enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis ipv6 enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis ipv6 enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] ipv6 enable
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis ipv6 enable 1
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 301
[SwitchD-Vlan-interface301] isis ipv6 enable 1
[SwitchD-Vlan-interface301] quit
```

Table of Contents

1 IPv6 BGP Configuration

Note

This chapter describes only configuration for IPv6 BGP. For BGP related information, refer to *BGP Configuration* in the *IP Routing Volume*.

When configuring IPv6 BGP, go to these sections for information you are interested in:

IPv6 BGP Overview

BGP-4 was designed to carry only IPv4 routing information, and thus other network layer protocols such as IPv6 are not supported.

To support multiple network layer protocols, IETF extended BGP-4 by introducing Multiprotocol BGP (MP-BGP), which is defined in RFC 2858 (multiprotocol extensions for BGP-4).

MP-BGP for IPv6 is referred to as IPv6 BGP for short.

IPv6 BGP puts IPv6 network layer information into the attributes of network layer reachable information (NLRI) and NEXT_HOP.

The NLRI attribute of IPv6 BGP involves:

- MP_REACH_NLRI: Multiprotocol Reachable NLRI, for advertising reachable route and next hop information.
- MP_UNREACH_NLRI: Multiprotocol Unreachable NLRI, for withdrawal of unreachable routes.

The NEXT_HOP attribute of IPv6 BGP is identified by an IPv6 unicast address or IPv6 local link address.

IPv6 BGP has the same messaging and routing mechanisms as BGP.

Configuration Task List

Complete the following tasks to configure IPv6 BGP:

Task		Remarks
Configuring IPv6 BGP Basic Functions	Specifying an IPv6 BGP Peer	Required
	Injecting a Local IPv6 Route	Optional
	Configuring a Preferred Value for Routes from a Peer/Peer Group	Optional
	Specifying the Source Interface for Establishing TCP Connections	Optional
	Allowing the establishment of a Non-Direct eBGP connection	Optional
	Configuring a Description for an IPv6 Peer/Peer Group	Optional
	Disabling Session Establishment to an IPv6 Peer/Peer Group	Optional
	Logging IPv6 Peer/Peer Group State Changes	Optional
Controlling Route Distribution and Reception	Configuring IPv6 BGP Route Redistribution	Optional
	Configuring IPv6 BGP Route Summarization	Optional
	Advertising a Default Route to an IPv6 Peer/Peer Group	Optional
	Configuring Outbound Route Filtering	Optional
	Configuring Inbound Route Filtering	Optional
	Configuring IPv6 BGP and IGP Route Synchronization	Optional
	Configuring Route Dampening	Optional
Configuring IPv6 BGP Route Attributes	Configuring IPv6 BGP Preference and Default LOCAL_PREF and NEXT_HOP Attributes	Optional
	Configuring the MED Attribute	Optional
	Configuring the AS_PATH Attribute	Optional
Tuning and Optimizing IPv6 BGP Networks	Configuring IPv6 BGP Timers	Optional
	Configuring IPv6 BGP Soft Reset	Optional
	Configuring the Maximum Number of Load-Balanced Routes	Optional
Configuring a Large Scale IPv6 BGP Network	Configuring IPv6 BGP Peer Group	Optional
	Configuring IPv6 BGP Community	Optional
	Configuring an IPv6 BGP Route Reflector	Optional

Configuring IPv6 BGP Basic Functions

Prerequisites

Before configuring this task, you need to:

- Specify IP addresses for interfaces.
- Enable IPv6.

Note

You need create a peer group before configuring basic functions for it. For related information, refer to Configuring IPv6 BGP Peer Group.

Specifying an IPv6 BGP Peer

Follow these steps to configure an IPv6 BGP peer:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify a router ID	**router-id** *router-id*	Optional Required if no IP addresses are configured for any interfaces.
Enter IPv6 address family view	**ipv6-family**	—
Specify an IPv6 peer and its AS number	**peer** *ipv6-address* **as-number** *as-number*	Required Not configured by default.

Injecting a Local IPv6 Route

Follow these steps to configure advertise a local route into the routing table:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Inject a local route into the IPv6 BGP routing table	**network** *ipv6-address prefix-length* [**short-cut** | **route-policy** *route-policy-name*]	Required Not added by default

Configuring a Preferred Value for Routes from a Peer/Peer Group

Follow these steps to configure a preferred value for routes received from a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure a preferred value for routes received from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value*	Optional By default, the preferred value is 0.

Note

If you both reference a routing policy and use the command **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* to set a preferred value for routes from a peer/peer group, the routing policy sets a non-zero preferred value for routes matching it. Other routes not matching the routing policy uses the value set with the command. If the preferred value in the routing policy is zero, the routes matching it will also use the value set with the command. For information about using a routing policy to set a preferred value, refer to the command **peer** { *group-name* | *ipv4-address* | *ipv6-address* } **route-policy** *route-policy-name* { **import** | **export** } in this document, and the command **apply preferred-value** *preferred-value* in *Routing Policy Commands* of the *IP Routing Volume*.

Specifying the Source Interface for Establishing TCP Connections

Follow these steps to specify the source interface for establishing TCP connections to a BGP peer or peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Specify the source interface for establishing TCP connections to an IPv6 BGP peer or peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **connect-interface** *interface-type interface-number*	Required By default, IPv6 BGP uses the outbound interface of the best route to the IPv6 BGP peer or peer group as the source interface for establishing a TCP connection.

Note

- To improve stability and reliability, you can specify a loopback interface as the source interface for establishing TCP connections to a BGP peer. By doing so, a connection failure upon redundancy availability will not affect TCP connection establishment.
- To establish multiple BGP connections to a BGP router, you need to specify on the local router the respective source interfaces for establishing TCP connections to the peers on the peering BGP router; otherwise, the local BGP router may fail to establish TCP connections to the peers when using the outbound interfaces of the best routes as the source interfaces.

Allowing the establishment of a Non-Direct eBGP connection

Follow these steps to allow the establishment of eBGP connection to a non-directly connected peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Allow the establishment of eBGP connection to a non directly connected peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ebgp-max-hop** [*hop-count*]	Required Not configured by default.

Caution

In general, direct links should be available between eBGP peers. If not, you can use the **peer ebgp-max-hop** command to establish a multi-hop TCP connection in between. However, you need not use this command for direct eBGP connections with loopback interfaces.

Configuring a Description for an IPv6 Peer/Peer Group

Follow these steps to configure description for an IPv6 peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure a description for an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **description** *description-text*	Optional Not configured by default.

Note

The peer group to be configured with a description must have been created.

Disabling Session Establishment to an IPv6 Peer/Peer Group

Follow these steps to disable session establishment to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Disable session establishment to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ignore**	Optional Not disabled by default

Logging IPv6 Peer/Peer Group State Changes

Follow these steps to configure to log on the session and event information of an IPv6 peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable logging of peer changes globally	**log-peer-change**	Optional Enabled by default.
Enter IPv6 address family view	**ipv6-family**	—
Enable the state change logging for an IPv6 peer or peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **log-change**	Optional Enabled by default.

Note

Refer to *BGP Commands* in the *IP Routing Volume* for information about the **log-peer-change** command.

Controlling Route Distribution and Reception

The task includes routing information filtering, routing policy application and route dampening.

Prerequisites

Before configuring this task, you need to:

- Enable IPv6
- Configure the IPv6 BGP basic functions

Configuring IPv6 BGP Route Redistribution

Follow these steps to configure IPv6 BGP route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Enable default route redistribution into the IPv6 BGP routing table	**default-route imported**	Optional Not enabled by default.
Enable route redistribution from another routing protocol	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required Not enabled by default.

Note

If the **default-route imported** command is not configured, using the **import-route** command cannot redistribute any IGP default route.

Configuring IPv6 BGP Route Summarization

To reduce the routing table size on medium and large BGP networks, you need to configure route summarization on BGP routers. BGP supports only manual summarization of IPv6 routes.

Follow these steps to configure IPv6 BGP route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure manual route summarization	**aggregate** *ipv6-address prefix-length* [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	Required Not configured by default.

Advertising a Default Route to an IPv6 Peer/Peer Group

Follow these steps to advertise a default route to an IPv6 peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Advertise a default route to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	Required Not advertised by default.

Note

With the **peer default-route-advertise** command executed, the local router advertises a default route with itself as the next hop to the specified IPv6 peer/peer group, regardless of whether the default route is available in the routing table.

Configuring Outbound Route Filtering

Follow these steps to configure outbound route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure the filtering of outgoing routes	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol process-id*]	Required Not configured by default.
Apply a routing policy to routes advertised to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	Required Not applied by default.
Specify an IPv6 ACL to filer routes advertised to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **export**	Required Not specified by default.
Specify an AS path ACL to filer routes advertised to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **export**	Required Not specified by default.
Specify an IPv6 prefix list to filer routes advertised to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **export**	Required Not specified by default.

Note

IPv6 BGP advertises routes passing the specified policy to peers. Using the *protocol* argument can filter only the routes redistributed from the specified protocol. If no *protocol* is specified, IPv6 BGP filters all routes to be advertised, including redistributed routes and routes imported with the **network** command.

Configuring Inbound Route Filtering

Follow these steps to configure inbound route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure inbound route filtering	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	Required Not configured by default.
Apply a routing policy to routes from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **import**	Required Not applied by default.
Specify an ACL to filter routes imported from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **import**	Required Not specified by default.
Specify an AS path ACL to filter routing information imported from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **import**	Required Not specified by default.
Specify an IPv6 prefix list to filter routing information imported from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **import**	Required Not specified by default.
Specify the upper limit of prefixes allowed to receive from an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-limit** *limit* [*percentage*]	Optional Unlimited by default.

Note

- Only routes passing the configured filtering can be added into the local IPv6 BGP routing table.
- Members of a peer group can have different inbound route filtering policies.

Configuring IPv6 BGP and IGP Route Synchronization

With this feature enabled and when a non-BGP router is responsible for forwarding packets in an AS, IPv6 BGP speakers in the AS cannot advertise routing information to outside ASs unless all routers in the AS know the latest routing information.

By default, when a BGP router receives an iBGP route, it only checks the reachability of the route's next hop before advertisement. If the synchronization feature is configured, only the iBGP route is advertised by IGP can the route be advertised to eBGP peers.

Follow these steps to configure IPv6 BGP and IGP route synchronization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Enable route synchronization between IPv6 BGP and IGP	**synchronization**	Required Not enabled by default.

Configuring Route Dampening

Follow these steps to configure BGP route dampening:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure IPv6 BGP route dampening parameters	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*]*	Optional Not configured by default.

Configuring IPv6 BGP Route Attributes

This section describes how to use IPv6 BGP route attributes to modify BGP routing policy. These attributes are:

- IPv6 BGP protocol preference
- Default LOCAL_PREF attribute
- MED attribute
- NEXT_HOP attribute
- AS_PATH attribute

Prerequisites

Before configuring this task, you have:

- Enabled IPv6 function
- Configured IPv6 BGP basic functions

Configuring IPv6 BGP Preference and Default LOCAL_PREF and NEXT_HOP Attributes

Follow these steps to perform this configuration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure preference values for IPv6 BGP external, internal, local routes	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	Optional The default preference values of external, internal and local routes are 255, 255, 130 respectively.
Configure the default local preference	**default local-preference** *value*	Optional The *value* defaults to 100.
Advertise routes to an IPv6 peer/peer group with the local router as the next hop	**peer** { *ipv6-group-name* | *ipv6-address* } **next-hop-local**	Required By default, IPv6 BGP specifies the local router as the next hop for routes sent to an IPv6 eBGP peer/peer group, but not for routes sent to an IPv6 iBGP peer/peer group.

Note

- To make sure an iBGP peer can find the correct next hop, you can configure routes advertised to the IPv6 iBGP peer/peer group to use the local router as the next hop. If BGP load balancing is configured, the local router specifies itself as the next hop of routes sent to an IPv6 iBGP peer/peer group regardless of whether the peer next-hop-local command is configured.
- In a “third party next hop" network, that is, the two IPv6 eBGP peers reside in a common broadcast subnet, the router does not specify itself as the next hop for routes sent to the IPv6 eBGP peer/peer group by default, unless the peer next-hop-local command is configured.

Configuring the MED Attribute

Follow these steps to configure the MED attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Configure a default MED value	**default med** *med-value*	Optional Defaults to 0
Enable the comparison of MED for routes from different eBGP peers	**compare-different-as-med**	Optional Not enabled by default.
Enable the comparison of MED for routes from each AS	**bestroute compare-med**	Optional Disabled by default

To do…	Use the command…	Remarks
Enable the comparison of MED for routes from confederation peers	**bestroute med-confederation**	Optional Disabled by default

Configuring the AS_PATH Attribute

Follow these steps to configure the AS_PATH attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Allow the local AS number to appear in AS_PATH of routes from a peer/peer group and specify the repeat times	**peer** { *ipv6-group-name* | *ipv6-address* } **allow-as-loop** [*number*]	Optional Not allowed by default
Specify a fake AS number for an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **fake-as** *as-number*	Optional Not specified by default.
Disable IPv6 MBGP from considering the AS_PATH during best route selection	**bestroute as-path-neglect**	Optional Enabled by default.
Configure to carry only the public AS number in updates sent to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **public-as-only**	Optional By default, IPv6 BGP updates carry private AS number.
Substitute the local AS number for the AS number of an IPv6 peer/peer group identified in the AS_PATH attribute	**peer** { *ipv6-group-name* | *ipv6-address* } **substitute-as**	Optional Not substituted by default

Tuning and Optimizing IPv6 BGP Networks

This section describes configurations of IPv6 BGP timers, IPv6 BGP connection soft reset and the maximum number of load balanced routes.

- IPv6 BGP timers

After establishing an IPv6 BGP connection, two routers send keepalive messages periodically to each other to keep the connection. If a router receives no keepalive message from the peer after the holdtime elapses, it tears down the connection.

When establishing an IPv6 BGP connection, the two parties compare their holdtimes, taking the shorter one as the common holdtime. If the holdtime is 0, neither keepalive massage is sent, nor holdtime is checked.

- IPv6 BGP connection soft reset

After modifying a route selection policy, you have to reset IPv6 BGP connections to make the new one take effect, causing a short time disconnection. The current IPv6 BGP implementation supports the

route-refresh feature that enables dynamic IPv6 BGP routing table refresh without needing to disconnect IPv6 BGP links.

With this feature enabled on all IPv6 BGP routers in a network, when a routing policy modified on a router, the router advertises a route-refresh message to its peers, which then send their routing information back to the router. Therefore, the local router can perform dynamic routing information update and apply the new policy without tearing down connections.

If a peer not supporting route-refresh exists in the network, you need to configure the **peer keep-all-routes** command on the router to save all routes from the peer. When the routing policy is changed, the system will update the IPv6 BGP routing table and apply the new policy.

Prerequisites

Before configuring IPv6 BGP timers, you need to:

- Enable IPv6
- Configure IPv6 BGP basic functions

Configuring IPv6 BGP Timers

Follow these steps to configure IPv6 BGP timers:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter BGP view		**bgp** *as-number*	—
Enter IPv6 address family view		**ipv6-family**	—
Configure IPv6 BGP timers	Specify keepalive interval and holdtime	**timer keepalive** *keepalive* **hold** *holdtime*	Optional The keepalive interval defaults to 60 seconds, holdtime defaults to 180 seconds.
	Configure keepalive interval and holdtime for an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **timer keepalive** *keepalive* **hold** *holdtime*	
Configure the interval for sending the same update to an IPv6 peer/peer group		**peer** { *ipv6-group-name* | *ipv6-address* } **route-update-interval** *interval*	Optional The interval for sending the same update to an iBGP peer or an eBGP peer defaults to 15 seconds or 30 seconds

Note

- Timers configured using the **timer** command have lower priority than timers configured using the **peer timer** command.
- The holdtime interval must be at least three times the keepalive interval.

Configuring IPv6 BGP Soft Reset

Enable route refresh

Follow these steps to enable route refresh:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Enable route refresh	**peer** { *ipv6-group-name* | *ipv6-address* } **capability-advertise route-refresh**	Optional Enabled by default.

Perform manual soft-reset

Follow these steps to perform manual soft reset:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Save all routes from an IPv6 peer/peer group, not letting them go through the inbound policy	**peer** { *ipv6-group-name* | *ipv6-address* } **keep-all-routes**	Optional Not saved by default.
Return to user view	**return**	Required
Soft-reset BGP connections manually	**refresh bgp ipv6** { **all** | *ipv6-address* | **group** *ipv6-group-name* | **external** | **internal** } { **export** | **import** }	

Note

If the **peer keep-all-routes** command is used, all routes from the peer/peer group will be saved regardless of whether the filtering policy is available. These routes will be used to generate IPv6 BGP routes after soft-reset is performed.

Configuring the Maximum Number of Load-Balanced Routes

Follow these steps to configure the maximum number of load balanced routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—

To do…	Use the command…	Remarks
Configure the maximum number of load balanced routes	**balance** *number*	Required By default, no load balancing is enabled.

Configuring a Large Scale IPv6 BGP Network

In a large-scale IPv6 BGP network, configuration and maintenance become no convenient due to too many peers. In this case, configuring peer groups makes management easier and improves route distribution efficiency. Peer group includes iBGP peer group, where peers belong to the same AS, and eBGP peer group, where peers belong to different ASs. If peers in an eBGP group belong to the same external AS, the eBGP peer group is a pure eBGP peer group, and if not, a mixed eBGP peer group.

In a peer group, all members enjoy a common policy. Using the community attribute can make a set of IPv6 BGP routers in multiple ASs enjoy the same policy, because sending of community between IPv6 BGP peers is not limited by AS.

To guarantee connectivity between iBGP peers, you need to make them fully meshed, but it becomes unpractical when there are too many iBGP peers. Using route reflectors or confederation can solve it. In a large-scale AS, both of them can be used.

Confederation configuration of IPv6 BGP is identical to that of BGP4, so it is not mentioned here. The following describes:

- Configuring IPv6 BGP peer groups
- Configuring IPv6 BGP community
- Configuring IPv6 BGP route reflectors

Prerequisites

Before configuring IPv6 BGP peer groups, you need to:

- Make peer nodes accessible to each other at the network layer
- Enable BGP and configure a router ID.

Configuring IPv6 BGP Peer Group

Configuring an iBGP peer group

Follow these steps to configure an iBGP group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Create an iBGP peer group	**group** *ipv6-group-name* [**internal**]	Required
Add a peer into the group	**peer** *ipv6-address* **group** *ipv6-group-name* [**as-number** *as-number*]	Required Not added by default

Creating a pure eBGP peer group

Follow these steps to configure a pure eBGP group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Create an eBGP peer group	**group** *ipv6-group-name* **external**	Required
Configure the AS number for the peer group	**peer** *ipv6-group-name* **as-number** *as-number*	Required Not configured by default.
Add an IPv6 peer into the peer group	**peer** *ipv6-address* **group** *ipv6-group-name*	Required Not added by default

Note

- To create a pure eBGP peer group, you need to specify an AS number for the peer group.
- If a peer was added into an eBGP peer group, you cannot specify any AS number for the peer group.

Creating a mixed eBGP peer group

Follow these steps to create a mixed eBGP peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Create an eBGP peer group	**group** *ipv6-group-name* **external**	Required
Specify the AS number of an IPv6 peer	**peer** *ipv6-address* **as-number** *as-number*	Required Not specified by default.
Add the IPv6 peer into the peer group	**peer** *ipv6-address* **group** *ipv6-group-name*	Required Not added by default

Note

When creating a mixed eBGP peer group, you need to create a peer and specify its AS number that can be different from AS numbers of other peers, but you cannot specify AS number for the eBGP peer group.

Configuring IPv6 BGP Community

Advertise community attribute to an IPv6 peer/peer group

Follow these steps to advertise community attribute to an IPv6 peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Advertise community attribute to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-community**	Required Not advertised by default.
Advertise extended community attribute to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-ext-community**	Required Not advertised by default.

Apply a routing policy to routes advertised to a peer/peer group

Follow these steps to apply a routing policy to routes advertised to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Apply a routing policy to routes advertised to an IPv6 peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	Required Not applied by default.

Note

- When configuring IPv6 BGP community, you need to configure a routing policy to define the community attribute, and apply the routing policy to route advertisement.
- For routing policy configuration, refer to Routing Policy Configuration in the IP Routing Volume.

Configuring an IPv6 BGP Route Reflector

Follow these steps to configure an IPv6 BGP route reflector:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—

To do…	Use the command…	Remarks
Configure the router as a route reflector and specify an IPv6 peer/peer group as a client	**peer** { *ipv6-group-name* | *ipv6-address* } **reflect-client**	Required Not configured by default.
Enable route reflection between clients	**reflect between-clients**	Optional Enabled by default.
Configure the cluster ID of the route reflector	**reflector cluster-id** *cluster-id*	Optional By default, a route reflector uses its router ID as the cluster ID

Note

- In general, since the route reflector forwards routing information between clients, it is not required to make clients of a route reflector fully meshed. If clients are fully meshed, it is recommended to disable route reflection between clients to reduce routing costs.
- If a cluster has multiple route reflectors, you need to specify the same cluster ID for these route reflectors to avoid routing loops.

Displaying and Maintaining IPv6 BGP

Displaying BGP

To do…	Use the command…	Remarks
Display IPv6 BGP peer group information	**display bgp ipv6 group** [*ipv6-group-name*]	Available in any view
Display IPv6 BGP advertised routing information	**display bgp ipv6 network**	
Display IPv6 BGP AS path information	**display bgp ipv6 paths** [*as-regular-expression*]	
Display IPv6 BGP peer/peer group information	**display bgp ipv6 peer** [*group-name* **log-info** | *ipv4-address* **verbose** | *ipv6-address* { **log-info** | **verbose** } | **verbose**]	
Display IPv6 BGP routing table information	**display bgp ipv6 routing-table** [*ipv6-address prefix-length*]	
Display IPv6 BGP routing information matching an AS path ACL	**display bgp ipv6 routing-table as-path-acl** *as-path-acl-number*	
Display IPv6 BGP routing information with the specified community attribute	**display bgp ipv6 routing-table community** [*aa:nn*<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**]* [**whole-match**]	
Display IPv6 BGP routing information matching an IPv6 BGP community list	**display bgp ipv6 routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	
Display dampened IPv6 BGP routing information	**display bgp ipv6 routing-table dampened**	
Display IPv6 BGP dampening parameter information	**display bgp ipv6 routing-table dampening parameter**	
Display IPv6 BGP routing information originated from different ASs	**display bgp ipv6 routing-table different-origin-as**	
Display IPv6 BGP routing flap statistics	**display bgp ipv6 routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *network-address* [*prefix-length* [**longer-match**]]]	
Display BGP routing information to or from an IPv4 or IPv6 peer	**display bgp ipv6 routing-table peer** { *ipv4-address* | *ipv6-address* } { **advertised-routes** | **received-routes** } [*network-address prefix-length* | **statistic**]	
Display IPv6 BGP routing information matching a regular expression	**display bgp ipv6 routing-table regular-expression** *as-regular-expression*	
Display IPv6 BGP routing statistics	**display bgp ipv6 routing-table statistic**	

Resetting IPv6 BGP Connections

To do…	Use the command…	Remarks
Perform soft reset on IPv6 BGP connections	**refresh bgp ipv6** { *ipv4-address* | *ipv6-address* | **all** | **external** | **group** *ipv6-group-name* | **internal** } { **export** | **import** }	Available in user view
Reset IPv6 BGP connections	**reset bgp ipv6** { *as-number* | *ipv4-address* | *ipv6-address* [**flap-info**] | **all** | **external** | **group** *group-name* | **internal** }	

Clearing IPv6 BGP Information

To do…	Use the command…	Remarks
Clear dampened IPv6 BGP routing information and release suppressed routes	**reset bgp ipv6 dampening** [*ipv6-address prefix-length*]	Available in user view
Clear IPv6 BGP route flap information	**reset bgp ipv6 flap-info** [*ipv6-address/prefix-length* | **as-path-acl** *as-path-acl-number* | **regexp** *as-path-regexp*]	

IPv6 BGP Configuration Examples

Note

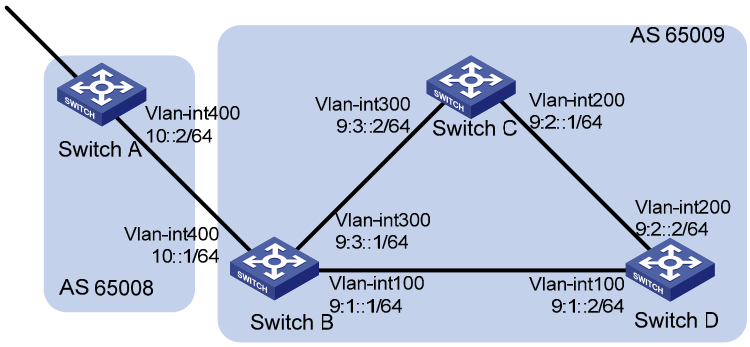
Some examples for IPv6 BGP configuration are similar to those of BGP4, so refer to *BGP Configuration* in the *IP Routing Volume* for related information.

IPv6 BGP Basic Configuration

Network requirements

In the following figure are all IPv6 BGP switches. Between Switch A and Switch B is an eBGP connection. Switch B, Switch C and Switch D are fully meshed through iBGP connections.

Figure 1-1 IPv6 BGP basic configuration network diagram



Configuration procedure

1) Configure IPv6 addresses for interfaces (omitted)
2) Configure iBGP connections

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 9:1::2 as-number 65009
[SwitchB-bgp-af-ipv6] peer 9:3::2 as-number 65009
[SwitchB-bgp-af-ipv6] quit
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] ipv6-family
[SwitchC-bgp-af-ipv6] peer 9:3::1 as-number 65009
[SwitchC-bgp-af-ipv6] peer 9:2::2 as-number 65009
[SwitchC-bgp-af-ipv6] quit
[SwitchC-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] bgp 65009
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] ipv6-family
[SwitchD-bgp-af-ipv6] peer 9:1::1 as-number 65009
[SwitchD-bgp-af-ipv6] peer 9:2::1 as-number 65009
[SwitchD-bgp-af-ipv6] quit
[SwitchD-bgp] quit
```

3) Configure the eBGP connection

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] ipv6-family
[SwitchA-bgp-af-ipv6] peer 10::1 as-number 65009
[SwitchA-bgp-af-ipv6] quit
[SwitchA-bgp] quit
```

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 10::2 as-number 65008
```

Display IPv6 peer information on Switch B.

```
[SwitchB] display bgp ipv6 peer

 BGP local router ID : 2.2.2.2
 Local AS number : 65009
 Total number of peers : 3                 Peers in established state : 3

  Peer       V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  10::2      4 65008        3        3     0       0 00:01:16 Established
  9:3::2     4 65009        2        3     0       0 00:00:40 Established
  9:1::2     4 65009        2        4     0       0 00:00:19 Established
```

Display IPv6 peer information on Switch C.

```
[SwitchC] display bgp ipv6 peer

 BGP local router ID : 3.3.3.3
 Local AS number : 65009
 Total number of peers : 2                 Peers in established state : 2

  Peer       V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  9:3::1     4 65009        4        4     0       0 00:02:18 Established
  9:2::2     4 65009        4        5     0       0 00:01:52 Established
```

Switch A and B has established an eBGP connection; Switch B, C and D have established iBGP connections with each other.

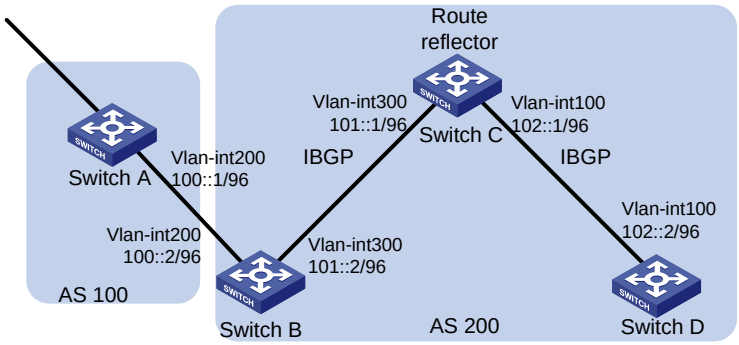
IPv6 BGP Route Reflector Configuration

Network requirements

As shown in the following figure, Switch B receives an eBGP update and sends it to Switch C, which is configured as a route reflector with two clients: Switch B and Switch D.

Switch B and Switch D need not establish an iBGP connection because Switch C reflects updates between them.

Figure 1-2 Network diagram for IPv6 BGP route reflector configuration



Configuration procedure

1) Configure IPv6 addresses for VLAN interfaces (omitted)
2) Configure IPv6 BGP basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] ipv6-family
[SwitchA-bgp-af-ipv6] peer 100::2 as-number 200
[SwitchA-bgp-af-ipv6] network 1:: 64
```

#Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 100::1 as-number 100
[SwitchB-bgp-af-ipv6] peer 101::1 as-number 200
[SwitchB-bgp-af-ipv6] peer 101::1 next-hop-local
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] bgp 200
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] ipv6-family
[SwitchC-bgp-af-ipv6] peer 101::2 as-number 200
[SwitchC-bgp-af-ipv6] peer 102::2 as-number 200
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] bgp 200
[SwitchD-bgp] router-id 4.4.4.4
```

```
[SwitchD-bgp] ipv6-family
[SwitchD-bgp-af-ipv6] peer 102::1 as-number 200
```

3) Configure route reflector

Configure Switch C as a route reflector, Switch B and Switch D as its clients.

```
[SwitchC-bgp-af-ipv6] peer 101::2 reflect-client
[SwitchC-bgp-af-ipv6] peer 102::2 reflect-client
```

Use the **display bgp ipv6 routing-table** command on Switch B and Switch D respectively, you can find both of them have learned the network 1::/64.

Troubleshooting IPv6 BGP Configuration

No IPv6 BGP Peer Relationship Established

Symptom

Display BGP peer information using the **display bgp ipv6 peer** command. The state of the connection to the peer cannot become established.

Analysis

To become IPv6 BGP peers, any two routers need to establish a TCP session using port 179 and exchange open messages successfully.

Processing steps

1) Use the **display current-configuration** command to verify the peer's AS number.
2) Use the **display bgp ipv6 peer** command to verify the peer's IPv6 address.
3) If the loopback interface is used, check whether the **peer connect-interface** command is configured.
4) If the peer is not directly connected, check whether the **peer ebgp-max-hop** command is configured.
5) Check whether a route to the peer is available in the routing table.
6) Use the **ping** command to check connectivity.
7) Use the **display tcp ipv6 status** command to check the TCP connection.
8) Check whether an ACL for disabling TCP port 179 is configured.

Table of Contents

1 Route Policy Configuration

A route policy is used on a router for route filtering and attributes modification when routes are received, advertised, or redistributed.

When configuring route policy, go to these sections for information you are interested in:

Note

Route policy in this chapter involves both IPv4 route policy and IPv6 route policy.

Introduction to Route Policy

Route Policy

A route policy is used on a router for route filtering and attributes modification when routes are received, advertised, or redistributed.

To configure a route policy, you need to define some filters based on the attributes of routing information, such as destination address, advertising router's address and so on. The filters can be set beforehand and then applied to the route policy.

Filters

There are six types of filters: ACL, IP prefix list, AS path ACL, community list, extended community list and route policy.

ACL

ACL involves IPv4 ACL and IPv6 ACL. An ACL is configured to match the destinations or next hops of routing information.

For ACL configuration, refer to *ACL configuration* in the *Security Volume*.

IP prefix list

IP prefix list involves IPv4 and IPv6 prefix list.

An IP prefix list is configured to match the destination address of routing information. Moreover, you can use the **gateway** option to allow only routing information from certain routers to be received. For **gateway** option information, refer to *RIP Commands* and *OSPF Commands* in the *IP Routing Volume.*

An IP prefix list, identified by name, can comprise multiple items. Each item, identified by an index number, can specify a prefix range to match. An item with a smaller index number is matched first. If one item is matched, the IP prefix list is passed, and the packet will not go to the next item.

AS-PATH list

An AS-PATH list, configured based on the BGP AS PATH attribute, can only be used to match BGP routing information.

Community list

A community list, configured based on the BGP community attribute, can only be used to match BGP routing information.

Extended community list

An extended community list, configured based on the BGP extended community attribute (Route-Target for VPN, and Source of Origin), can only be used to match BGP routing information.

Route policy

A route policy is used to match routing information and modify the attributes of permitted routes. It can reference the above mentioned filters to define its own match criteria.

A route policy can comprise multiple nodes, which are in logic OR relationship. Each route policy node is a match unit, and a node with a smaller number is matched first. Once a node is matched, the route policy is passed and the packet will not go to the next node.

A route policy node comprises a set of **if-match** and **apply** clauses. The **if-match** clauses define the match criteria. The matching objects are some attributes of routing information. The **if-match** clauses of a route policy node is in logical AND relationship. That is, a packet must match all the **if-match** clauses of the node to pass it. The **apply** clauses of the node specify the actions to be taken on the permitted packets, such as route attribute modification.

Route Policy Application

A route policy is applied on a router to filter routes when they are received, advertised or redistributed and to modify some attributes of permitted routes.

Route Policy Configuration Task List

Complete the following tasks to configure a route policy:

Task	
Defining Filters	Defining an IP-prefix List
	Defining an AS Path List
	Defining a Community List
	Defining an Extended Community List

Task	
Configuring a Route Policy	Creating a Route Policy
	Defining if-match Clauses
	Defining apply Clauses

Defining Filters

Prerequisites

Before configuring this task, you need to decide on:

- IP-prefix list name
- Matching address range
- Extcommunity list sequence number

Defining an IP-prefix List

Define an IPv4 prefix list

Identified by name, an IPv4 prefix list can comprise multiple items. Each item specifies a prefix range to match and is identified by an index number.

An item with a smaller index number is matched first. If one item is matched, the IP prefix list is passed, and the routing information will not go to the next item.

Follow these steps to define an IPv4 prefix list:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define an IPv4 prefix list	**ip ip-prefix** *ip-prefix-name* [**index** *index-number*] { **permit** | **deny** } *ip-address mask-length* [**greater-equal** *min-mask-length*] [**less-equal** *max-mask-length*]	Required Not defined by default.

Note

If all the items are set to the **deny** mode, no routes can pass the IPv4 prefix list. Therefore, you need to define the **permit** 0.0.0.0 0 **less-equal** 32 item following multiple **deny** items to allow other IPv4 routing information to pass.

For example, the following configuration filters routes 10.1.0.0/16, 10.2.0.0/16 and 10.3.0.0/16, but allows other routes to pass.

```
<Sysname> system-view
[Sysname] ip ip-prefix abc index 10 deny 10.1.0.0 16
[Sysname] ip ip-prefix abc index 20 deny 10.2.0.0 16
[Sysname] ip ip-prefix abc index 30 deny 10.3.0.0 16
[Sysname] ip ip-prefix abc index 40 permit 0.0.0.0 0 less-equal 32
```

Define an IPv6 prefix list

Identified by name, each IPv6 prefix list can comprise multiple items. Each item specifies a prefix range to match and is identified by an index number.

An item with a smaller index number is matched first. If one item is matched, the IPv6 prefix list is passed, and the routing information will not go to the next item.

Follow these steps to define an IPv6 prefix list:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define an IPv6 prefix list	**ip ipv6-prefix** *ipv6-prefix-name* [**index** *index-number*] { **deny** | **permit** } *ipv6-address prefix-length* [**greater-equal** *min-prefix-length*] [**less-equal** *max-prefix-length*]	Required Not defined by default.

Note

If all items are set to the **deny** mode, no routes can pass the IPv6 prefix list. Therefore, you need to define the **permit** :: 0 **less-equal** 128 item following multiple **deny** items to allow other IPv6 routing information to pass.

For example, the following configuration filters routes 2000:1::/48, 2000:2::/48 and 2000:3::/48, but allows other routes to pass.

```
<Sysname> system-view
[Sysname] ip ipv6-prefix abc index 10 deny 2000:1:: 48
[Sysname] ip ipv6-prefix abc index 20 deny 2000:2:: 48
[Sysname] ip ipv6-prefix abc index 30 deny 2000:3:: 16
[Sysname] ip ipv6-prefix abc index 40 permit :: 0 less-equal 128
```

Defining an AS Path List

You can define multiple items for an AS path list that is identified by number. The relation between items is logical OR, that is, if a route matches one of these items, it passes the AS path list.

Follow these steps to define an AS path list:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define an AS path ACL	**ip as-path** *as-path-number* { **deny** | **permit** } *regular-expression*	Required Not defined by default.

Defining a Community List

You can define multiple items for a community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the community list.

Follow these steps to define a community list:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Define a community list	Define a basic community list	**ip community-list** *basic-comm-list-num* { **deny** | **permit** } [*community-number-list*] [**internet** | **no-advertise** | **no-export** | **no-export-subconfed**] *	Required to define either; Not defined by default.
	Define an advanced community list	**ip community-list** *adv-comm-list-num* { **deny** | **permit** } *regular-expression*	

Defining an Extended Community List

You can define multiple items for an extended community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the extended community list.

Follow these steps to define an extended community list:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define an extended community list	**ip extcommunity-list** *ext-comm-list-number* { **deny** | **permit** } { **rt** *route-target* }&<1-16>	Required Not defined by default

Configuring a Route Policy

A route policy is used to filter routing information, and modify attributes of matching routing information. The match criteria of a route policy can be configured by referencing filters above mentioned.

A route policy can comprise multiple nodes, and each route policy node contains:

- **if-match** clauses: Define the match criteria that routing information must satisfy. The matching objects are some attributes of routing information.
- **apply** clauses: Specify the actions to be taken on routing information that has satisfied the match criteria, such as route attribute modification.

Prerequisites

Before configuring this task, you need to configure:

- Filters
- Routing protocols

You also need to decide on:

- Name of the route policy, and node numbers
- Match criteria
- Attributes to be modified

Creating a Route Policy

Follow these steps to create a route policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a route policy, specify a node for it and enter route policy node view	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required

Note

- If a route policy node has the **permit** keyword specified, routing information matching all the **if-match** clauses of the node will be handled using the **apply** clauses of this node, without needing to match against the next node. If routing information does not match the node, it will go to the next node for a match.
- If a route policy node has the **deny** keyword specified, the **apply** clauses of the node will not be executed. When routing information matches all the **if-match** clauses of the node, it cannot pass the node, or go to the next node. If route information cannot match all the **if-match** clauses of the node, it will go to the next node for a match.
- When a route policy has more than one node, at least one node should be configured with the **permit** keyword. If the route policy is used to filter routing information, routing information that does not meet any node cannot pass the route policy. If all nodes of the route policy are set with the **deny** keyword, no routing information can pass it.

Defining if-match Clauses

Follow these steps to define if-match clauses for a route-policy node:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter route policy node view		**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required
Define match criteria for IPv4 routes	Match IPv4 routing information specified in the ACL	**if-match acl** *acl-number*	Optional Not configured by default.
	Match IPv4 routing information specified in the IP prefix list	**if-match ip-prefix** *ip-prefix-name*	
	Match IPv4 routing information whose next hop or source is specified in the ACL or IP prefix list	**if-match ip** { **next-hop** | **route-source** } { **acl** *acl-number* | **ip-prefix** *ip-prefix-name* }	Optional Not configured by default.

To do…	Use the command…	Remarks
Match IPv6 routing information whose next hop or source is specified in the ACL or IP prefix list	**if-match ipv6** { **address** | **next-hop** | **route-source** } { **acl** *acl-number* | **prefix-list** *ipv6-prefix-name* }	Optional Not configured by default.
Match BGP routing information whose AS path attribute is specified in the AS path list (s)	**if-match as-path** *AS-PATH-number*&<1-16>	Optional Not configured by default.
Match BGP routing information whose community attribute is specified in the community list(s)	**if-match community** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	Optional Not configured by default.
Match routes having the specified cost	**if-match cost** *value*	Optional Not configured by default.
Match BGP routing information whose extended community attribute is specified in the extended community list(s)	**if-match extcommunity** *ext-comm-list-number*&<1-16>	Optional Not configured by default.
Match routing information having specified outbound interface(s)	**if-match interface** { *interface-type interface-number* }&<1-16>	Optional Not configured by default.
Match routing information having the specified route type	**if-match route-type** { **internal** | **external-type1** | **external-type2** | **external-type1or2** | **is-is-level-1** | **is-is-level-2** | **nssa-external-type1** | **nssa-external-type2** | **nssa-external-type1or2** } *	Optional Not configured by default.
Match RIP, OSPF, and IS-IS routing information having the specified tag value	**if-match tag** *value*	Optional Not configured by default.

Note

- The **if-match** clauses of a route policy node are in logic AND relationship, namely, routing information has to satisfy all its **if-match** clauses before being executed with its **apply** clauses.
- You can specify no or multiple **if-match** clauses for a route policy node. If no **if-match** clause is specified, and the route policy node is in **permit** mode, all routing information can pass the node. If it is in **deny** mode, no routing information can pass it.
- An ACL specified in an **if-match** clause should be a non-VPN ACL.
- The **if-match** commands for matching IPv4 destination, next hop and source address are different from those for matching IPv6 ones.

Defining apply Clauses

Follow these steps to define **apply** clauses for a route policy:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter route policy node view		**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required Not created by default.
Set the AS-PATH attribute for BGP routing information		**apply as-path** *as-number*&<1-10> [**replace**]	Optional Not set by default.
Delete the community attribute of BGP routing information using the community list		**apply comm-list** *comm-list-number* **delete**	Optional Not configured by default.
Set the community attribute for BGP routing information		**apply community** { **none** | **additive** | { *community-number*&<1-16> | *aa:nn*&<1-16> | **internet** | **no-export-subconfed** | **no-export | no-advertise** } * [**additive**] }	Optional Not set by default.
Set a cost for routing information		**apply cost** [**+** | **-**] *value*	Optional Not set by default.
Set a cost type for routing information		**apply cost-type** [**external** | **internal** | **type-1** | **type-2**]	Optional Not set by default.
Set the extended community attribute for BGP routing		**apply extcommunity** { **rt** { *as-number*:*nn* | *ip-address*:*nn* } }&<1-16> [**additive**]	Optional Not set by default.
Set the next hop	for IPv4 routes	**apply ip-address next-hop** *ip-address*	Optional Not set by default. The setting does not apply to redistributed routing information.
	for IPv6 routes	**apply ipv6 next-hop** *ipv6-address*	Optional Not set by default. The setting does not apply to redistributed routing information.
Inject routing information to a specified ISIS level		**apply isis** { **level-1** | **level-1-2** | **level-2** }	Optional Not configured by default.
Set the local preference for BGP routing information		**apply local-preference** *preference*	Optional Not set by default.
Set the origin attribute for BGP routing information		**apply origin** { **igp** | **egp** *as-number* | **incomplete** }	Optional Not set by default.
Set the preference for the routing protocol		**apply preference** *preference*	Optional Not set by default.

To do…	Use the command…	Remarks
Set a preferred value for BGP routing information	**apply preferred-value** *preferred-value*	Optional Not set by default.
Set a tag value for RIP, OSPF or IS-IS routing information	**apply tag** *value*	Optional Not set by default.

Note

- The difference between IPv4 and IPv6 **apply** clauses is the command for setting the next hop for routing information.
- The **apply ip-address next-hop** and **apply ipv6 next-hop** commands do not apply to redistributed IPv4 and IPv6 routes respectively.

Displaying and Maintaining the Route Policy

To do…	Use the command…	Remarks
Display BGP AS-PATH list information	**display ip as-path** [*as-path-number*]	Available in any view
Display BGP community list information	**display ip community-list** [*basic-community-list-number* | *adv-community-list-number*]	
Display BGP extended community list information	**display ip extcommunity-list** [*ext-comm-list-number*]	
Display IPv4 prefix list statistics	**display ip ip-prefix** [*ip-prefix-name*]	
Display IPv6 prefix list statistics	**display ip ipv6-prefix** [*ipv6-prefix-name*]	
Display route policy information	**display route-policy** [*route-policy-name*]	
Clear IPv4 prefix list statistics	**reset ip ip-prefix** [*ip-prefix-name*]	Available in user view
Clear IPv6 prefix list statistics	**reset ip ipv6-prefix** [*ipv6-prefix-name*]	

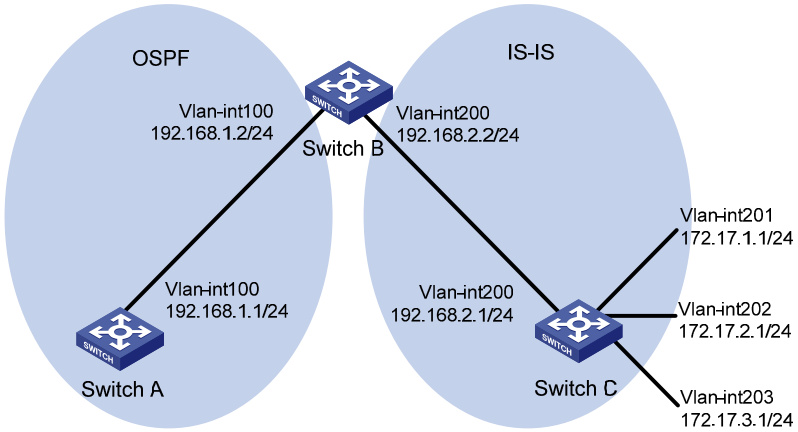
Route Policy Configuration Example

Applying a Route Policy to IPv4 Route Redistribution

Network Requirements

- As shown in the following figure, Switch B exchanges routing information with Switch A using OSPF, and with Switch C using IS-IS.
- On Switch B, enable route redistribution from IS-IS to OSPF and apply a route policy to set the cost of route 172.17.1.0/24 to 100, and the tag of route 172.17.2.0/24 to 20.

Figure 1-1 Network diagram for route policy application to route redistribution



Configuration procedure

1) Specify IP addresses for interfaces (omitted).
2) Configure IS-IS.

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis
[SwitchC-isis-1] is-level level-2
[SwitchC-isis-1] network-entity 10.0000.0000.0001.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 201
[SwitchC-Vlan-interface201] isis enable
[SwitchC-Vlan-interface201] quit
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] isis enable
[SwitchC-Vlan-interface202] quit
[SwitchC] interface vlan-interface 203
[SwitchC-Vlan-interface203] isis enable
[SwitchC-Vlan-interface203] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis
[SwitchB-isis-1] is-level level-2
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable
[SwitchB-Vlan-interface200] quit
```

3) Configure OSPF and route redistribution

Configure OSPF on Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

On Switch B, configure OSPF and enable route redistribution from IS-IS.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] import-route isis 1
[SwitchB-ospf-1] quit
```

Display the OSPF routing table on Switch A to view redistributed routes.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop        AdvRouter     Area
 192.168.1.0/24     1562     Stub    192.168.1.1    192.168.1.1   0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag        NextHop        AdvRouter
 172.17.1.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.2.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 172.17.3.0/24      1        Type2   1          192.168.1.2    192.168.2.2
 192.168.2.0/24     1        Type2   1          192.168.1.2    192.168.2.2

 Total Nets: 5
 Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0
```

4) Configure filtering lists

Configure ACL 2002 to permit route 172.17.2.0/24.

```
[SwitchB] acl number 2002
[SwitchB-acl-basic-2002] rule permit source 172.17.2.0 0.0.0.255
[SwitchB-acl-basic-2002] quit
```

Configure IP prefix list prefix-a to permit route 172.17.1.0/24.

```
[SwitchB] ip ip-prefix prefix-a index 10 permit 172.17.1.0 24
```

5) Configure a route policy.

```
[SwitchB] route-policy isis2ospf permit node 10
[SwitchB-route-policy] if-match ip-prefix prefix-a
[SwitchB-route-policy] apply cost 100
[SwitchB-route-policy] quit
[SwitchB] route-policy isis2ospf permit node 20
```

```
[SwitchB-route-policy] if-match acl 2002
[SwitchB-route-policy] apply tag 20
[SwitchB-route-policy] quit
[SwitchB] route-policy isis2ospf permit node 30
[SwitchB-route-policy] quit
```

6) Apply the route policy to route redistribution.

On Switch B, apply the route policy when redistributing routes.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route isis 1 route-policy isis2ospf
[SwitchB-ospf-1] quit
```

Display the OSPF routing table on Switch A. The cost of route 172.17.1.0/24 is 100, the tag of route 172.17.1.0/24 is 20.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
                  Routing Tables

 Routing for Network
 Destination        Cost     Type    NextHop         AdvRouter     Area
 192.168.1.0/24       1     Transit    192.168.1.1     192.168.1.1   0.0.0.0

 Routing for ASEs
 Destination        Cost     Type    Tag         NextHop       AdvRouter
 172.17.1.0/24      100      Type2   1           192.168.1.2   192.168.2.2
 172.17.2.0/24      1        Type2   20          192.168.1.2   192.168.2.2
 172.17.3.0/24      1        Type2   1           192.168.1.2   192.168.2.2
 192.168.2.0/24     1        Type2   1           192.168.1.2   192.168.2.2

 Total Nets: 5
 Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0
```

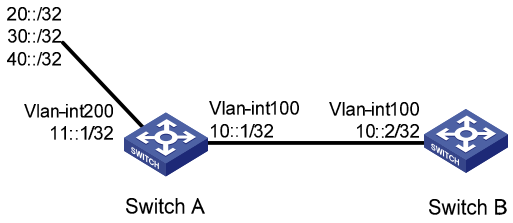
Applying a Route Policy to IPv6 Route Redistribution

Network requirements

As shown in the following figure:

- Enable RIPng on Switch A and Switch B.
- On Switch A, configure three static routes, and apply a route policy to static route redistribution to permit routes 20::0/32 and 40::0/32, and deny route 30::0/32.
- Display RIPng routing table information on Switch B to verify the configuration.

Figure 1-2 Network diagram for route policy application to route redistribution



Configuration procedure

1) Configure Switch A.

Configure IPv6 addresses for VLAN-interface 100 and VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 address 10::1 32
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ipv6 address 11::1 32
[SwitchA-Vlan-interface200] quit
```

Enable RIPng on VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
```

Configure three static routes.

```
[SwitchA] ipv6 route-static 20:: 32 11::2
[SwitchA] ipv6 route-static 30:: 32 11::2
[SwitchA] ipv6 route-static 40:: 32 11::2
```

Configure a route policy.

```
[SwitchA] ip ipv6-prefix a index 10 permit 30:: 32
[SwitchA] route-policy static2ripng deny node 0
[SwitchA-route-policy] if-match ipv6 address prefix-list a
[SwitchA-route-policy] quit
[SwitchA] route-policy static2ripng permit node 10
[SwitchA-route-policy] quit
```

Enable RIPng and apply the route policy to static route redistribution.

```
[SwitchA] ripng
[SwitchA-ripng-1] import-route static route-policy static2ripng
```

2) Configure Switch B.

Configure the IPv6 address for VLAN-interface 100.

```
[SwitchB] ipv6
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipv6 address 10::2 32
```

Enable RIPng on VLAN-interface 100.

```
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Enable RIPng.

```
[SwitchB] ripng
```

Display RIPng routing table information.

```
[SwitchB-ripng-1] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::7D58:0:CA03:1  on Vlan-interface 100
 Dest 10::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 18 Sec
 Dest 20::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 8 Sec
 Dest 40::/32,
     via FE80::7D58:0:CA03:1, cost  1, tag 0, A, 3 Sec
```

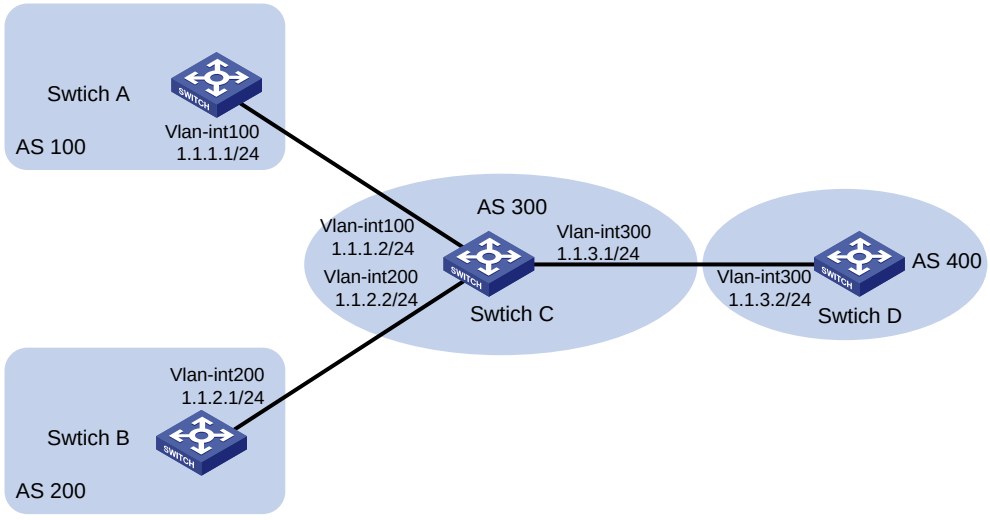
Applying a Route Policy to Filter Received BGP Routes

Network requirements

As shown in the following figure:

- All the switches run BGP. Switch C establishes eBGP connections with other switches.
- Configure a route policy on Switch D to reject routes from AS 200.

Figure 1-3 Route policy configuration to filter received BGP routes (on switches)



Configuration procedure

1) Configure IP addresses for the interfaces (omitted).
2) Configure BGP.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
```

```
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 1.1.1.2 as-number 300
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 1.1.2.2 as-number 300
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 300
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 1.1.1.1 as-number 100
[SwitchC-bgp] peer 1.1.2.1 as-number 200
[SwitchC-bgp] peer 1.1.3.2 as-number 400
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 400
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] peer 1.1.3.1 as-number 300
[SwitchD-bgp] quit
```

On Switch A, inject routes 4.4.4.4/24, 5.5.5.5/24, and 6.6.6.6/24 to BGP.

```
[SwitchA-bgp] network 4.4.4.4 24
[SwitchA-bgp] network 5.5.5.5 24
[SwitchA-bgp] network 6.6.6.6 24
```

On Switch B, inject routes 7.7.7.7/24, 8.8.8.8/24, and 9.9.9.9/24 to BGP.

```
[SwitchB-bgp] network 7.7.7.7 24
[SwitchB-bgp] network 8.8.8.8 24
[SwitchB-bgp] network 9.9.9.9 24
```

Display the BGP routing table information of Switch D.

```
[SwitchD-bgp] display bgp routing-table

 Total Number of Routes: 6

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

 *>  4.4.4.0/24         1.1.3.1                               0       300 100i
 *>  5.5.5.0/24         1.1.3.1                               0       300 100i
 *>  6.6.6.0/24         1.1.3.1                               0       300 100i
 *>  7.7.7.0/24         1.1.3.1                               0       300 200i
 *>  8.8.8.0/24         1.1.3.1                               0       300 200i
```

```
 *>  9.9.9.0/24         1.1.3.1                                0       300 200i
```

The display above shows that Switch D has learned routes 4.4.4.0/24, 5.5.5.0/24, and 6.6.6.0/24 from AS 100 and 7.7.7.0/24, 8.8.8.0/24, and 9.9.9.0/24 from AS 200.

3) Configure Switch D to reject routes from AS 200.

Configure AS_PATH list 1 on Switch D.

```
[SwitchD] ip as-path 1 permit .*200.*
```

Configure a route policy named **rt1** on Switch D.

```
[SwitchD] route-policy rt1 deny node 1
[SwitchD] if-match as-path 1
[SwitchD] route-policy rt2 permit node 2
```

On Switch D, specify route policy **rt1** to filter routes received from peer 1.1.3.1.

```
[SwitchD] bgp 400
[SwitchD] peer 1.1.3.1 route-policy rt1 import
```

Display the BGP routing table information of Switch D.

```
[SwitchD] display bgp routing-table


 Total Number of Routes: 3

 BGP Local router ID is 4.4.4.4
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
     Network            NextHop         MED        LocPrf     PrefVal Path/Ogn

 *>  4.4.4.0/24         1.1.3.1                                0       300 100i
 *>  5.5.5.0/24         1.1.3.1                                0       300 100i
 *>  6.6.6.0/24         1.1.3.1                                0       300 100i
```

The display above shows that Switch D has learned only routes 4.4.4.0/24, 5.5.5.0/24, and 6.6.6.0/24 from AS 100.

Troubleshooting Route Policy Configuration

IPv4 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while the routing protocol runs normally.

Analysis

At least one item of the IP prefix list should be configured as permit mode, and at least one node in the Route policy should be configured as permit mode.

Solution

1) Use the **display ip ip-prefix** command to display IP prefix list information.
2) Use the **display route-policy** command to display route policy information.

IPv6 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while the routing protocol runs normally.

Analysis

At least one item of the IPv6 prefix list should be configured as permit mode, and at least one node of the Route policy should be configured as permit mode.

Solution

1) Use the **display ip ipv6-prefix** command to display IP prefix list information.
2) Use the **display route-policy** command to display route policy information.

Table of Contents

1 BFD Configuration

When configuring BFD, go to these sections for information you are interested in:

Note

- The term “router” or router icon in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.
- Switche 4800G is centralized device that support IRF. Multiple 4800G switches can form a distributed chassis switch in a logical sense, in which the master and slave switches are like the master and slave boards of a distributed switch. The centralized 4800G series switches are implemented in a distributed way after they form an IRF stack.

Introduction to BFD

Bidirectional forwarding detection (BFD) provides a single mechanism to quickly detect and monitor the connectivity of links in networks. To improve network performance, devices must quickly detect communication failures to restore communication through backup paths as soon as possible. Normally, devices in a network may employ the following detection methods:

- Detect link failures by sending hardware detection signals, such as SDH (synchronous digital hierarchy) transmission system alarms.
- If no hardware detection signals are provided or failures cannot be detected through hardware detection signals, devices can use the hello mechanism of a routing protocol for failure detection, which has a slower failure detection rate of more than one second. In Gigabit data transmission, such a rate will cause a large quantity of data to be dropped.
- Implement real-time detection for all media types and protocols through a uniform mechanism.

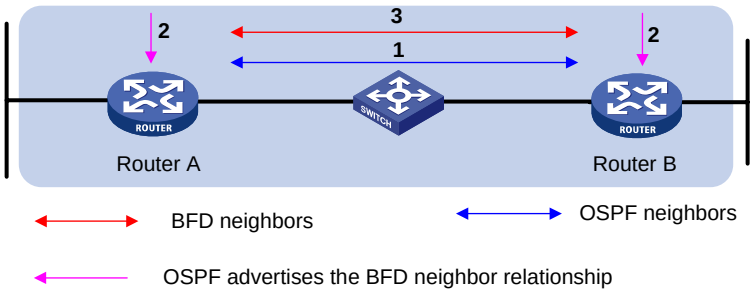
How BFD Works

BFD provides a general-purpose, standard, medium- and protocol-independent fast failure detection mechanism. It can uniformly and quickly detect the failures of the bidirectional forwarding paths between two routers for protocols, such as routing protocols and Multiprotocol Label Switching (MPLS).

BFD provides no neighbor discovery mechanism. Protocols that BFD services notify BFD of routers to which it needs to establish sessions. After a session is established, if no BFD control packet is received from the peer within the negotiated BFD interval, BFD notifies a failure to the protocol, which takes appropriate measures.

Operation of BFD

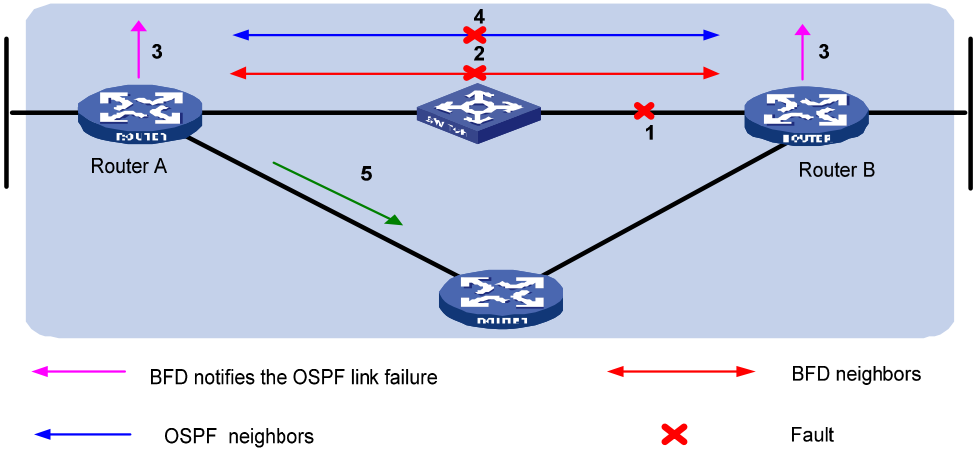
Figure 1-1 BFD session establishment



BFD session establishment (as shown in the above figure):

- A protocol sends Hello messages to discover neighbors and establish neighborships.
- After establishing neighborships, the protocol notifies BFD of the neighbor information, including destination and source addresses.
- BFD uses the information to establish BFD sessions.

Figure 1-2 BFD fault detection



BFD fault detection (as shown in the above figure):

- Upon detection of a link failure, BFD clears the session and notifies the protocol of the failure.
- The protocol terminates the neighborship on the link.
- If a backup link is available, the protocol will use it to forward packets.

Note

No detection time resolution is defined in the BFD draft. At present, most devices supporting BFD provide detection measured in milliseconds.

BFD session modes

- Control packet mode: Both ends of the link exchange BFD control packets to monitor link status.
- Echo mode: One end of the link sends Echo packets to the other end, which then forwards the packets back to the originating end, thereby monitoring link status in both directions.

BFD operating modes

Before a BFD session is established, there are two BFD operating modes: active and passive.

- Active mode: In this mode, BFD actively sends BFD control packets regardless of whether any BFD control packet is received from the peer.
- Passive mode: In this mode, BFD does not send control packets until a BFD control packet is received from the peer.

At least one end must operate in the active mode for a BFD session to be established.

After a BFD session is established, both ends must operate in one of the following two BFD operating modes: asynchronous and command.

- Asynchronous mode: A device operating in the asynchronous mode periodically sends BFD control packets. It tears down the BFD session if it receives no BFD control packet from the peer within the BFD interval.
- Demand mode: In this mode, it is assumed that a system has an independent way of verifying that it has connectivity to the other system. Once a BFD session is established, such a system may ask the other system to stop sending BFD Control packets, except when the system feels the need to verify connectivity explicitly, in which case a short sequence of BFD Control packets is exchanged, and then the far system quiesces. Demand mode may operate independently in each direction, or simultaneously.

Note

- At present, only the asynchronous mode is supported.
- When a BFD session operates in Echo mode, the session is independent of the operating mode.
- When the connectivity to another system needs to be verified explicitly, a system sends several BFD control packets that have the Poll (P) bit set at the negotiated transmit interval. If no response is received within the detection interval, the session is considered down. If the connectivity is found to be up, no more BFD control packets are sent until the next command is issued.

Dynamic BFD parameter changes

After a BFD session is established, both ends can negotiate the related BFD parameters, such as the minimum transmit interval, minimum receive interval, initialization mode, and packet authentication mode. After that, both ends use the negotiated parameters, without affecting the current session state.

Authentication modes

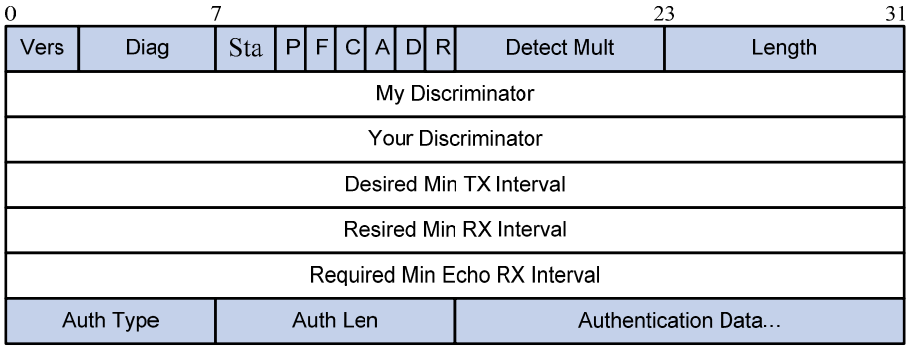
BFD provides the following authentication methods:

- Simple: Plain text authentication
- MD5: MD5 (Message Digest 5) authentication
- SHA1: SHA1 (Secure Hash Algorithm 1) authentication

BFD Packet Format

As illustrated in Figure 1-3, BFD control packets use UDP and port number 3784.

Figure 1-3 BFD control packet format



- Vers: Protocol version. The protocol version is 1.
- Diag: This bit indicates the reason for the last transition of the local protocol from **up** to some other state. Table 1-1 lists the states.

Table 1-1 Diag bit values

Diag	Description
0	No Diagnostic
1	Control Detection Time Expired
2	Echo Function Failed
3	Neighbor Signaled Session Down
4	Forwarding Pane Reset
5	Path Down
6	Concatenated Path Down
7	Administratively Down
8~31	Reserved for future use

- State (Sta): Current BFD session state. Its value can be 0 for AdminDown, 1 for Down, 2 for Init, and 3 for Up.

- Demand (D): If set, Demand mode is active in the transmitting system (the system wishes to operate in Demand mode, knows that the session is up in both directions, and is directing the remote system to cease the periodic transmission of BFD Control packets). If clear, Demand mode is not active in the transmitting system.
- Poll (P): If set, the transmitting system is requesting verification of connectivity, or of a parameter change, and is expecting a packet with the Final (F) bit in reply. If clear, the transmitting system is not requesting verification.
- Final (F): If set, the transmitting system is responding to a received BFD Control packet that had the Poll (P) bit set. If clear, the transmitting system is not responding to a Poll.
- Control Plane Independent(C): If set, the transmitting system's BFD implementation does not share fate with its control plane (in other words, BFD is implemented in the forwarding plane and can continue to function through disruptions in the control plane.) If clear, the transmitting system's BFD implementation shares fate with its control plane.
- Authentication Present (A): If set, the Authentication Section is present and the session is to be authenticated.
- Reserved (R): This byte must be set to zero on transmit, and ignored on receipt.
- Detect Mult: Detection time multiplier.
- Length: Length of the BFD Control packet, in bytes.
- My Discriminator: A unique, nonzero discriminator value generated by the transmitting system, used to demultiplex multiple BFD sessions between the same pair of systems.
- Your Discriminator: It is the discriminator received from the corresponding remote system. This field reflects back the received value of My Discriminator, or is 0 if that value is unknown.
- Desired Min Tx Interval: This is the minimum interval, in microseconds, that the local system would like to use when transmitting BFD Control packets. The value zero is reserved.
- Required Min Rx Interval: This is the minimum interval, in microseconds, between received BFD Control packets that this system is capable of supporting. If this value is zero, the transmitting system does not want the remote system to send any periodic BFD Control packets.
- Required Min Echo Rx Interval: This is the minimum interval, in microseconds, between received BFD Echo packets that this system is capable of supporting. If this value is zero, the transmitting system does not support the receipt of BFD Echo packets.
- Auth Type: The authentication type in use, if the Authentication Present (A) bit is set.
- Auth Len: The length, in bytes, of the authentication section, including the Auth Type and Auth Len fields.

Protocols and Standards

- draft-ietf-bfd-base-05: *Protocol Independent Bidirectional Forwarding Detection*
- draft-ietf-bfd-v4v6-1hop-05: *BFD for IPv4 and IPv6 (Single Hop)*

BFD Configuration Task List

Complete the following tasks to configure BFD:

Task		Remarks
Configuring BFD Basic Functions		Optional
Configuring Protocol-based BFD	Configuring BFD for OSPF	Required
	Configuring BFD for IS-IS	Required
	Configuring BFD for RIP	Required
	Configuring BFD for BGP	Required
	Configuring BFD for VRRP	Required
	Configuring BFD for Static Routes	Required

Configuring BFD Basic Functions

Configuration Prerequisites

Before configuring BFD detection modes, complete the following tasks:

- Configure the network layer addresses of the interfaces so that adjacent nodes are reachable to each other at the network layer;
- Configure the routing protocols that support BFD

Configuration Procedure

Follow these steps to configure BFD basic functions:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify a BFD session initiation mode	**bfd session init-mode** { **active** | **passive** }	Optional **active** by default
Enter interface view	**interface** *interface-type interface-number*	—
Configure the minimum BFD transmit interval	**bfd min-transmit-interval** *value*	Optional 400 seconds by default
Configure the minimum echo receive interval	**bfd min-echo-receive-interval** *value*	Optional 400 seconds by default
Configure the minimum packet receive interval	**bfd min-receive-interval** *value*	Optional 400 seconds by default
Configure the detect time multiplier	**bfd detect-multiplier** *value*	Optional 5 by default

To do…	Use the command…	Remarks
Configure the authentication type	**bfd authentication-mode** { **md5** *key-id key* | **sha1** *key-id key* | **simple** *key-id password* }	Optional By default, the interface operates in the non-authentication mode.

Configuring Protocol-based BFD

Configuring BFD for OSPF

After discovering neighbors by sending hello packets, OSPF notifies BFD of the neighbor addresses, and BFD uses theses addresses to establish sessions. Before a BFD session is established, it is in the **Down** state. In this state, BFD control packets are sent at an interval of not less than one second to reduce BFD control packet traffic. After the BFD session is established, BFD control packets are sent at the negotiated interval, thereby implementing fast fault detection. To configure BFD for OSPF, you need to configure OSPF first.

Follow these steps to enable BFD on an OSPF interface:

To do…	Use the command…	Description
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable BFD on the interface	**ospf bfd enable**	Required Not enabled by default

Note

- One network segment can only belong to one area and you must specify each OSPF interface to belong to the specific area.
- Both ends of a BFD session must be on the same network segment and in the same area.
- For OSPF configuration, refer to *OSPF Configuration* in the *IP Routing Volume*.

Configuring BFD for IS-IS

Before the following configuration, configure IS-IS basic functions on each node.

Follow these steps to enable BFD on an IS-IS interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable IS-IS on the interface	**isis enable** [*process-id*]	Required Disabled by default

To do…	Use the command…	Remarks
Enable BFD on the IS-IS interface	**isis bfd enable**	Required Not enabled by default

Note

For details about IS-IS, refer to *IS-IS Configuration* in the *IP Routing Volume*.

Configuring BFD for RIP

RIP periodically sends route update requests to neighbors. If no route update response for a route is received within the specified interval, RIP considers the route unreachable. This mechanism cannot detect link faults quickly.

After BFD is configured for RIP, when BFD detects a broken link, RIP can quickly age out the unreachable route before the update timer expires.

RIP with BFD support provides two link detection modes:

- Single-hop detection in BFD echo packet mode for a directly connected neighbor. In this mode, a BFD session is established only when the neighbor has route information to send.
- Bidirectional detection in BFD control packet mode for an indirectly connected neighbor. In this mode, a BFD session is established only when both ends have routes to send and BFD is enabled on the receiving interface.

Single-hop detection in BFD echo packet mode

Follow these steps to configure BFD for RIP (single-hop detection in BFD echo packet mode):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the source IP address of BFD echo packets	**bfd echo-source-ip** *ip-address*	Required By default, no source IP address is configured for BFD echo packets.
Enter interface view	**interface** *interface-type interface-number*	—
Enable BFD on the RIP interface	**rip bfd enable**	Required Disabled by default.

Bidirectional detection in BFD control packet mode

Follow these steps to configure BFD for RIP (bidirectional detection in BFD control packet mode):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RIP process and enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	Required By default, RIP is disabled.
Specify a RIP neighbor	**peer** *ip-address*	Required By default, RIP does not unicast updates to any peer.
Enter interface view	**interface** *interface-type interface-number*	—
Enable BFD on the RIP interface	**rip bfd enable**	Required Disabled by default

Note

- Unidirectional detection in BFD echo packet mode only works for RIP neighbors that are directly connected, namely, one hop away from each other.
- Using the **undo peer** command does not remove the neighbor relationship at once and therefore cannot bring down the BFD session at once.
- For RIP configuration, refer to *RIP Configuration* in the *IP Routing Volume*.

Configuring BFD for BGP

The default BGP keepalive interval is 60 seconds and can be configured as 1 second at least. Thus, the holdtime is 180 seconds by default and three seconds at least. This makes the detection of neighbor relationships rather slow, and a large quantity of packets will be dropped when being transmitted or received through a high-speed interface. BFD quickly detects neighbor relationships and reduces network convergence time. Before configuring BFD for BGP, you need to enable BGP.

Follow these steps to enable BFD for a BGP peer:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable BFD for the specified BGP peer	**peer** *ip-address* **bfd**	Required Not enabled by default

Note

- At present, you can configure BFD for IPv4 BGP neighbors only.
- If GR capability is enabled for BGP, use BFD with caution.
- For BGP configuration, refer to *BGP Configuration* in the *IP Routing Volume*.

Configuring BFD for VRRP

To configure BFD for VRRP, you need to configure a BFD track entry and then bind the track entry to a VRRP group.

Track starts and stops the BFD session. Upon detecting a neighbor failure, BFD notifies Track of the failure. Then, Track notifies it to VRRP for quick VRRP master/backup switchover.

Before associating a VRRP group with a track entry, you need to create the VRRP group on the interface and assign a virtual IP address to it.

Follow these steps to configure BFD for VRRP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the source address of echo packets	**bfd echo-source-ip** *ip-address*	Required Not configured by default
Configure a BFD track entry	**track** *track-entry-number* **bfd echo interface** *interface-type interface-number* **remote ip** *remote-ip* **local ip** *local-ip*	Required Not configured by default
Enter interface view	**interface** *interface-type interface-number*	—
Bind a VRRP group with the Track entry	**vrrp vrid** *virtual-router-id* **track** *track-entry-number* { **switchover** | [**reduced** *priority-reduced*] }	Required By default, no monitored interface is specified.

Note

- The *local-ip* of the track entry must be the IP address of the specified outgoing interface and must be on the network segment as the *remote-ip*.
- For VRRP configuration, refer to *VRRP Configuration* in the *System Volume*.
- For Track configuration, refer to *Track Configuration* in the System Volume.

Configuring BFD for Static Routes

A dynamic routing protocol notifies BFD of its neighbor information. BFD uses such information to establish sessions with neighbors by sending BFD control packets. Static routing, which has no neighbor discovery mechanism, implements BFD as follows:

BFD control packet mode

To use BFD control packets for bidirectional detection between two devices, you need to enable BFD control packet mode for each device's static route destined to the peer.

Follow these steps to configure BFD control packet mode for static routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable BFD control packet mode for static routes	**ip route-static** *dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Use either command
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* **bfd control-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

BFD echo packet mode

With BFD echo packet mode enabled for a static route, the local device sends BFD echo packets to the peer, which loops it back to test the link in between.

Follow these steps to configure BFD echo packet mode for static routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the source address of echo packets	**bfd echo-source-ip** *ip-address*	Required Not configured by default
Enable BFD echo packet mode for static routes	**ip route-static** *dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Use either command
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* **|** *mask-length* } *interface-type interface-number next-hop-address* **bfd echo-packet** [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	

Caution

- If route flaps occur, enabling BFD may worsen the route flaps. Therefore, enable BFD with care in such cases.
- The source address of echo packets must be configured if the BFD session operates in the echo mode.
- If you configure BFD for a static route, you need to specify the outbound interface and next hop IP address for the route.
- BFD cannot be used for a static route with the outbound interface having the spoofing attribute.
- BFD can be used for static routes with direct nexthops rather than non-direct nexthops.
- In the draft, the BFD echo function is revised to specify that a BFD session is established at only one end when the echo mode is used.
- For static route configuration, refer to *Static Routing Configuration* in the *IP Routing Volume*.

Enabling Trap

When the trap function is enabled on the BFD module, the module will generate trap messages at the notifications level to report the important events of the module. The generated trap messages are sent to the device's information center, which determines the output rules for the trap messages, namely, whether to output the trap messages and the output destinations). For the information center configuration, refer to *Information Center Configuration* in the *System Volume*.

Follow these steps to enable BFD trap:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable BFD trap	**snmp-agent trap enable bfd**	Optional Enabled by default

Note

For the description of the **snmp-agent trap enable isdn** command, refer to the **snmp-agent trap enable** command in *SNMP Commands* in the *System Volume*.

Displaying and Maintaining BFD

To do…	Use the command…	Remarks
Display information about BFD-enabled interfaces	**display bfd interface** [**verbose**]	Available in any view
Display information about enabled BFD debugging	**display bfd debugging-switches**	Available in any view
Display PAF configuration information	**display bfd paf**	Available in any view

To do…		Use the command…	Remarks
Display BFD session information	On a centralized device	**display bfd session** [**verbose**]	Available in any view
	On a distributed device	**display bfd session** [**verbose**] [**slot** *slot-number* [**all** | **verbose**]]	Available in any view
Clear BFD session statistics	On a centralized device	**reset bfd session statistics**	Available in user view
	On a distributed device	**reset bfd session statistics** [**slot** *slot-number*]	Available in user view

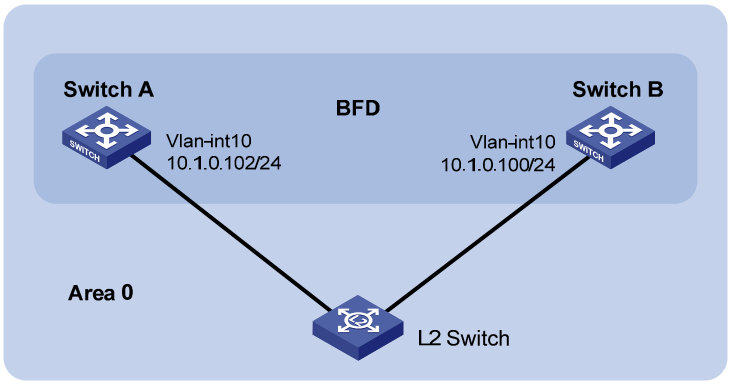
BFD Configuration Examples

Configuring BFD for OSPF

Network requirements

- Switch A and Switch B are interconnected through a Layer 2 switch. BFD is enabled on the switch interfaces. OSPF is enabled on the switches that are reachable to each other at the network layer.
- When the link between Switch B and the Layer 2 switch fails, BFD can quickly detect the failure and notify OSPF of the failure.

Figure 1-4 Network diagram for BFD configuration on an OSPF link



Configuration procedure

1) Configure VLAN interfaces.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] ip address 10.1.0.102 24
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] ip address 10.1.0.100 24
[SwitchB-Vlan-interface10] quit
```

2) Configure OSPF basic functions.

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.0.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] ospf bfd enable
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.0.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
[SwitchB] interface vlan-interface 10
[SwitchB-Vlan-interface10] ospf bfd enable
[SwitchB-Vlan-interface10] quit
```

3) Configure BFD parameters.

Configure Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] bfd min-transmit-interval 300
[SwitchA-Vlan-interface10] bfd min-receive-interval 300
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] bfd authentication-mode simple 1 zhang
[SwitchA-Vlan-interface10] quit
[SwitchA] quit
```

Configure Switch B.

```
[SwitchB] bfd session init-mode active
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] bfd min-transmit-interval 300
[SwitchB-Vlan-interface10] bfd min-receive-interval 300
[SwitchB-Vlan-interface10] bfd detect-multiplier 6
[SwitchB-Vlan-interface10] bfd authentication-mode simple 1 zhang
```

4) Verify the configuration.

Display BFD information of Switch A.

```
<SwitchA> display bfd session
 Total Session Num: 1            Init Mode: Active
 Session Working Under Ctrl Mode:
 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 3/1           10.1.0.102      10.1.0.100      Up    1700ms   vlan10
```

Display OSPF neighbor information of Switch A.

```
<SwitchA> display ospf peer

                  OSPF Process 1 with Router ID 192.168.1.40
                       Neighbor Brief Information

 Area: 0.0.0.1
 Router ID       Address         Pri Dead-Time Interface       State
 10.1.0.102      10.1.0.100      1   31        vlan10          Full/DR
```

Enable BFD debugging on Switch A.

```
<SwitchA> debugging bfd scm
<SwitchA> debugging bfd event
<SwitchA> debugging ospf event
<SwitchA> terminal debugging
```

When the link between Switch B and the Layer 2 switch fails, you can see that Switch A can quickly detect the changes on Switch B.

```
%Nov 12 18:34:48:823 2005 SwitchA BFD/5/LOG: Sess[10.1.0.102/10.1.0.100, vlan10], Sta :
UP->DOWN, Diag: 1
%Nov 12 18:34:48:824 2005 SwitchA RM/4/RMLOG:OSPF-NBRCHANGE: Process 1, Neighbour 10.1.0.102
(vlan10) from Full to Down
*0.50673825 SwitchA BFD/8/SCM:Sess[10.1.0.102/10.1.0.100, vlan10],Oper: Reset
*0.50673825 SwitchA BFD/8/EVENT:Send sess-down Msg, [Src:10.1.0.102, Dst:10.1.0.100, vlan10]
Protocol: OSPF
*0.50673826 SwitchA RM/7/RMDEBUG:OSPF-BFD: Message Type rcv BFD down, Connect Type
direct-connect, Src IP Address 10.1.0.102, Src IFIndex 5, Dst IP Address 10.1.0.100
*0.50673827 SwitchA RM/7/RMDEBUG:OSPF-BFD: Message Type delete session, Connect Type
direct-connect, Src IP Address 10.1.0.102, Src IFIndex 5, Dst IP Address 10.1.0.100
OSPF 1: Nbr 10.1.0.100 Rcv KillNbr State Full -> Down.
*0.50673829 SwitchA BFD/8/EVENT:Receive Delete-sess, [Src:10.1.0.102, Dst:10.1.0.100,
vlan10], Direct, Proto:OSPF
*0.50673830 SwitchA BFD/8/SCM:Sess[10.1.0.102/10.1.0.100, vlan10], Oper: Del
application(OSPF)
*0.50673831 SwitchA BFD/8/SCM:No application in session, delete
session[10.1.0.102/10.1.0.100, vlan10]
*0.50673831 SwitchA BFD/8/SCM:Sess[10.1.0.102/10.1.0.100, vlan10], Oper: Delete
*0.50673832 SwitchA BFD/8/SCM:Delete send-packet timer
*0.50673833 SwitchA BFD/8/SCM:Delete session entry
*0.50673833 SwitchA BFD/8/SCM:Delete session from IP hash table
*0.50673834 SwitchA BFD/8/SCM:Delete session from bfd interface
*0.50673834 SwitchA BFD/8/SCM:No session under bfd-int[vlan10] with default configuration,
delete bfd-if
*0.50673835 SwitchA BFD/8/SCM:Bfd-if[vlan10], Oper: Delete
*0.50673840 SwitchA BFD/8/SCM:No bfd session exists, stop receiving any bfd packets
```

Display BFD information of Switch A.

Because Switch A has removed its neighbor relationship with Switch B, no information is output.

```
<SwitchA> display bfd session
```

Display OSPF neighbor information of Switch A.

Because Switch A has removed its neighbor relationship with Switch B, no information is output.

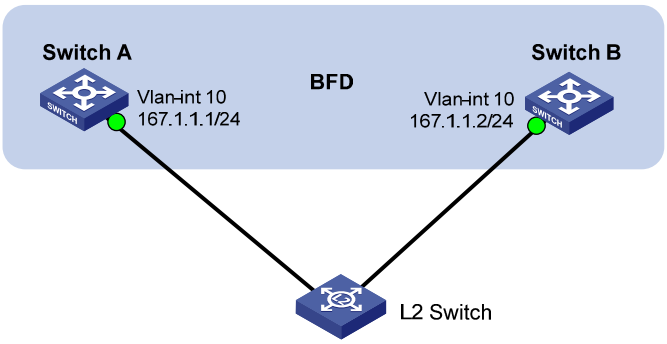
```
<SwitchA> display ospf peer
                  OSPF Process 1 with Router ID 192.168.1.40
                       Neighbor Brief Information
```

Configuring BFD for IS-IS

Network requirements

- Switch A and Switch B are interconnected through a Layer-2 switch. BFD is enabled on the switch interfaces. IS-IS is enabled on the switches that are reachable to each other at the network layer.
- When the link between Switch B and the Layer-2 switch fails, BFD can quickly detect the failure and notify IS-IS of the failure.

Figure 1-5 Network diagram for BFD configuration on an IS-IS link



Configuration procedure

1) Configure VLAN interfaces.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] ip address 167.1.1.1 24
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] ip address 167.1.1.2 24
[SwitchB-Vlan-interface10] quit
```

2) Configure IS-IS basic functions.

Configure Switch A.

```
[SwitchA] isis
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] isis enable
[SwitchA-Vlan-interface10] isis bfd enable
```

```
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
[SwitchB] isis
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] isis enable
[SwitchB-Vlan-interface10] isis bfd enable
[SwitchB-Vlan-interface10] quit
```

3) Configure BFD parameters.

Configure Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] bfd min-receive-interval 300
[SwitchA-Vlan-interface10] bfd min-transmit-interval 300
[SwitchA-Vlan-interface10] bfd authentication-mode simple 1 zhang
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] quit
[SwitchA] quit
```

Configure Switch B.

```
[SwitchB] bfd session init-mode active
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] bfd min-receive-interval 500
[SwitchB-Vlan-interface10] bfd min-transmit-interval 500
[SwitchB-Vlan-interface10] bfd authentication-mode simple 1 zhang
[SwitchB-Vlan-interface10] bfd detect-multiplier 8
```

4) Verify the configuration.

Display BFD information of Switch A.

```
<SwitchA> display bfd session
 Total Session Num: 1            Init Mode: Active
 Session Working Under Ctrl Mode:
 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 5/3           167.1.1.1       167.1.1.2       Up    1900ms   vlan10
```

Display IS-IS neighbor information of Switch A.

```
<SwitchA> display isis peer 1
  System Id: 0000.0000.0002
  Interface: vlan10                 Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 24s       Type: L1(L1L2)     PRI: 64
  System Id: 0000.0000.0002
  Interface: vlan10                 Circuit Id: 0000.0000.0001.01
  State: Up     HoldTime: 29s       Type: L2(L1L2)     PRI: 64
```

Enable BFD debugging on Switch A.

```
<SwitchA> debugging bfd scm
<SwitchA> debugging bfd event
<SwitchA> debugging isis event bfd
```

```
<SwitchA> terminal debugging
```

When the link between Switch B and the Layer 2 switch fails, you can see that Switch A can quickly detect the changes on Switch B.

```
#Aug  8 14:54:05:362 2008 SwitchA IFNET/4/INTERFACE UPDOWN:
 Trap 1.3.6.1.6.3.1.1.5.3<linkDown>: Interface 983041 is Down, ifAdminStatus is
1, ifOperStatus is 2
#Aug  8 14:54:05:363 2008 SwitchA ISIS/4/ADJ_CHANGE:TrapID(1.3.6.1.2.1.138.0.17<
isisAdjacencyChange>), ISIS Level-2 Adjencency IN Circuit-983041 State Change.
#Aug  8 14:54:05:364 2008 SwitchA ISIS/4/ADJ_CHANGE:TrapID(1.3.6.1.2.1.138.0.17<
isisAdjacencyChange>), ISIS Level-1 Adjencency IN Circuit-983041 State Change.
%Aug  8 14:54:05:365 2008 SwitchA IFNET/4/LINK UPDOWN:
 vlan10: link status is DOWN
%Aug  8 14:54:05:366 2008 SwitchA IFNET/4/UPDOWN:
 Line protocol on the interface Ethernet0/1 is DOWN
%Aug  8 14:54:05:367 2008 SwitchA ISIS/4/ADJLOG:ISIS-1-ADJCHANGE: Adjacency To 0
000.0000.0002 (vlan10) DOWN, Level-2 Circuit Down.
%Aug  8 14:54:05:367 2008 SwitchA ISIS/4/ADJLOG:ISIS-1-ADJCHANGE: Adjacency To 0
000.0000.0002 (vlan10) DOWN, Level-2 Adjacency clear.
%Aug  8 14:54:05:368 2008 SwitchA ISIS/4/ADJLOG:ISIS-1-ADJCHANGE: Adjacency To 0
000.0000.0002 (vlan10) DOWN, Level-1 Circuit Down.
%Aug  8 14:54:05:369 2008 SwitchA ISIS/4/ADJLOG:ISIS-1-ADJCHANGE: Adjacency To 0
000.0000.0002 (vlan10) DOWN, Level-1 Adjacency clear.
*Aug  8 14:54:05:370 2008 SwitchA ISIS/6/ISIS:
 ISIS-1-BFD: Success to send msg. Msg type 1 delete session. IfPhyIndex: 5 ,DstI
PAddr: 192.168.0.100 , SrcIPAddr:192.168.0.102. NeighborType:Level-2.
*Aug  8 14:54:05:370 2008 SwitchA ISIS/6/ISIS:
 ISIS-1-BFD: Success to send msg. Msg type 1 delete session. IfPhyIndex: 5 ,DstI
PAddr: 192.168.0.100 , SrcIPAddr:192.168.0.102. NeighborType:Level-1.
```

Display BFD information of Switch A.

Because Switch A has removed its neighbor relationship with Switch B, no information is output.

```
<SwitchA> display bfd session
```

Display IS-IS neighbor information of Switch A.

Because Switch A has removed its neighbor relationship with Switch B, no information is output.

```
<SwitchA> display isis peer 1
```

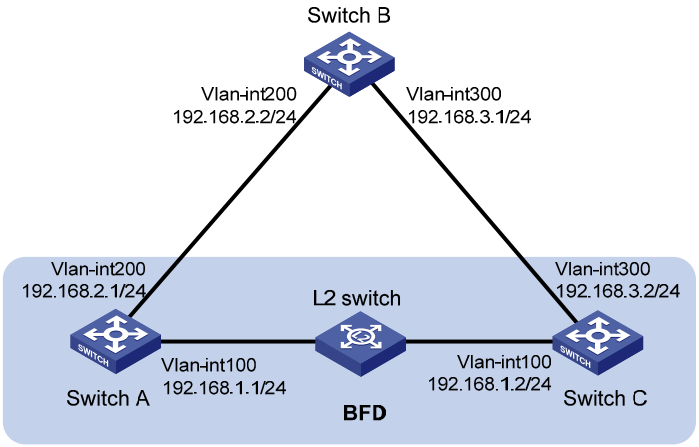
Configuring BFD for RIP (Single-Hop Detection in BFD Echo Packet Mode)

Network requirements

- Switch A and Switch C are interconnected through a Layer 2 switch. VLAN-interface 100 of the two switches runs RIP process 1, BFD is enabled on VLAN-interface 100 of Switch A.
- Switch A is connected to Switch C through Switch B. VLAN-interface 200 on Switch A runs RIP process 2; VLAN-interface 300 on Switch C, and VLAN-interface 200 and VLAN-interface 300 on Switch B run RIP process 1.
- Configure a static route and enable static route redistribution into RIP on Switch C. Switch A learns the static route sent by Switch C, the outbound interface of the route is the interface connected to the Layer 2 switch.

- When the link between Switch C and the Layer 2 switch fails, BFD can quickly detect the link failure and notify it to RIP, and the BFD session goes down. In response, RIP deletes the neighbor relationship with Switch C and the route information received from Switch C. Then, Switch A learns the static route sent by Switch C with the outbound interface being the interface connected to Switch B.

Figure 1-6 Network diagram for configuring BFD for RIP (single-hop detection in BFD echo packet mode)



Configuration procedure

1) Configure VLAN interfaces.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.1.1 24
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ip address 192.168.2.1 24
[SwitchA-Vlan-interface200] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ip address 192.168.2.2 24
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 300
[SwitchB-Vlan-interface300] ip address 192.168.3.1 24
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ip address 192.168.1.2 24
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] ip address 192.168.3.2 24
[SwitchC-Vlan-interface300] quit
```

2) Configure RIP basic functions.

Configure Switch A.

```
[SwitchA] rip 1
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] rip bfd enable
[SwitchA-Vlan-interface100] quit
[SwitchA] rip 2
[SwitchA-rip-2] network 192.168.2.0
```

Configure Switch B.

```
[SwitchB] rip 1
[SwitchB-rip-1] network 192.168.2.0
[SwitchB-rip-1] network 192.168.3.0
[SwitchB-rip-1] quit
```

Configure Switch C.

```
[SwitchC] rip 1
[SwitchC-rip-1] network 192.168.1.0
[SwitchC-rip-1] network 192.168.3.0
[SwitchC-rip-1] import-route static
[SwitchC-rip-1] quit
```

3) Configure BFD parameters.

Configure Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] bfd echo-source-ip 11.11.11.11
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] bfd min-transmit-interval 500
[SwitchA-Vlan-interface100] bfd min-receive-interval 500
[SwitchA-Vlan-interface100] bfd detect-multiplier 7
[SwitchA-Vlan-interface100] quit
[SwitchA] quit
```

4) Configure a static route on Switch C.

```
[SwitchC] ip route-static 100.1.1.1 24 null 0
```

5) Verify the configuration.

Display the BFD session information of Switch A.

```
<SwitchA> display bfd session
 Total Session Num: 1            Init Mode: Active
 Session Working Under Echo Mode:
 LD          SourceAddr      DestAddr        State   Holdtime    Interface
 5           192.168.1.1     192.168.1.2     Up      2000ms      Vlan100
```

Display the RIP route 100.1.1.0/24 learned on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 2
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 1
```

```
   Preference: 100                  Cost: 1
      NextHop: 192.168.1.2      Interface: vlan-interface 100
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.1.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h00m47s
          Tag: 0
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                  Cost: 2
      NextHop: 192.168.2.2      Interface: vlan-interface 200
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.2.2
    Tunnel ID: 0x0                  Label: NULL
        State: Inactive Adv           Age: 00h12m50s
        Tag: 0
```

Enable RIP event debugging on Switch A.

```
<SwitchA> debugging rip 1 event
<SwitchA> terminal debugging
```

When the link between Switch C and the switch fails, you can see that Switch A can quickly detect the change.

```
%Jan 19 10:41:51:203 2008 SwitchA BFD/4/LOG:Sess[192.168.1.1/192.168.1.2, Vlan-interface
100,Ctrl], Sta: UP->DOWN, Diag: 1
*Jan 19 10:33:12:813 2008 SwitchA RM/6/RMDEBUG: RIP-BFD: Message Type Disable, Connect Type
Direct-connect, Pkt Type Echo, Src IP Address 192.168.1.1, Src IFIndex4, Nbr IP Address
192.168.1.2.
```

Display the BFD information of Switch A.

You can see that Switch A has deleted the neighbor relationship with Switch C and thus no output information is displayed.

```
<SwitchA> display bfd session
```

Display the route information of RIP process 1 on Switch A.

The RIP route learned from Switch C is no longer existent.

```
<SwitchA> display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
```

Display the RIP route 100.1.1.0/24 learned on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 1
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                  Cost: 2
      NextHop: 192.168.2.2      Interface: vlan-interface 200
    BkNextHop: 0.0.0.0         BkInterface:
```

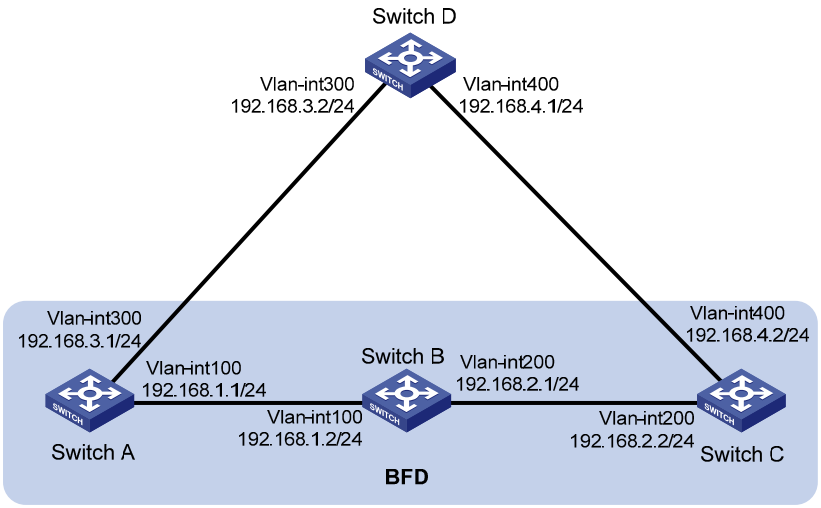
```
RelyNextHop: 0.0.0.0          Neighbor : 192.168.2.2
  Tunnel ID: 0x0                  Label: NULL
      State: Active Adv             Age: 00h18m40s
        Tag: 0
```

Configuring BFD for RIP (Bidirectional Detection in BFD Control Packet Mode)

Network requirements

- Switch A is connected to Switch C through Switch B. VLAN-interface 100 on Switch A, VLAN-interface 200 on Switch C, and VLAN-interface 200 and VLAN-interface 100 on Switch B run RIP process 1.
- Configure a static route to Switch C on Switch A, and configure a static route to Switch A on Switch C. Enable BFD on VLAN-interface 100 of Switch A and VLAN-interface 200 of Switch C.
- Switch A is connected to Switch C through Switch D. VLAN-interface 300 on Switch A runs RIP process 2; VLAN-interface 400 on Switch C, and VLAN-interface 300 and VLAN-interface 400 on Switch D run RIP process 1.
- Enable static route redistribution into RIP on Switch A and Switch C so that Switch A and Switch C have routes to send to each other. Switch A learns the static route sent by Switch C, the outbound interface is the interface connected to Switch B.
- When the link between Switch B and Switch C fails, BFD can quickly detect the link failure and notify it to RIP, and the BFD session goes down. In response, RIP deletes the neighbor relationship with Switch C and the route information received from Switch C. Then, Switch A learns the static route sent by Switch C, the outbound interface of the route is the interface connected to Switch D.

Figure 1-7 Network diagram for configuring BFD for RIP (bidirectional detection in BFD control packet mode)



Configuration procedure

1) Configure VLAN interfaces.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.1.1 24
[SwitchA-Vlan-interface100] quit
```

```
[SwitchA] interface vlan-interface 300
[SwitchA-Vlan-interface300] ip address 192.168.3.1 24
[SwitchA-Vlan-interface300] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.168.1.2 24
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ip address 192.168.2.1 24
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] interface vlan 200
[SwitchC-Vlan-interface200] ip address 192.168.2.2 24
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ip address 192.168.4.2 24
[SwitchC-Vlan-interface400] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] ip address 192.168.3.2 24
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] ip address 192.168.4.1 24
[SwitchD-Vlan-interface400] quit
```

2) Configure RIP basic functions and enable static route redistribution into RIP so that Switch A and Switch C have routes to send to each other.

Configure Switch A.

```
[SwitchA] rip 1
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] peer 192.168.2.2
[SwitchA-rip-1] undo validate-source-address
[SwitchA-rip-1] import-route static
[SwitchA-rip-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] rip bfd enable
[SwitchA-Vlan-interface100] quit
[SwitchA] rip 2
[SwitchA-rip-2] network 192.168.3.0
[SwitchA-rip-2] quit
```

Configure Switch C.

```
[SwitchC] rip 1
[SwitchC-rip-1] network 192.168.2.0
[SwitchC-rip-1] network 192.168.4.0
[SwitchC-rip-1] peer 192.168.1.1
```

```
[SwitchC-rip-1] undo validate-source-address
[SwitchC-rip-1] import-route static
[SwitchC-rip-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] rip bfd enable
[SwitchC-Vlan-interface200] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] rip 1
[SwitchD-rip-1] network 192.168.3.0
[SwitchD-rip-1] network 192.168.4.0
```

3) Configure BFD parameters.

Configure Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] bfd min-transmit-interval 500
[SwitchA-Vlan-interface100] bfd min-receive-interval 500
[SwitchA-Vlan-interface100] bfd detect-multiplier 7
[SwitchA-Vlan-interface100] quit
```

Configure Switch C.

```
[SwitchC] bfd session init-mode active
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] bfd min-transmit-interval 500
[SwitchC-Vlan-interface200] bfd min-receive-interval 500
[SwitchC-Vlan-interface200] bfd detect-multiplier 7
[SwitchC-Vlan-interface200] quit
```

4) Configure static routes.

Configure a static route to Switch C on Switch A.

```
[SwitchA] ip route-static 192.168.2.0 24 vlan-interface 100 192.168.1.2
[SwitchA] quit
```

Configure a static route to Switch A on Switch C.

```
[SwitchC] ip route-static 192.168.1.0 24 vlan-interface 200 192.168.2.1
[SwitchC] ip route-static 100.1.1.0 24 null 0
```

5) Verify the configuration.

Display the BFD session information of Switch A.

```
<SwitchA> display bfd session
Total Session Num: 1            Init Mode: Active
 Session Working Under Ctrl Mode:
 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 6/3           192.168.1.1     192.168.2.2     Up    1700ms   vlan100
```

Display the RIP route 100.1.1.0/24 learned on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 2
```

```
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 1
   Preference: 100                   Cost: 1
      NextHop: 192.168.1.2      Interface: vlan-interface 100
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.1.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h00m47s
          Tag: 0
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
      NextHop: 192.168.3.2      Interface: vlan-interface 300
    BkNextHop: 0.0.0.0         BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.3.2
    Tunnel ID: 0x0                  Label: NULL
        State: Inactive Adv           Age: 00h12m50s
        Tag: 0
```

Enable RIP event debugging on Switch A.

```
<SwitchA> debugging rip 1 event
<SwitchA> terminal debugging
```

When the link between Switch B and Switch C fails, you can see that Switch A quickly detects the link state change.

```
%Jan 19 10:41:51:203 2008 SwitchA BFD/4/LOG:Sess[192.168.1.1/192.168.2.2, Vlan-interface
100, Ctrl], Sta: UP->DOWN, Diag: 1
*Jan 19 10:41:51:203 2008 SwitchA RM/6/RMDEBUG: RIP-BFD: Message Type Disable, Connect Type
Indirect-connect, Pkt Type Control, Src IP Address 192.168.1.1, Src IFIndex 4, Nbr IP Address
192.168.2.2.
```

Display the BFD information of Switch A.

You can see that Switch A has deleted the neighbor relationship with Switch C and thus no output information is displayed.

```
<SwitchA> display bfd session
```

Display the route information of RIP process 1 on Switch A.

The RIP route learned from Switch C is no longer existent.

```
<SwitchA> display rip 1 route
 Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
 ------------------------------------------------------------------------------
```

Display the RIP route 100.1.1.0/24 learned on Switch A.

```
<SwitchA> display ip routing-table 100.1.1.0 24 verbose
Routing Table : Public
Summary Count : 1
  Destination: 100.1.1.0/24
     Protocol: RIP             Process ID: 2
   Preference: 100                   Cost: 2
```

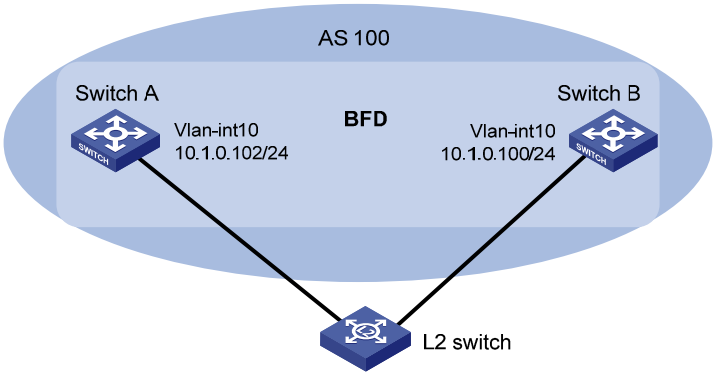
```
      NextHop: 192.168.3.2      Interface: vlan-interface 300
    BkNextHop: 0.0.0.0        BkInterface:
  RelyNextHop: 0.0.0.0          Neighbor : 192.168.3.2
    Tunnel ID: 0x0                  Label: NULL
        State: Active Adv             Age: 00h18m40s
          Tag: 0
```

Configuring BFD for BGP

Network requirements

- Switch A and Switch B are interconnected through a Layer 2 switch. BFD is enabled on the connected interfaces. BGP is enabled on the switches that are reachable to each other at the network layer.
- When the link between Switch A and Switch B fails, BFD can quickly detect the failure and notify BGP of the failure.

Figure 1-8 Network diagram for BFD configuration on a BGP link



Configuration procedure

1) Configure VLAN interfaces.

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10]ip address 10.1.0.102 24
[SwitchA-Vlan-interface10] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] ip address 10.1.0.100 24
[SwitchB-Vlan-interface10] quit
```

2) Configure BGP basic functions.

Configure Switch A.

```
[SwitchA] bgp 100
[SwitchA-bgp] peer 10.1.0.102 as-number 100
[SwitchA-bgp] peer 10.1.0.102 bfd
[SwitchA-bgp] quit
```

Configure Switch B.

```
[SwitchB] bgp 100
[SwitchB-bgp] peer 10.1.0.100 as-number 100
[SwitchB-bgp] peer 10.1.0.100 bfd
[SwitchB-bgp] quit
```

3) Configure BFD parameters.

Configure Switch A.

```
[SwitchA] bfd session init-mode active
[SwitchA-vlan10] interface Vlan-interface 10
[SwitchA-Vlan-interface10] bfd min-transmit-interval 300
[SwitchA-Vlan-interface10] bfd min-receive-interval 300
[SwitchA-Vlan-interface10] bfd detect-multiplier 7
[SwitchA-Vlan-interface10] bfd authentication-mode simple 1 zhang
[SwitchA-Vlan-interface10] quit
[SwitchA] quit
```

Configure Switch B.

```
[SwitchB] bfd session init-mode active
[SwitchB-vlan10] interface Vlan-interface 10
[SwitchB-Vlan-interface10] bfd min-transmit-interval 300
[SwitchB-Vlan-interface10] bfd min-receive-interval 300
[SwitchB-Vlan-interface10] bfd detect-multiplier 6
[SwitchB-Vlan-interface10] bfd authentication-mode simple 1 zhang
```

4) Verify the configuration.

Enable BFD debugging on Switch A.

```
<SwitchA> debugging bfd scm
<SwitchA> debugging bfd event
<SwitchA> debugging bgp bfd
```

Display the detailed BFD neighbor information of Switch A.

```
<SwitchA> display bfd session verbose

Total Session Num: 1            Init Mode: Active

 Session Working Under Ctrl Mode:

     Local Discr: 4                  Remote Discr: 4
       Source IP: 10.1.0.102        Destination IP: 10.1.0.100
   Session State: Up                      Interface: Vlan10
 Min Trans Inter: 300ms            Act Trans Inter: 300ms
  Min Recv Inter: 300ms           Act Detect Inter: 2000ms
    Recv Pkt Num: 52                  Send Pkt Num: 50
       Hold Time: 1600ms              Connect Type: Direct
  Establish Time: 02:42:53          Last Down Time: 02:42:53
    Last Up Time: 02:42:53               Auth mode: None
        Protocol: BGP
       Diag Info: No Diagnostic
```

When the link between Switch A and Switch B fails, display the detailed BGP neighbor information of Switch A. Switch A has removed its neighbor relationship with Switch B.

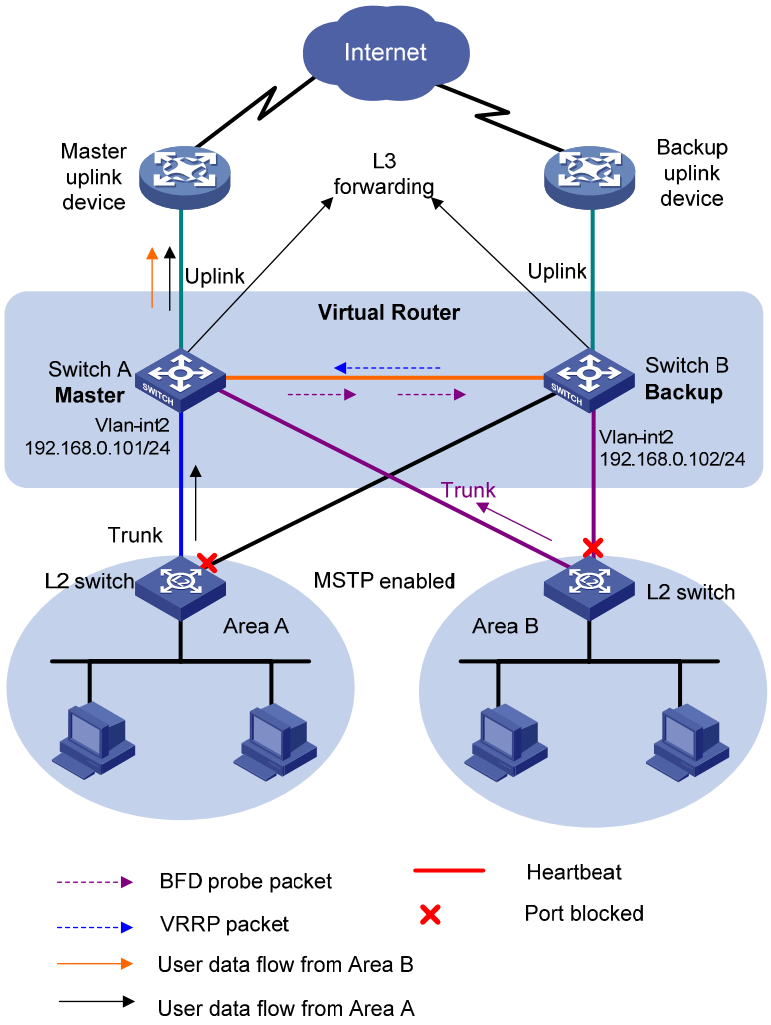
```
<SwitchA> display bgp peer 10.1.0.100 verbose
          Peer: 10.1.0.100  Local: 1.1.1.1
          Type: IBGP link
          BGP version 4, remote router ID 2.2.2.2
          BGP current state: Idle
          BGP current event: Stop
          BGP last state: Established
 Received: Total 1 messages, Update messages 0
 Sent: Total 0 messages, Update messages 0
 Maximum allowed prefix number: 4294967295
 Threshold: 75%
 Minimum time between advertisement runs is 15 seconds
 Optional capabilities:
 Route refresh capability has been enabled
 Peer Preferred Value: 0
 BFD: Enabled

 Routing policy configured:
 No routing policy is configured
```

Configuring BFD for the VRRP Backup to Monitor the Master

Network requirements

If BFD is not configured, when the master in a VRRP group fails, the backup cannot become the master until the configured timeout timer expires. The timeout is generally three to four seconds and therefore the switchover is slow. To solve this problem, VRRP uses BFD to probe the state of the master. Once the master fails, the backup can become the new master within 100 milliseconds.

Figure 1-9 Network diagram for monitoring the master on the backup



Configuration procedure

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 2
[SwitchA-vlan-interface2] ip address 192.168.0.101 24
[SwitchA-vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
[SwitchA-vlan-interface2] vrrp vrid 1 priority 110
[SwitchA-vlan-interface2] return
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bfd session init-mode active
[SwitchB] bfd echo-source-ip 10.10.10.10
[SwitchB] interface vlan-interface 2
[SwitchB-vlan-interface2] ip address 192.168.0.102 24
[SwitchB-vlan-interface2] bfd min-echo-receive-interval 10
[SwitchB-vlan-interface2] bfd detect-multiplier 3
[SwitchB-vlan-interface2] quit
[SwitchB] track 1 bfd echo interface vlan-interface 2 remote ip 192.168.0.101 local ip
192.168.0.102
```

```
[SwitchB] interface vlan-interface 2
[SwitchB–vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
[SwitchB–vlan-interface2] vrrp vrid 1 track 1 switchover
[SwitchB–vlan-interface2] return
```

Use the **display vrrp verbose** command to display the configuration.

Display the detailed information of VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 110                  Run Pri        : 110
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.101
```

Display the detailed information of VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Switchover
 Virtual IP        : 192.168.0.10
 Master IP         : 192.168.0.101
```

The display above shows that, in backup group 1, Switch A is the master router and Switch B the backup router.

When Switch A goes down, use the **display vrrp** command to display the detailed information of VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
```

```
Track Object      : 1                   Switchover
Virtual IP        : 192.168.0.10
Virtual MAC       : 0000-5e00-0101
Master IP         : 192.168.0.102
```

Display the track entry information of Switch B.

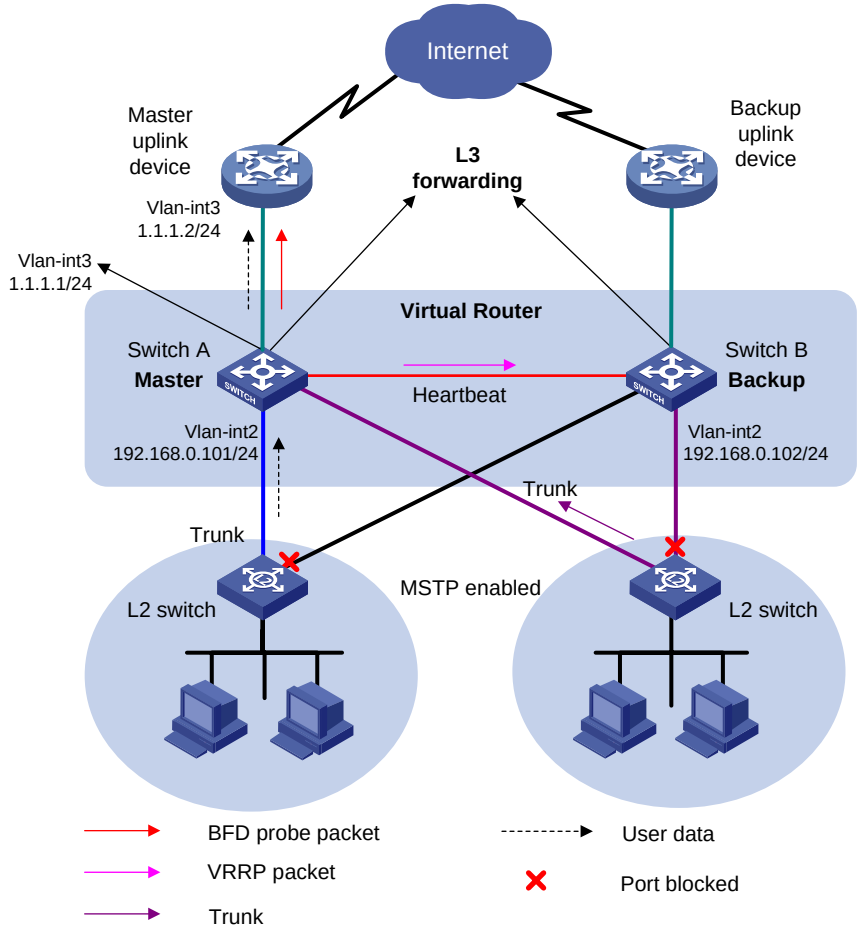
```
<SwitchB> display track 1
Track ID: 1
Status: Negative
Reference Object:
BFD Session:
Packet type: Echo
Interface  : vlan-interface2
Remote IP  : 192.168.0.101
Local IP   : 192.168.0.102
```

Configuring BFD for the VRRP Master to Monitor the Uplinks

Network requirements

- The master monitors the state of its uplink. When the uplink is down, the mater decreases its priority and sends a VRRP packet with the new priority. Upon receiving the packet with a lower priority, the backup becomes the new master after a very short delay.
- The backup monitors the state of its uplink. Once the uplink state goes down, the backup lowers its priority.

Figure 1-10 Network diagram for monitoring the uplink through VRRP



Configuration procedure

Configure Switch A. The IP address of the uplink device is 1.1.1.2/24.

```
<SwitchA> system-view
[SwitchA] bfd session init-mode active
[SwitchA] bfd echo-source-ip 10.10.10.10
[SwitchA] interface vlan-interface 3
[SwitchA-vlan-interface3] ip address 1.1.1.1 24
[SwitchA-vlan-interface3] bfd min-echo-receive-interval 10
[SwitchA-vlan-interface3] bfd detect-multiplier 3
[SwitchA-vlan-interface3] quit
[SwitchA] track 1 bfd echo interface vlan-interface 3 remote ip 1.1.1.2 local ip 1.1.1.1
[SwitchA] interface vlan-interface 2
[SwitchA-vlan-interface3] ip address 192.168.0.101 24
[SwitchA-vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
[SwitchA-vlan-interface2] vrrp vrid 1 priority 110
[SwitchA-vlan-interface2] vrrp vrid 1 track 1 reduced 20
[SwitchA-vlan-interface2] return
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
[SwitchB-vlan-interface2] ip address 192.168.0.102 24
```

```
[SwitchB–vlan-interface2] vrrp vrid 1 virtual-ip 192.168.0.10
[SwitchB–vlan-interface2] return
```

Use the **display vrrp verbose** command to display the configuration.

Display the detailed information of VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 110                  Run Pri        : 110
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Pri Reduced    : 20
 Virtual IP        : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.101
```

Display the detailed information of VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Master IP         : 192.168.0.101
```

The display above shows that, in VRRP group 1, Switch A is the master router and Switch B the backup router.

When the uplink of Switch A goes down, use the **display vrrp** command to display the detailed information of VRRP group 1 on Switch A.

```
<SwitchA> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Backup
 Config Pri        : 110                  Run Pri        : 90
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
 Track Object      : 1                    Pri Reduced    : 20
 Virtual IP        : 192.168.0.10
```

```
 Master IP         : 192.168.0.102
```

When the uplink of Switch A goes down, display the detailed information of VRRP group 1 on Switch B.

```
<SwitchB> display vrrp verbose
IPv4 Standby Information:
 Run Method        : VIRTUAL-MAC
 Total number of virtual routers: 1
 Virtual IP Ping   : Enable
 Interface         : vlan-interface2
 VRID              : 1                    Adver. Timer   : 1
 Admin Status      : UP                   State          : Master
 Config Pri        : 100                  Run Pri        : 100
 Preempt Mode      : YES                  Delay Time     : 0
 Auth Type         : NONE
Virtual IP     : 192.168.0.10
 Virtual MAC       : 0000-5e00-0101
 Master IP         : 192.168.0.102
```

Display the track entry information of Switch A.

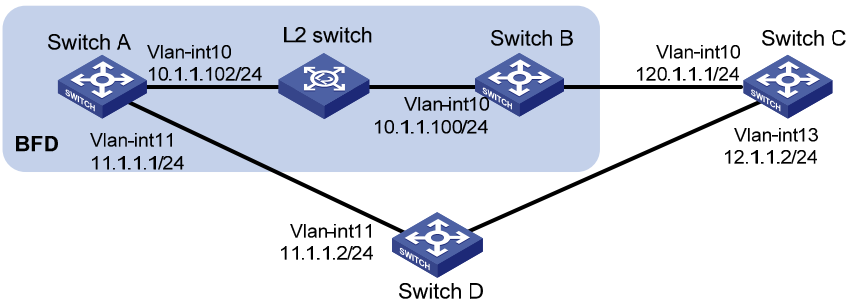
```
<SwitchA> display track 1
Track ID: 1
Status: Negative
Reference Object:
BFD Session:
Packet type: Echo
Interface  : vlan-interface3
Remote IP  : 1.1.1.2
Local IP   : 1.1.1.1
```

Configuring BFD Echo Packet Mode for Static Routing

Network requirements

Configure a static route on Switch A to Switch C and enable BFD. When the link between Switch A and Switch B fails, Switch A selects Switch D to reach Switch C.

Figure 1-11 Network diagram for BFD configuration on a static route



Configuration procedure

1) Configure BFD

Configure a static route on Switch A and enable BFD on it. Implement BFD through BFD echo packets.

```
<SwitchA> system-view
[SwitchA] bfd echo-source-ip 123.1.1.1
[SwitchA] interface vlan-interface 10
[SwitchA-vlan-interface10] bfd min-echo-receive-interval 300
[SwitchA-vlan-interface10] bfd detect-multiplier 7
[SwitchA-vlan-interface10] quit
[SwitchA] ip route-static 120.1.1.1 24 vlan-interface 10 10.1.1.100 bfd echo-packet
[SwitchA] ip route-static 120.1.1.1 24 vlan-interface 11 11.1.1.2 preference 65
[SwitchA] quit
```

2) Verify the configuration

Display BFD session information on Switch A.

```
<SwitchA> display bfd session

Total Session Num: 1            Init Mode: Active

 Session Working Under Echo Mode:

 LD            SourceAddr      DestAddr        State Holdtime Interface
 7             10.1.1.102       10.1.1.100       Up    1700ms   Vlan10
```

3) Display static route information on Switch A.

```
<SwitchA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 2

Static Routing table Status : <Active>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24         Static 65   0           10.1.1.100      Vlan10

Direct Routing table Status : <Inactive>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24        Static 60   0           11.1.1.2        Vlan11
```

Enable BFD debugging on Switch A.

```
<SwitchA> debugging bfd event
<SwitchA> debugging bfd scm
<SwitchA> terminal debugging
```

When the link between Switch B and the Layer-2 switch goes down, Switch A can quickly detect the changes on Switch B.

```
%Nov 12 19:28:28:592 2005 SwitchA BFD/5/LOG:Sess[123.1.1.1/10.1.1.100, Vlan10], Sta:
UP->DOWN, Diag: 1
*0.53892593 SwitchA BFD/8/SCM:Sess[123.1.1.1/10.1.1.100, Vlan10], Oper: Reset
*0.53892593 SwitchA BFD/8/EVENT:Send sess-down Msg, [Src:123.1.1.1, Dst:10.1.1.100, Vlan10]
Protocol: STATIC
*0.53892595 SwitchA RM/7/LOG:static route [Dest:120.1.1.1/24,Nexthop:10.1.1.100,ExitIf:
Vlan10] became invalid
```

Execute the **display ip routing-table protocol static** command, and you can see Switch A selects Switch D to reach Switch C.

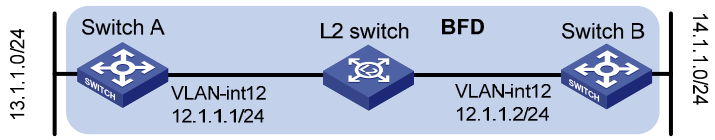
```
<SwitchA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 2

Static Routing table Status : < Active>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24         Static 65   0            11.1.1.2        Vlan11

Static Routing table Status : < Inactive>
Summary Count : 1

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

120.1.1.1/24        Static 60   0            10.1.1.100      Vlan10
```

Configuring BFD Control Packet Mode for Static Routing

Network requirements

Configure a static route to subnet 14.1.1.0/24 on Switch A and configure a static route to subnet 13.1.1.0/24 on Switch B. Both routes have BFD control packet mode enabled. When the link between Switch A and Switch B fails, BFD can detect it immediately.

Figure 1-12 Configure BFD control packet mode for static routing



Configuration procedure

1) Configure BFD

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 12
[SwitchA-vlan-interface12] ip address 12.1.1.1 24
```

```
[SwitchA-vlan-interface12] bfd min-transmit-interval 500
[SwitchA-vlan-interface12] bfd min-receive-interval 500
[SwitchA-vlan-interface12] bfd detect-multiplier 9
[SwitchA-vlan-interface12] quit
[SwitchA] ip route-static 14.1.1.0 24 vlan-interface 12 12.1.1.2 bfd control-packet
[SwitchA] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 12
[SwitchB-vlan-interface12] ip address 12.1.1.2 24
[SwitchB-vlan-interface12] bfd min-transmit-interval 500
[SwitchB-vlan-interface12] bfd min-receive-interval 500
[SwitchB-vlan-interface12] bfd detect-multiplier 9
[SwitchB-vlan-interface12]] quit
[SwitchB] ip route-static 13.1.1.0 24 vlan-interface 12 12.1.1.1 bfd control-packet
[SwitchB] quit
```

2) Verify the configuration.

```
<SwitchA> display bfd session

 Total Session Num: 1            Init Mode: Active

 Session Working Under Ctrl Mode:

 LD/RD         SourceAddr      DestAddr        State Holdtime Interface
 4/7           12.1.1.1        12.1.1.2        Up    2000ms   Vlan12
```

Display static routes on Switch A.

```
<SwitchA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 1

Static Routing table Status : < Active>
Summary Count : 1
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
14.1.1.0/24         Static 60   0            12.1.1.2        Vlan12

Static Routing table Status : < Inactive>
Summary Count : 0
```

Enable BFD debugging on Switch A.

```
<SwitchA> debugging bfd event
<SwitchA> debugging bfd scm
<SwitchA> terminal debugging
```

When the link between Switch A and Layer-2 switch fails, Switch A can detect the failure.

```
%Jul 27 10:18:18:672 2007 SwitchA BFD/4/LOG:Sess[12.1.1.1/12.1.1.2, Vlan12,Ctrl],
 Sta: UP->DOWN, Diag: 1
*Jul 27 10:18:18:672 2007 SwitchA BFD/7/EVENT:Send sess-down Msg, [Src:12.1.1.1,
Dst:12.1.1.2, Vlan12,Ctrl], instance:0, protocol:STATIC
```

```
*Jul 27 10:18:19:172 2007 SwitchA BFD/7/EVENT:Receive Delete-sess, [Src:12.1.1.1
,Dst:12.1.1.2, Vlan12,Ctrl], Direct, Instance:0x0, Proto:STATIC
*Jul 27 10:18:19:172 2007 SwitchA BFD/7/EVENT:Notify driver to stop receiving bf
```

Display the static route on Switch A, which is in the inactive state.

```
<SwitchA> display ip routing-table protocol static
Public Routing Table : Static
Summary Count : 1

Static Routing table Status : < Active>
Summary Count : 0

Static Routing table Status : < Inactive>
Summary Count : 1
Destination/Mask    Proto  Pre  Cost         NextHop         Interface
14.1.1.0/24         Static 60   0            12.1.1.2        Vlan12
```

Table of Contents

1 MCE Overview

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running routing protocols.

MCE Overview

Multi-CE (MCE) enables a switch to function as the CEs of multiple VPN instances in a BGP/MPLS VPN network, thus reducing the investment on network equipment.

Introduction to BGP/MPLS VPN

BGP/MPLS VPN is a kind of PE-based L3VPN technology for service provider VPN solutions. It uses BGP to advertise VPN routes and uses MPLS to forward VPN packets on the service provider backbone.

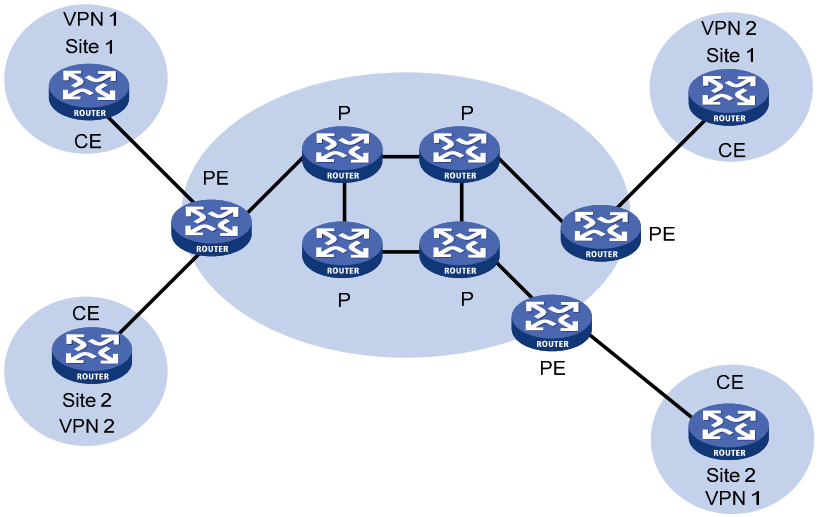
BGP/MPLS VPN provides flexible networking modes, excellent scalability, and convenient support for MPLS QoS and MPLS TE. Hence, it is widely used.

The BGP/MPLS VPN model consists of three kinds of devices:

- Customer edge device (CE): A CE resides on a customer network and has one or more interfaces directly connected with service provider networks. It can be a router, a switch, or a host. It neither can "sense" the existence of any VPN nor needs to support MPLS.
- Provider edge router (PE): A PE resides on a service provider network and connects one or more CEs to the network. On an MPLS network, all VPN processing occurs on the PEs.
- Provider (P) router: A P router is a backbone router on a service provider network. It is not directly connected with any CE. It only needs to be equipped with basic MPLS forwarding capability.

Figure 1-1 illustrates a BGP/MPLS VPN implementation.

Figure 1-1 A BGP/MPLS VPN implementation



CEs and PEs mark the boundary between the service providers and the customers.

A CE is usually a router. After a CE establishes adjacency with a directly connected PE, it redistributes its VPN routes to the PE and learns remote VPN routes from the PE. A CE and a PE use BGP/IGP to exchange routing information. You can also configure static routes between them.

After a PE learns the VPN routing information of a CE, it uses BGP to exchange VPN routing information with other PEs. A PE maintains routing information about only VPNs that are directly connected, rather than all VPN routing information on the provider network.

A P router maintains only routes to PEs. It does not need to know anything about VPN routing information.

When VPN traffic travels over the MPLS backbone, the ingress PE functions as the ingress LSR, the egress PE functions as the egress LSR, while P routers function as the transit LSRs.

You can use 3Com Switch 4800G as the CEs in a BGP/MPLS VPN implementation.

BGP/MPLS VPN Concepts

Site

Site is often mentioned in the VPN, whose meanings are described as follows:

- A site is a group of IP systems with IP connectivity that does not rely on any service provider network to implement.
- The classification of a site depends on the topology relationship of the devices, rather than the geographical positions, though the devices at a site are adjacent to each other geographically in most cases.
- The devices at a site can belong to multiple VPNs, namely, a site can belong to multiple VPNs.
- A site is connected to a provider network through one or more CEs. A site can contain many CEs, but a CE can belong to only one site.

Sites connected to the same provider network can be classified into different sets by policies. Only the sites in the same set can access each other through the provider network. Such a set is called a VPN.

Address space overlapping

Each VPN independently manages the addresses that it uses. The assembly of such addresses for a VPN is called an address space.

The address spaces of VPNs may overlap. For example, if both VPN 1 and VPN 2 use the addresses in network segment 10.110.10.0/24, address space overlapping occurs.

VPN instance

In MPLS VPN, route separation between VPNs is implemented by VPN instance.

A PE creates and maintains a separate VPN instance for each directly connected site. Each VPN instance contains the VPN membership and routing rules of the corresponding site. If a user at a site belongs to multiple VPNs at the same time, the VPN instance of the site contains information about all the VPNs.

For independency and security of VPN data, each VPN instance on a PE maintains a relatively independent routing table and a separate label forwarding information base (LFIB). VPN instance information contains these items: the LFIB, IP routing table, interfaces bound to the VPN instance, and administration information of the VPN instance. The administration information of the VPN instance includes the route distinguisher (RD), route filtering policy, and member interface list.

Note

LFIBs of VPN instances exist on only PEs supporting MPLS. No LFIBs of VPN instances exist on MCE-capable devices.

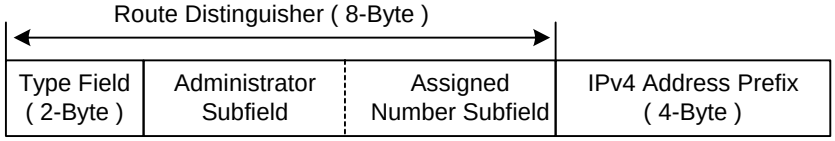
VPN-IPv4 address

Traditional BGP cannot process VPN routes which have overlapping address spaces. If, for example, both VPN 1 and VPN 2 use addresses in the segment 10.110.10.0/24 and advertise a route to the segment, BGP selects only one of them, which results in loss of the other route.

PEs use MP-BGP to advertise VPN routes, and use VPN-IPv4 address family to solve the problem with traditional BGP.

A VPN-IPv4 address consists of 12 bytes. The first eight bytes represent the RD, followed by a 4-byte IPv4 address prefix, as shown in.

Figure 1-2 Structure of a VPN-IPv4 address



When a PE receives an ordinary IPv4 route from a CE, it must redistribute the VPN route to the peer PE. The uniqueness of a VPN route is implemented by adding an RD to the route.

A service provider can independently assign RDs provided the assigned RDs are unique. In this way, a PE can advertise different routes to VPNs even if the VPNs are from different service providers and are using the same IPv4 address space.

You are recommended to configure a distinct RD for each VPN instance on a PE, guaranteeing that routes to the same CE use the same RD. The VPN-IPv4 address with an RD of 0 is in fact a globally unique IPv4 address.

By prefixing a distinct RD to a specific IPv4 address prefix, you make it a globally unique VPN IPv4 address prefix.

An RD can be related to an autonomous system (AS) number, in which case it is the combination of an AS number and a discretionary number; or be related to an IP address, in which case it is the combination of an IP address and a discretionary number.

An RD can be in either of the following two formats distinguished by the Type field:

- When the value of the Type field is 0, the Administrator subfield occupies two bytes, the Assigned number subfield occupies four bytes, and the RD format is: 16-bit AS number:32-bit user-defined number. For example, 100:1.
- When the value of the Type field is 1, the Administrator subfield occupies four bytes, the Assigned number subfield occupies two bytes, and the RD format is: 32-bit IPv4 address:16-bit user-defined number. For example, 172.1.1.1:1.

For the global uniqueness of an RD, you are not recommended to set the Administrator subfield to any private AS number or private IP address.

VPN target attributes

BGP/MPLS VPN uses the BGP extended community attributes called VPN target attributes, or route target attributes, to control the advertisement of VPN routing information.

A VPN instance on a PE supports two types of VPN target attributes:

- Export target attribute: A local PE sets this type of VPN target attribute for VPN-IPv4 routes learnt from directly connected sites before advertising them to other PEs.
- Import target attribute: A PE checks the export target attribute of VPN-IPv4 routes advertised by other PEs. If the export target attribute matches the import target attribute of the VPN instance, the PE adds the routes to the VPN routing table.

In other words, VPN target attributes define which sites can receive a VPN-IPv4 route, and from which sites a PE can receive routes.

Like RDs, VPN target attributes can be of two types of formats:

- 16-bit AS number:32-bit user-defined number. For example, 100:1.
- 32-bit IPv4 address:16-bit user-defined number. For example, 172.1.1.1:1.

Introduction to MCE

With BGP/MPLS VPN, data of private networks can be transmitted in the public network securely through tunnels. However, in a typical BGP/MPLS VPN network, each VPN is connected to the PE through a CE, as shown in Figure 1-1.

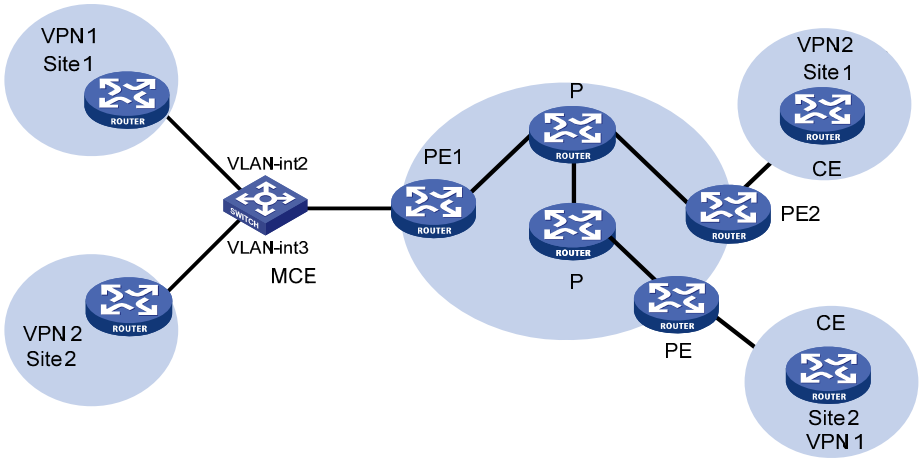
With the users' increasing demand for service segmentation and security, a private network may be divided into multiple VPNs, and the users of different VPN are usually isolated from each other. In a private network containing multiple VPNs, users may be in such a dilemma: equipment investment and the maintenance cost increment caused by assigning a CE for each of the VPNs; and potential data security risks introduced by sharing one CE among multiple VPNs (because the same routing entry may be used in multiple VPNs in this case).

An 3Com Switch 4800G switch with MCE enabled can solve this problem. By binding the VLAN interfaces to the VPNs in a network on an 3Com Switch 4800G of this kind, you can create and maintain a routing table for each of the VPNs. In this way, packets of different VPNs in the private network can be isolated. Moreover, with the cooperation of the PE, the routes of each VPN can be advertised to the corresponding remote PE properly, so that packets of each VPN in the private network can be transmitted securely through the public network.

How MCE Works

Figure 1-3 illustrates how MCE creates and maintains routing entries of multiple VPNs and how the MCE exchanges VPN routes with PEs.

Figure 1-3 How MCE works



In Figure 1-3, the two VPN sites on the left side (Site 1 and Site 2) are connected to the backbone network through an MCE device. Two VPN tunnels are expected between them and the remote VPNs at Site 2 and Site 1.

With MCE enabled, routing tables can be created for VPN 1 and VPN 2 individually, VLAN-interface 2 can be bound to VPN 1, and VLAN-interface 3 can be bound to VPN 2. When receiving a piece of routing information, MCE determines the source of the routing information according to the number of the interface receiving the information and then maintains the corresponding routing table accordingly.

You need to also to bind the interfaces to the VPNs on PE 1 in the same way as those on the MCE device. The MCE device is connected to PE 1 through a trunk, which permits packets of VLAN 2 and VLAN 3 with VLAN tags carried. In this way, PE 1 can determine the VPN a received packet belongs to according to the VLAN tag of the packet and passes the packet to the corresponding tunnel.

Routing Information Exchange for MCE

Interface-to-VPN-instance binding enables CEs and PEs to determine the sources of received packets and then forward the packets according to the routing information concerning the corresponding VPNs. The following sections describe the way how MCE transmits the private routing information of multiple VPNs to PEs properly.

Route Exchange between a CE and the Private Network

A CE can adopt the following routing protocols to exchange VPN routes with a site:

- Static route
- RIP
- OSPF
- IS-IS
- EBGP

Note

This introduces the cooperation of routing protocols and MCE in brief. For details on routing protocols, see the IPv4 Routing module of this manual.

Static routes

A CE can communicate with a site through static routes. As static routes configure for traditional CEs take effect globally, address overlapping between multiple VPNs remains a problem till the emergence of MCE. MCE allows static-route-to-VPN-instance binding, which isolates the static routes of different VPNs.

RIP

An 3Com Switch 4800G can bind RIP processes to VPN instances. With the same binding configured on CE and site, private network routes of different VPNs can be exchanged between CEs and sites through different RIP processes, thus isolating and securing VPN routes.

OSPF

An 3Com Switch 4800G can bind OSPF processes to VPN instances and isolate the routes of different VPNs.

Note that:

For an OSPF process bound to a VPN instance, the router ID of the public network configured in system view is invalid. So you need to specify the router ID when creating an OSPF process.

An OSPF process can be bound to only one VPN instance, however, a VPN instance can use multiple OSPF processes for private network route transmission. To make sure routes can be advertised properly, you need to configure the same domain ID for all the OSPF processes bound to a VPN instance.

Note

Normally, when an OSPF route is imported to the BGP routing table as a BGP route on a PE, some attributes of the OSPF route get lost. When the BGP route is imported to the OSPF routing table on the remote CE, not all the attributes of the original OSPF routes can be restored. As a result, the route cannot be distinguished from the routes imported from other domains. In order to distinguish OSPF routes imported from different OSPF domains, the OSPF routes to be imported to the BGP routing tables on PEs must carry an attribute (the OSPF domain ID) used to identify the OSPF domains. The domain ID of an OSPF process is contained in the routes generated by the process. When an OSPF route is imported to BGP, the domain ID is added to BGP VPN routes as the extended BGP community.

In cases where a VPN have multiple MCE devices attached to it, when a MCE device advertises the routes learned from BGP within the VPN, the routes may be learned by other MCE devices, thus generating route loops. To prevent route loops, you can configure route tags for different VPN instances on each MCE. It is recommended that a VPN be assigned the same route tag on multiple MCEs.

IS-IS

Similar to those in OSPF, IS-IS processes can be bound to VPN instances for private network routes to be exchanged between CEs and sites. An IS-IS process can be bound to only one VPN instance.

EBGP

To use EBGP to exchange private routes between a CE and a site, you need to configure BGP peers for VPN instances on CEs and import IGP routing information from corresponding VPNs. Normally, sites reside in different ASs, so EBGP is used for route exchange. In this case, the following configurations are needed.

1) Configuring to use EBGP to import IGP routes from each site

To advertise private network routes to PEs properly, IGP routes in the sites directly connected to an MCE device need to be first imported to the BGP routing table of the MCE device.

2) Configuring a peer group for each VPN instance

For proper route exchange between a CE and a site, you need to configure a peer group for each VPN instance and assign AS numbers for these peer groups in BGP IPv4 address family view.

3) Applying filtering policies for route filtering

To make sure that routing information is exchanged between sites and PE devices properly, filtering policies are applied to filter routes received or to be advertised.

Route Exchange between CE and PE

Routing information entries are bound to specific VPN instances on a MCE device, and packets of each VPN instance are forwarded between CE and PE according to interface. As a result, VPN routing information can be transmitted by performing relatively simple configurations between CE and PE, such as importing the VPN routing entries on MCE devices to the routing table of the routing protocol running between CEs and PEs.

The following routing protocols can be used between CE and PE for routing formation exchange:

- Static route

- RIP
- OSPF
- IS-IS
- EBGP

For information on how to configure the routing protocols and how to import routes, refer to the *IPv4 Routing* module of this manual.

2 MCE Configuration

Note

For detailed information on the routing protocol configuration mentioned in this chapter, see the *IPv4 Routing* module of this manual.

Configuring a VPN Instance

VPN Instance Configuration Task List

Complete the following tasks to configure a VPN instance:

Task	Remarks
Creating a VPN Instance	Required
Associating an VPN Instance with an Interface	Required
Configuring the Route-related Attributes for a VPN Instance	Required

Creating a VPN Instance

A VPN instance needs to be associated with a site. A VPN instance does not correspond to a VPN directly. Instead, a VPN instance is an integration of the VPN membership and routing rules of its corresponding site.

A VPN instance takes effect only after a route distinguisher (RD) is configured for it. For a VPN instance with the RD not configured, all the other settings (except the description information) are inaccessible.

The description information of a VPN instance can be used to record the relationship between the VPN instance and a VPN.

Follow these steps to create a VPN instance:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a VPN instance and enter VPN instance view	**ip vpn-instance** *vpn-instance-name*	Required By default, no VPN instance is present.
Configure an RD for the VPN instance	**route-distinguisher** *route-distinguisher*	Required By default, a VPN instance has no RD configured.

To do…	Use the command…	Remarks
Set the description information for the VPN instance	**description** *text*	Optional By default, a VPN instance has no description configured.

Caution

The RD configured for a VPN instance on the MCE device must be same as that configured for the VPN instance on the PE device.

Associating an VPN Instance with an Interface

After creating a VPN instance, you need to associate it with the interface connecting a site to a PE.

Follow these steps to associate an VPN instance with an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view of the interface to be associated	**interface** *interface-type interface-number*	—
Associate the current interface with the VPN instance	**ip binding vpn-instance** *vpn-instance-name*	Required By default, no interface is associated with a VPN instance.

Note

Executing the **ip binding vpn-instance** command invalidates the IP address configured for the current interface, so you need to configure an IP address for an interface again after associating the interface with a VPN instance.

Configuring the Route-related Attributes for a VPN Instance

The process of advertising VPN routes is as follows:

- When the switch learns a VPN route from a site and injects it into BGP, BGP associates the route with a VPN target extended community attribute list, which is normally the export route attribute list of the VPN instance.
- A VPN instance determines whether to accept a received route according to the VPN target import extended community attribute list associated with the route and the setting specified by the **vpn-target** command.
- A VPN instance modifies the export VPN target attribute according to the setting specified by the **vpn-target** command.

Follow these steps to configure the route-related attributes for a VPN instance:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VPN instance view	**ip vpn-instance** *vpn-instance-name*	—
Associate the current VPN instance with one or multiple VPN targets	**vpn-target** *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]	Required By default, a VPN instance has no VPN target associated with it.
Configure the maximum number of routes a VPN instance can accommodate	**routing-table limit** *number* { *warn-threshold* | **simply-alert** }	Optional Not configured by default
Apply a route policy for the routes received	**import route-policy** *route-policy*	Optional By default, all routes matching the VPN target attribute are permitted.
Apply a route policy for the routes to be advertised	**export route-policy** *route-policy*	Optional By default, all routes matching the VPN target attribute are permitted.

Note

- This attribute can be advertised with a route only when BGP runs between the MCE and the PE. Otherwise, this attribute is of no sense.
- The VPN target specified for a VPN instance on the MCE device must be same as that specified for the VPN instance on the PE device.

Configuring Route Exchange between a MCE and a Site

Configuring Route Exchange between a MCE and a Site

Complete the following tasks to configure route exchange between a MCE and a site:

Task	Remarks
Configuring to Use Static Routes between a MCE and a Site	You can choose one or multiple configurations as required.
Configuring to Use RIP between a MCE and a Site	
Configuring to Use OSPF between a MCE and a Site	
Configuring to Use IS-IS between a MCE and a Site	
Configuring to Use EBGP between a MCE and a Site	

Configuring to Use Static Routes between a MCE and a Site

Follow these steps to configure static routes between a MCE and a site:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define a static route for a VPN instance	**ip route-static vpn-instance** *s-vpn-instance-name&<1-5> dest-address* { *mask* | *mask-length* } { *gateway-address* [**public**] | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Required This operation is performed on the MCE device. The corresponding configuration on the site is the same as configuring a normal static route.

Configuring to Use RIP between a MCE and a Site

A RIP process can be bound to only one VPN instance. RIP processes not bound to any VPN instances belong to the public network.

Follow these steps to configure RIP between a MCE and a site:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable RIP for a VPN instance (This operation also leads you to RIP view)	**rip** [*process-id*] **vpn-instance** *vpn-instance-name*	Required This operation is performed on the MCE device. As for the corresponding configuration on the site, you can just enable RIP as usual.
Redistribute routes from the remote site advertised by the PE	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	Required By default, RIP does not redistribute routes from other protocols.

Note

After enabling RIP for a VPN instance, you need also to configure to use RIP for routing information exchange.

Configuring to Use OSPF between a MCE and a Site

An OSPF process can be bound to only one VPN instance. OSPF processes not bound to any VPN instances belong to the public network.

Follow these steps to configure OSPF between a MCE and a site:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF for a VPN instance (this operation also leads you to OSPF view)	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *vpn-instance-name*] *	Required This operation is performed on the MCE device. As for the corresponding configuration on the site, you can just enable OSPF as usual.
Enable multi-VPN-instance CE for the OSPF process	**vpn-instance-capability simple**	Required Disabled by default
Redistribute routes from the remote site advertised by the PE	**import-route** *protocol* [*process-id* | **allow-ibgp**] [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] *	Required Not redistributed by default
Configure the OSPF domain ID	**domain-id** *domain-id* [**secondary**]	Optional By default, the OSPF domain ID is 0. This operation is performed on the MCE device. As for the corresponding configuration on the site, you can just enable OSPF as usual.
Configure the type codes of OSPF extended community attributes	**ext-community-type** { **domain-id** *type-code1* | **router-id** *type-code2* | **route-type** *type-code3* }	Optional The defaults are as follows: 0x0005 for Domain ID, 0x0107 for Router ID, and 0x0306 for Route Type.

Note

- Router IDs of the public network configured in system view do not applies to OSPF processes bound to VPN instances. So you need to configure the Router ID after enabling OSPF for a VPN instance.
- To make sure routes can be advertised properly, you need to configure the same domain ID for all the OSPF processes bound to a VPN.
- After enabling OSPF for a VPN instance, you need also to configure to use OSPF for routing information exchange, which is described in the *IPv4 Routing* module of this manual.

Configuring to Use IS-IS between a MCE and a Site

An IS-IS process can be bound to only one VPN instance. IS-IS processes not bound to any VPN instances belong to the public network.

Follow these steps to configure IS-IS between a MCE and a site:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable IS-IS for a VPN instance and enter IS-IS view	**isis** [*process-id*] **vpn-instance** *vpn-instance-name*	Required This operation is performed on the MCE device. As for the corresponding configuration on the site, you can just enable IS-IS as usual.
Redistribute routes from the remote site advertised by the PE	**import-route** { **isis** [*process-id*] | **ospf** [*process-id*] | **rip** [*process-id*] | **bgp** [**allow-ibgp**] | **direct** | **static** } [**cost** *cost* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	Required By default, IS-IS does not redistribute routes from other protocols. Without the level specified, the command redistributes routes to the Level-2 routing table by default.

Note

After enabling IS-IS for a VPN instance, you need also to configure to use IS-IS for routing information exchange.

Configuring to Use EBGP between a MCE and a Site

1) Configuration on the MCE device

Follow these steps to configure an MCE device:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter BGP-VPN instance view	**ipv4-family vpn-instance** *vpn-instance-name*	Required
Configure a CE as a VPN peer	**peer** { *group-name* | *ip-address* } [**as-number** *as-number*]	Required
Redistribute routes from the remote site advertised by the PE	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required By default, EBGP does not redistribute routes from other protocols.
Apply a filter policy to the routes to be advertised	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	Optional By default, routes to be advertised are not filtered.

To do…	Use the command…	Remarks
Apply a filter policy to routes received	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	Optional By default, received routes are not filtered.
Configure to permit the routes with their AS numbers contained in their AS_PATH attributes being the local AS number (this operation can also specify the number of the AS number occurrences allowed)	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	Optional By default, routes with their AS numbers contained in their AS_PATH attributes being the local AS number are denied.

Note

AS number contained in the AS_PATH attribute can be used for route loop detect. With EBGP running between a MCE and a site, the routes advertised by an MCE device to the site carry the local AS number. So do the routes advertised by the site. In this case, you need to configure to permit the routes with their AS numbers contained in their AS_PATH attributes being the local AS number on MCE devices for the routes advertised by the site to be received and processed by the MCE device.

2) Configuration on the site

Note

The site configuration procedures vary with device model. The following takes an 3Com Switch 4800G as an example. As for switches from other vendors, refer to the corresponding user manuals.

Follow these steps to configure the site:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Configure the PE as a peer	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	Required
Configure to import routes	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	Optional The site must advertise the addresses of the reachable VPN segments to the MCE connected to the site.

Note

In a VPN instance with BGP enabled, the BGP route exchange is processed in the same way as those in a normal BGP-enabled network.

Configuring Route Exchange between a MCE and a PE

Configuring Route Exchange between a MCE and a PE

Complete the following tasks to configure route exchange between a MCE and a PE:

Task	Remarks
Configuring to Use Static Routes between a MCE and a PE	You can choose one or multiple configurations as required.
Configuring to Use RIP between a MCE and a PE	
Configuring to Use OSPF between a MCE and a PE	
Configure to Use IS-IS between a MCE and a PE	
Configure to Use EBGP between a MCE and a PE	

Configuring to Use Static Routes between a MCE and a PE

Follow these steps to define a static route for a VPN instance:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Define a static route for a VPN instance	**ip route-static** *dest-address* { *mask* | *mask-length* } { *gateway-address* | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Required By default, for a static route, the preference value is 60, the tag value is 0, and no description information is configured.
	ip route-static vpn-instance *s-vpn-instance-name&<1-6> dest-address* { *mask* | *mask-length* } { *gateway-address* [**public**] | *interface-type interface-number* [*gateway-address*] | **vpn-instance** *d-vpn-instance-name gateway-address* } [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	
Set the default preference value of static routes	**ip route-static default-preference** *default-preference-value*	Optional By default, the preference value of a static route is 60.

Note

- A static route configured for a VPN instance does not take effect if you configure the next hop address of the route as the IP address of a local interface (such as Ethernet interface, VLAN interface).
- If the default static route preference is not configured, the preference of a newly defined static route adopts the system default preference value, which is 60. A customized default static route preference has no effect on existing static routes.
- Static routes can be controlled selectively by using routing policies according to the tag values set for static routes.
- The **ip route-static** command defines a default route if both the destination address and the mask are set to 0.0.0.0.

Configuring to Use RIP between a MCE and a PE

When configuring to use RIP between a MCE and a PE, you need to configure the RIP processes to be bound to the VPN instances and manually import the VPN routes in the site maintained by the MCE device to the routing table of the PE.

Follow these steps to enable RIP for a VPN instance:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable RIP for a VPN instance and enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	Required
Set the default cost for imported routes	**default cost** *value*	Optional By default, the cost for an imported route is 0.
Import the VPN routes of the site	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	Required By default, RIP does not import routes from other protocols.

Configuring to Use OSPF between a MCE and a PE

When configuring to use OSPF between a MCE and a PE, you need to configure the OSPF processes to be bound to VPN instances and router IDs; you also need to manually import the VPN routes in the site maintained by the MCE device to the routing table of the PE.

Follow these steps to configure OSPF between a MCE and a PE:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable OSPF for a VPN instance and enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required

To do…	Use the command…	Remarks
Enable OSPF to import routes of other protocols	**import-route** *protocol* [*process-id* | **allow-ibgp**] [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] *	Required By default, OSPF does not import the routes of other protocols.
Apply a filter policy for imported routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*]]	Optional By default, no filter policy is applied.
Configure the default values of the route attributes (including cost, limit, tag, and type) for the routes imported	**default** { **cost** *cost* | **limit** *limit* | **tag** *tag* | **type** *type* } *	Optional The default route attributes are as follows: the cost value is 1, the upper limit on the maximum number of imported external routes is 1,000, the tag value is 1, and the type is type2.

Configure to Use IS-IS between a MCE and a PE

When configuring to use IS-IS between a MCE and a PE, you need to configure the IS-IS processes to be bound to VPN instances and manually import the VPN routes in the site maintained by the MCE device to the routing table of the PE.

In IS-IS, routes discovered by other routing protocols are external routes. While importing routes of other protocols, you can specify the default cost value for the imported routes as well. You can also apply filter policies for imported routes.

Follow these steps to configure IS-IS to import external routes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable IS-IS for a VPN instance and enter IS-IS view	**isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]	—
Import routes of other protocols	**import-route** { **isis** [*process-id*] | **ospf** [*process-id*] | **rip** [*process-id*] | **bgp** [**allow-ibgp**] | **direct** | **static** } [**cost** *cost* | **cost-type** { **external** | **internal** } | [**level-1** | **level-1-2** | **level-2**] | **route-policy** *route-policy-name* | **tag** *tag*] *	Required By default, IS-IS does not import routes of other protocols. If none of the **level-1**, **level-2**, and **level-3** keywords is specified, the external routes are imported to level-2 routing table.
Apply a filter policy for imported routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export** [**isis** *process-id* **| ospf** *process-id* | **rip** *process-id* | **bgp** | **direct** | **static**]	Optional By default, no filter policy is applied.

Configure to Use EBGP between a MCE and a PE

To use EBGP to exchange routing information between a MCE and a PE, you need to configure the peer end as a peer in the BGP-VPNs on both ends, import VPN routes in the site to the MCE, and then advertise these routes to the PE.

Follow these steps to configure EBGP between a MCE and a PE:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter BGP-VPN instance view	**ipv4-family vpn-instance** *vpn-instance-name*	Required
Configure the PE as a VPN peer	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	Required
Import routes of the local site	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required The MCE device must import routes of the local site to the VPN routing table in order to advertise these routes to the PE device.
Apply a filter policy for routes to be advertised	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* | **ospf** *process-id* | **rip** *process-id* | **static**]	Optional By default, no filter policy is applied.
Apply a filter policy for received routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	Optional By default, no filter policy is applied.

Displaying and Maintaining MCE

To do…	Use the command…	Remarks
Display the IP routing tables associated with a VPN instance	**display ip routing-table vpn-instance** *vpn-instance-name* [**verbose**]	Available in any view
Display the information about a VPN instance	**display ip vpn-instance** [**instance-name** *vpn-instance-name*]	Available in any view
Display information about a specified BGP VPNv4 peer group	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **group** [*group-name*]	Available in any view
Display information about BGP VPNv4 routes injected into a specified VPN instance	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **network**	Available in any view
Display BGP VPNv4 AS path information	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **paths** [*as-regular-expression*]	Available in any view

To do…	Use the command…	Remarks
Display information about BGP VPNv4 peers	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **peer** [*group-name* **log-info** | *ip-address* { **log-info** | **verbose** } | **verbose**]	Available in any view
Display the BGP VPNv4 routing information of a specified VPN instance	**display bgp vpnv4 vpn-instance** *vpn-instance-name* **routing-table** [*network-address* [{ *mask* | *mask-length* } [**longer-prefixes**]] | **as-path-acl** *as-path-acl-number* | **cidr** | **community** [*aa:nn*]&<1-13>[**no-export-subconfed** | **no-advertise** | **no-export**]* [**whole-match**] | **community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16> | **dampened** | **dampening parameter** | **different-origin-as** | **flap-info** [**as-path-acl** *as-path-acl-number* | *network-address* [*mask* [**longer-match**]] | *mask-length* [**longer-match**]]] | **regular-expression** *as-regular-expression*] | **peer** *ip-address* { **advertised-routes** | **received-routes** } | **regular-expression** *as-regular-expression* | **statistic**]	Available in any view
Perform a soft reset of the BGP connections in a specified VPN instance	**refresh bgp vpn-instance** *vpn-instance-name* { *ip-address* | **all** | **external** | **group** *group-name* } { **export** | **import** }	Available in user view
Reset the BGP connections of a VPN instance	**reset bgp vpn-instance** *vpn-instance-name* { *as-number* | *ip-address* | **all** | **external** | **group** *group-name* }	Available in user view
Clear the route flap dampening information of a VPN instance	**reset bgp vpn-instance** *vpn-instance-name* **dampening** [*network-address* [*mask* | *mask-length*]]	Available in user view
Clear route flap history information about a BGP peer of a VPN instance	**reset bgp vpn-instance** *vpn-instance-name ip-address* **flap-info** **reset bgp vpn-instance** *vpn-instance-name* **flap-info** [*ip-address* [*mask* | *mask-length*] | **as-path-acl** *as-path-acl-number* | **regexp** *as-path-regexp*]	Available in user view

Note

The above table lists only the commands used to display VPN instance-related information is displayed. For information about the commands used to display routing protocol configuration, see relevant chapters in the IPv4 Routing module of this manual.

MCE Configuration Example

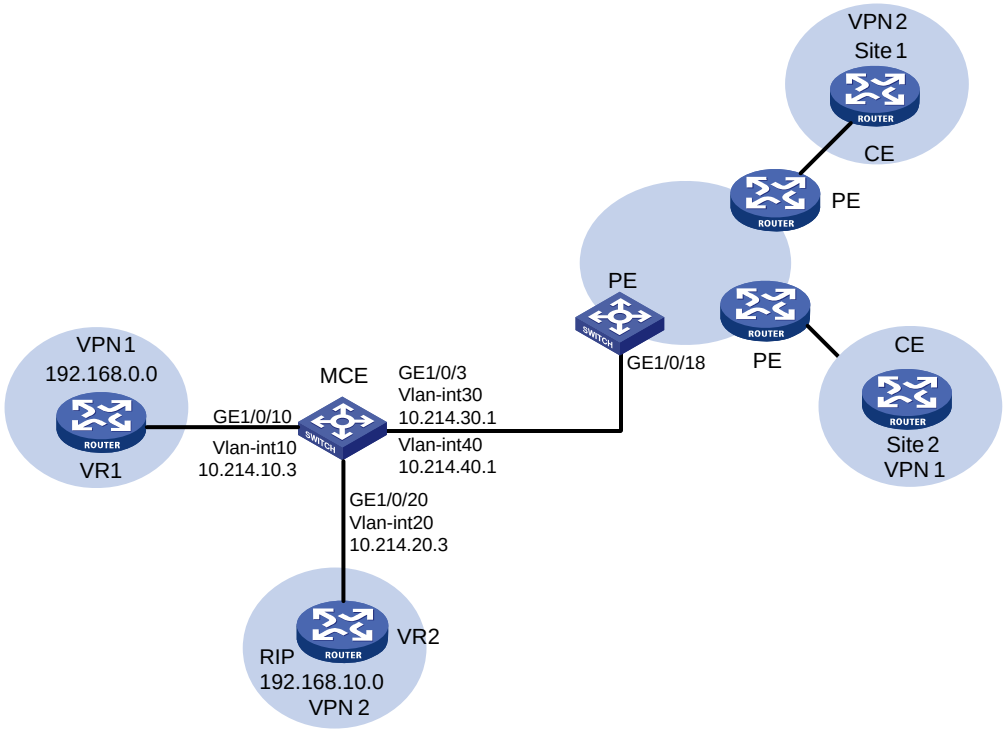
MCE Configuration Example (A)

Network requirements

- An MCE device connects to VPN1 (with the address range being 192.168.0.0/16) through VLAN-interface 10 (with the IP address being 10.214.10.3) and connects to VPN2 (with the address range being 192.168.10.0/24) through VLAN-interface 20 (with the IP address being 10.214.20.3). VPN2 has RIP enabled.
- The MCE device connects to the PE through VLAN-interface 30 and VLAN-interface 40, whose IP addresses are 192.168.30.1/30 and 192.168.40.1/30, respectively.
- It is required that the MCE device isolates routes of VPN1 from those of VPN2 and advertises all the VPN routes to the PE device using OSPF.

Network diagram

Figure 2-1 Network diagram for MCE configuration (A)



Configuration procedure

For distinguish devices, assume the system name of the MCE device is “MCE”, the names of the egress router of VPN1 and VPN2 are “VR1” and “VR2”, and the system name of the PE device is “PE”.

- Configure VPN instances

Configure two instances VPN1 and VPN2 on the MCE device, with the RD values of the two VPN instances being 10:1 and 20:1.

```
<MCE> system-view
[MCE] ip vpn-instance vpn1
[MCE-vpn-instance-vpn1] route-distinguisher 10:1
[MCE-vpn-instance-vpn1] quit
```

```
[MCE] ip vpn-instance vpn2
[MCE-vpn-instance-vpn2] route-distinguisher 20:1
```

Create VLAN 10, add GigabitEthernet 1/0/10 to VLAN 10, and create VLAN-interface 10.

```
[MCE-vpn-instance-vpn2] quit
[MCE] vlan 10
[MCE-vlan10] port GigabitEthernet 1/0/10
[MCE-vlan10] quit
[MCE] interface Vlan-interface 10
```

Bind VLAN-interface 10 to VPN1, and configure IP address 10.214.10.3/24 for VLAN-interface 10.

```
[MCE-Vlan-interface10] ip binding vpn-instance vpn1
[MCE-Vlan-interface10] ip address 10.214.10.3 24
```

Create VLAN 20, add GigabitEthernet 1/0/20 to VLAN 20, create VLAN-interface 20, bind VLAN-interface 20 to VPN2, and configure IP address 10.214.20.3/24 for VLAN-interface 20.

```
[MCE-Vlan-interface10] quit
[MCE] vlan 20
[MCE-vlan20] port GigabitEthernet 1/0/20
[MCE-vlan20] quit
[MCE] interface Vlan-interface 20
[MCE-Vlan-interface20] ip binding vpn-instance vpn2
[MCE-Vlan-interface20] ip address 10.214.20.3 24
[MCE-Vlan-interface20] quit
```

Create VLAN 30, VLAN 40 and the corresponding VLAN interfaces. Then bind VLAN 30 to VPN 1, and VLAN 40 to VPN 2, and configure IP addresses of the VLAN interfaces.

```
[MCE] vlan 30
[MCE-vlan30] quit
[MCE] interface Vlan-interface 30
[MCE-Vlan-interface30] ip binding vpn-instance vpn1
[MCE-Vlan-interface30] ip address 10.214.30.1 30
[MCE-Vlan-interface30] quit
[MCE] vlan 40
[MCE-vlan40] quit
[MCE] interface Vlan-interface 40
[MCE-Vlan-interface40] ip binding vpn-instance vpn2
[MCE-Vlan-interface40] ip address 10.214.40.1 30
[MCE-Vlan-interface40] quit
```

- Configure the routing protocol running between MCE and a site

MCE is directly connected to VPN1, which has no routing protocol enabled. You can configure to use static routes between MCE and a site.

Configuration on VR1: Assume VR1 is an 3Com Switch 4800G, configure IP address 10.214.10.2/24 for the interface connecting to MCE and IP address 192.168.0.1/24 for the interface connecting to VPN1. The operation of adding a port to a VLAN and configuring IP address for a VLAN-interface is omitted here.

Configure a default route on VR1, specifying the next hop address to 10.214.10.3.

```
<VR1> system-view
[VR1] ip route-static 0.0.0.0 0.0.0.0 10.214.10.3
```

Define a static route on MCE, specify the next hop address 10.214.10.2 for packets destined for the network segment 192.168.0.0, and bind this route to VPN1.

```
[MCE-Vlan-interface10] quit
[MCE] ip route-static vpn-instance vpn1 192.168.0.0 16 10.214.10.2
```

Display the information about the routes of VPN1 maintained on MCE.

```
[MCE] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0     0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0     0            127.0.0.1       InLoop0
10.214.10.0/24      Direct 0    0            10.214.10.3     Vlan10
10.214.10.3/32      Direct 0    0            127.0.0.1       InLoop0
192.168.0.0/16     Static 60    0            10.214.10.2     Vlan10
```

As shown in the displayed information above, a static route has been specified for VPN1.

On VR2, configure IP address 10.214.20.2/24 for the interface connecting to MCE (configuration procedures omitted), enable RIP and advertise the network segments 192.168.10.0 and 10.214.20.0.

```
<VR2> system-view
[VR2] rip 20
[VR2-rip-20] network 192.168.10.0
[VR2-rip-20] network 10.0.0.0
```

RIP is running within VPN2, so you can configure RIP on MCE and involve the RIP on MCE in the routing computation in the site to update the routing information automatically. Create RIP process 20, disable automatic route summarization, redistribute routes from OSPF process 20, and bind the RIP process to VPN2.

```
[MCE] rip 20 vpn-instance vpn2
```

Advertise the network segment 10.214.20.0 and 10.214.40.0.

```
[MCE-rip-20] network 10.0.0.0
[MCE-rip-20] undo summary
[MCE-rip-20] import-route ospf
```

Display the information about the routes of VPN2 on MCE.

```
[MCE-rip-20] display ip routing-table vpn-instance vpn2
Routing Tables: vpn2
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.214.20.0/24      Direct 0    0            10.214.20.3     Vlan20
10.214.20.3/32      Direct 0    0            127.0.0.1       InLoop0
10.214.40.0/30      Direct 0    0            10.214.40.1     Vlan40
10.214.40.1/32      Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

```
192.168.10.0/24     RIP    100  1            10.214.20.2     Vlan20
```

As shown in the displayed information above, MCE has obtained the routes of VPN2 through RIP, and maintains these routes in a routing table different from the routing table for routing information of VPN1 to the network segment 192.168.0.0, thus isolating the routes of VPN1 from the routes of VPN2.

- Configure the routing protocol running between the MCE and a PE

MCE uses GigabitEthernet 1/0/3 to connect to GigabitEthernet 1/0/18 of PE. Configure the two ports to be trunk ports and permit tagged packets of VLAN 30 and VLAN 40.

```
[MCE-rip-20] quit
[MCE] interface GigabitEthernet 1/0/3
[MCE-GigabitEthernet1/0/3] port link-type trunk
[MCE-GigabitEthernet1/0/3] port trunk permit vlan 30 40
```

Configure GigabitEthernet 1/0/18 of PE.

```
<PE> system-view
[PE] interface GigabitEthernet 1/0/18
[PE-GigabitEthernet1/0/18] port link-type trunk
[PE-GigabitEthernet1/0/18] port trunk permit vlan 30 40
```

Configure IP addresses 10.214.30.2 and 10.214.40.2 for VLAN-interface 30 and VLAN-interface 40 of PE respectively. The configuration procedures are omitted here.

Configure Loopback0 of MCE and CE to specify the router ID for MCE and PE respectively. The IP addresses for Loopback0 of MCE and CE are 101.101.10.1 and 100.100.10.1 respectively. Configuration procedures are omitted here.

Create OSPF process 10 on MCE, bind the process to VPN1, and set the OSPF domain ID to 10, and enable OSPF multi-instance.

```
[MCE-GigabitGigabitEthernet1/0/3] quit
[MCE] ospf 10 router-id 101.101.10.1 vpn-instance vpn1
[MCE-ospf-10] domain 10
[MCE-ospf-10] vpn-instance-capability simple
```

Advertise the network segment 10.214.30.0 within Area0, and import static routes of VPN1.

```
[MCE-ospf-10] area 0
[MCE-ospf-10-area-0.0.0.0] network 10.214.30.0 0.0.0.255
[MCE-ospf-10-area-0.0.0.0] quit
[MCE-ospf-10] import-route static
```

Create OSPF process 10 on PE, bind the process to VPN1, set the OSPF domain ID to 10, enable OSPF multi-instance, and advertise the network segment 10.214.30.0 within Area0.

```
[PE-GigabitEthernet1/0/18] quit
[PE] ospf 10 router-id 100.100.10.1 vpn-instance vpn1
[PE-ospf-10] domain-id 10
[PE-ospf-10] vpn-instance-capability simple
[PE-ospf-10] area 0
[PE-ospf-10-area-0.0.0.0] network 10.214.30.0 0.0.0.255
```

Display the information about the routes of VPN1 on PE.

```
[PE-ospf-10-area-0.0.0.0] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
```

```
          Destinations : 6        Routes : 6

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
10.214.30.0/24      Direct 0    0            10.214.30.1     Vlan30
10.214.30.2/32      Direct 0    0            127.0.0.1       InLoop0
100.100.10.1/32     Direct 0    0            127.0.0.1       InLoop0
192.168.0.0/16      O_ASE  150  1            10.214.30.1     Vlan30
```

As shown in the displayed information above, the static routes of VPN1 have been imported to the OSPF routing table between MCE and PE.

Create OSPF process 20 and import the routing information of VPN2, which is similar to the above procedure. The only difference lies in that RIP routes rather than static routes are imported to the OSPF routing table of MCE. The detailed configuration procedures are omitted here. The information displayed below verifies the configuration.

```
<PE> display ip routing-table vpn-instance vpn2
display ip routing-table vpn-instance vpn2
Routing Tables: vpn2
          Destinations : 6        Routes : 6

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
10.214.40.0/24      Direct 0    0            10.214.40.1     Vlan40
10.214.40.2/32      Direct 0    0            127.0.0.1       InLoop0
200.200.20.1/32     Direct 0    0            127.0.0.1       InLoop0
192.168.10.0/24     O_ASE  150  1            10.214.40.1     Vlan40
```

After the above configurations, the routing information of VPN1 and VPN2 can be advertised to PE properly.

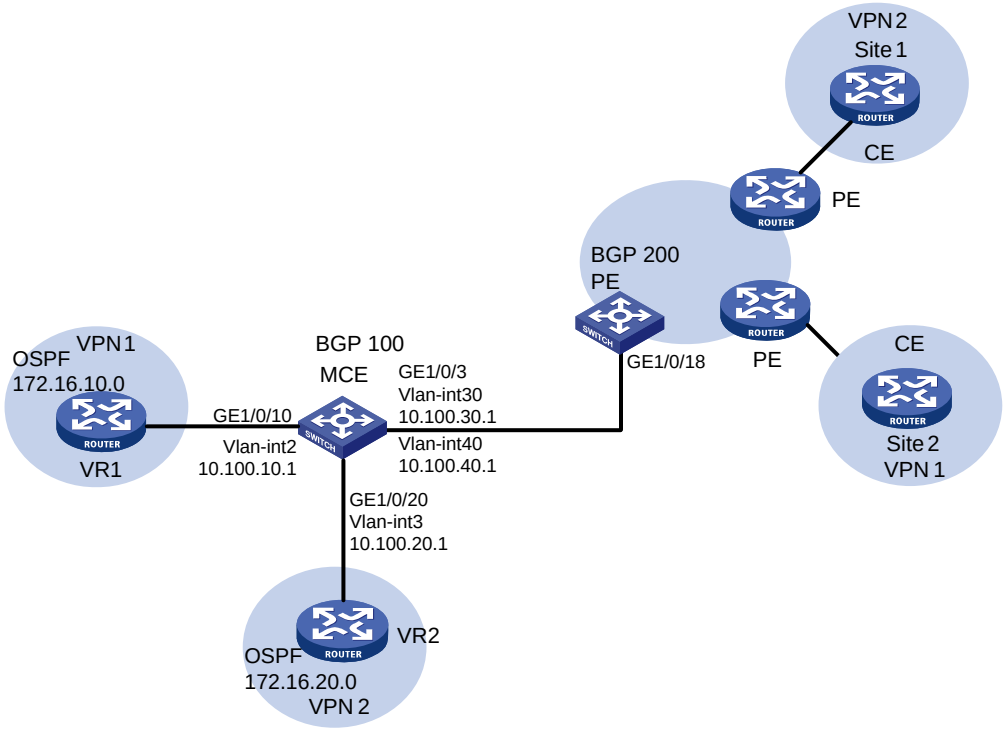
MCE Configuration Example (B)

Network requirements

- An 3Com Switch 4800G functions as MCE. It is required that VPN routes of site 1 and site 2 be advertised to the PE for the purpose that VPNs at both ends of the MPLS backbone network can communicate with each other properly.
- OSPF runs within both site 1 and site 2, and EBGP runs between MCE and PE.

Network diagram

Figure 2-2 Network diagram for MCE configuration (B)



Configuration procedure

- Configure VPN instances

Configure two instances VPN1 and VPN2 on the MCE device, with the RD values of the two VPN instances being 10:1 and 20:1. Configure the VPN target values of the two VPN instances as 10:1 and 20:1 for both the import and export extended community attribute list.

```
<MCE> system-view
[MCE] ip vpn-instance vpn1
[MCE-vpn-instance-vpn1] route-distinguisher 10:1
[MCE-vpn-instance-vpn1] vpn-target 10:1 both
[MCE-vpn-instance-vpn1] quit
[MCE] ip vpn-instance vpn2
[MCE-vpn-instance-vpn2] route-distinguisher 20:1
[MCE-vpn-instance-vpn2] vpn-target 20:1 both
```

Create VLAN 2, add GigabitEthernet 1/0/10 to VLAN 2, and create VLAN-interface 2.

```
[MCE-vpn-instance-vpn2] quit
[MCE] vlan 2
[MCE-vlan2] port GigabitEthernet 1/0/10
[MCE-vlan2] quit
[MCE] interface Vlan-interface 2
```

Bind VLAN-interface 2 to VPN1, and configure IP address 10.214.10.3/24 for VLAN-interface 2.

```
[MCE-Vlan-interface2] ip binding vpn-instance vpn1
[MCE-Vlan-interface2] ip address 10.214.10.3 24
```

Create VLAN 3, add GigabitEthernet 1/0/20 to VLAN 3, create VLAN-interface 3, bind VLAN-interface 3 to VPN2, and configure IP address 10.214.20.3/24 for VLAN-interface 3.

```
[MCE-Vlan-interface10] quit
[MCE] vlan 3
[MCE-vlan3] port GigabitEthernet 1/0/20
[MCE-vlan3] quit
[MCE] interface Vlan-interface 3
[MCE-Vlan-interface3] ip binding vpn-instance vpn2
[MCE-Vlan-interface3] ip address 10.214.20.3 24
[MCE-Vlan-interface3] quit
```

Create VLAN 30, VLAN 40 and the corresponding VLAN interfaces. Then bind VLAN 30 to VPN 1, and VLAN 40 to VPN 2, and configure IP addresses of the VLAN interfaces.

```
[MCE] vlan 30
[MCE-vlan30] quit
[MCE] interface Vlan-interface 30
[MCE-Vlan-interface30] ip binding vpn-instance vpn1
[MCE-Vlan-interface30] ip address 10.214.30.1 30
[MCE-Vlan-interface30] quit
[MCE] vlan 40
[MCE-vlan40] quit
[MCE] interface Vlan-interface 40
[MCE-Vlan-interface40] ip binding vpn-instance vpn2
[MCE-Vlan-interface40] ip address 10.214.40.1 30
[MCE-Vlan-interface40] quit
```

- Configure the routing protocol running between MCE and a site

The procedure of enabling OSPF in the two VPN instances and advertising the network segments is the same as that in normal OSPF and is omitted.

Create OSPF process 10 for MCE whose router ID is 10.10.10.1, bind the process to VPN1. Redistribute BGP routes from VPN1, enable OSPF multi-instance, and advertise the network segment 10.100.10.0.

```
<MCE> system-view
[MCE] ospf 10 router-id 10.10.10.1 vpn-instance vpn1
[MCE-ospf-10] vpn-instance capability simple
[MCE-ospf-10] import-route bgp
[MCE-ospf-10] area 0
[MCE-ospf-10-area-0.0.0.0] network 10.100.10.0 0.0.0.255
```

Display the information about the routes of VPN1.

```
[MCE-ospf-10-area-0.0.0.0] display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
10.100.10.0/24      Direct 0    0            10.100.10.1     Vlan2
```

```
10.100.10.1/32      Direct 0    0           127.0.0.1       InLoop0
172.16.10.0/24      OSPF   10   1           10.100.10.2     Vlan2
```

As shown in the displayed information above, MCE has obtained the routing information of VPN1 through OSPF process 10.

Create OSPF process 20 for MCE whose router ID is 10.10.20.1, bind the process to VPN2. Redistribute BGP routes from VPN2, enable OSPF multi-instance, and advertise the network segment 10.100.20.0. The procedure of configuring OSPF process 20 is similar to that of configuring OSPF process 10. Followed is the result of the above configuration.

```
[MCE] display ip routing-table vpn-instance vpn2
Routing Tables: vpn2
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0    0           127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0           127.0.0.1       InLoop0
10.100.20.0/24      Direct 0    0           10.100.20.1     Vlan3
10.100.20.1/32      Direct 0    0           127.0.0.1       InLoop0
172.16.20.0/24      OSPF   10   1           10.100.20.2     Vlan3
```

- Configure the routing protocol running between MCE and PE

The procedure of connecting MCE to PE through trunk ports is similar to that in MCE Configuration Example (A) and is omitted here.

Create BGP process 10 for MCE.

```
[MCE] bgp 100
[MCE-bgp]
```

Enter IPv4 address family view in VPN1.

```
[MCE-bgp] ipv4-family vpn-instance vpn1
[MCE-bgp-vpn1]
```

Configure PE as an EBGP peer and import the routing information of OSPF process 10 (assuming that the address of the interface bound to VPN1 is 10.100.30.3 and the ID of the BGP process is 200).

```
[MCE-bgp-vpn1] peer 10.100.30.3 as-number 200
[MCE-BGP-vpn1] import-route ospf 10
```

Create BGP process 200 on the PE, and configure MCE as an EBGP peer.

```
<PE> system-view
[PE] bgp 200
[PE-bgp] ipv4-family vpn-instance vpn1
[PE-bgp-vpn1] peer 10.100.30.1 as-number 100
```

Display the information about the routes of VPN1 on PE.

```
<PE-bgp-vpn1> display ip routing-table vpn-instance vpn1
Routing Tables: vpn1
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface
```

```
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
10.100.30.0/24      Direct 0    0            10.100.10.3     Vlan2
10.100.30.3/32      Direct 0    0            127.0.0.1       InLoop0
172.16.10.0/24      BGP    255  2            10.100.10.2     Vlan2
```

For VPN2, perform the configurations similar to the above on MCE and PE to import the OSPF routing information of VPN2 to the EBGP routing table. Configuration procedures are omitted here. Followed is the result of the above configurations.

```
<PE> display ip routing-table vpn-instance vpn2
Routing Tables: vpn2
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
10.100.40.0/24      Direct 0    0            10.100.20.3     Vlan3
10.100.40.3/32      Direct 0    0            127.0.0.1       InLoop0
172.16.20.0/24      BGP    255  2            10.100.20.2     Vlan3
```

After the above configurations, MCE has imported the OSPF routing information of VPN1 and VPN2 to the EBGP routing table of PE properly.

IP Multicast Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The IP Multicast Volume is organized as follows:

Features	Description
Multicast Overview	This document describes the main concepts in multicast: • Introduction to Multicast • Multicast Models • Multicast Architecture • Multicast Packets Forwarding Mechanism
Multicast Routing and Forwarding	Multicast routing and forwarding refer to some policies that filter RPF routing information for IP multicast support. This document describes: • Multicast routing and forwarding overview • Multicast routing and forwarding configuration
IGMP	Internet Group Management Protocol (IGMP) is a protocol in the TCP/IP suite responsible for management of IP multicast members. This document describes: • IGMP overview • Configuring basic functions of IGMP • Configuring IGMP performance parameters • Configuring IGMP SSM Mapping • Configuring IGMP Proxying
PIM	PIM leverages the unicast routing table created by any unicast routing protocol to provide routing information for IP multicast. This document describes: • Configuring PIM-DM • Configuring PIM-SM • Configuring PIM-SSM • Configuring PIM Common Features
MSDP	Multicast source discovery protocol (MSDP) describes interconnection mechanism of multiple PIM-SM domains. It is used is to discover multicast source information in other PIM-SM domains. This document describes: • MSDP configuration • Configuring an MSDP Peer Connection • Configuring SA Messages Related Parameters

Features	Description
MBGP	As a multicast extension of MP-BGP, MBGP enables BGP to provide routing information for multicast applications. This document describes: • Configuring MBGP Basic Functions • Configuring MBGP Route Attributes • Configuring a Large Scale MBGP Network
IGMP Snooping	Running at the data link layer, IGMP Snooping is a multicast control mechanism on the Layer 2 Ethernet switch and it is used for multicast group management and control. This document describes: • Configuring Basic Functions of IGMP Snooping • Configuring IGMP Snooping Port Functions • Configuring IGMP Snooping Querier • Configuring IGMP Snooping Policy
Multicast VLAN	Multicast VLAN configuration
IPv6 Multicast Routing and Forwarding	IPv6 multicast routing and forwarding refer to some policies that filter RPF routing information for IPv6 multicast support. This document describes: • IPv6 Multicast routing and forwarding overview • IPv6 Multicast routing and forwarding configuration
MLD	MLD is used by an IPv6 router or a Ethernet Switch to discover the presence of multicast listeners on directly-attached subnets. This document describes: • MLD overview • Configuring Basic Functions of MLD • Adjusting MLD Performance • Configuring MLD SSM Mapping • Configuring MLD Proxying
IPv6 PIM	IPv6 PIM discovers multicast source and delivers information to the receivers. This document describes: • Configuring IPv6 PIM-DM • Configuring IPv6 PIM-SM • Configuring IPv6 PIM-SSM • Configuring IPv6 PIM Common Features
IPv6 MBGP	As an IPv6 multicast extension of MP-BGP, IPv6 MBGP enables BGP to provide routing information for IPv6 multicast applications. This document describes: • Configuring IPv6 MBGP Basic Functions • Configuring IPv6 MBGP Route Attributes • Configuring a Large Scale IPv6 MBGP Network
MLD Snooping	Multicast Listener Discovery Snooping (MLD Snooping) is an IPv6 multicast constraining mechanism that runs on Layer 2 devices to manage and control IPv6 multicast groups. This document describes: • Configuring Basic Functions of MLD Snooping • Configuring MLD Snooping Port Functions • Configuring MLD Snooping Querier • Configuring MLD Snooping Policy
IPv6 Multicast VLAN	IPv6 Multicast VLAN configuration

Table of Contents

1 Multicast Overview

Note

This manual chiefly focuses on the IP multicast technology and device operations. Unless otherwise stated, the term “multicast” in this document refers to IP multicast.

Introduction to Multicast

As a technique coexisting with unicast and broadcast, the multicast technique effectively addresses the issue of point-to-multipoint data transmission. By allowing high-efficiency point-to-multipoint data transmission over a network, multicast greatly saves network bandwidth and reduces network load.

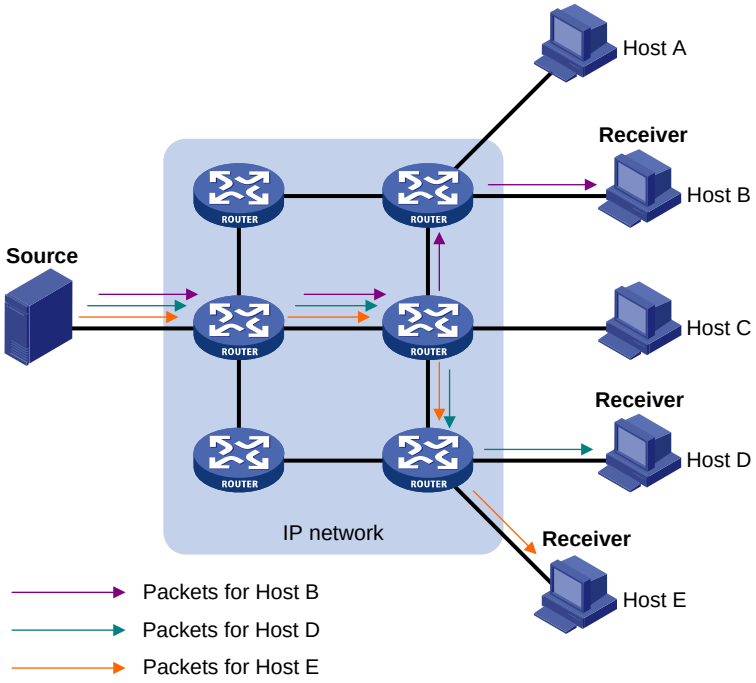
With the multicast technology, a network operator can easily provide new value-added services, such as live Webcasting, Web TV, distance learning, telemedicine, Web radio, real-time videoconferencing, and other bandwidth- and time-critical information services.

Comparison of Information Transmission Techniques

Unicast

In unicast, the information source (Source in the figure) needs to send a separate copy of information to each host (Receiver in the figure) that wants the information, as shown in Figure 1-1.

Figure 1-1 Unicast transmission



Assume that Host B, Host D and Host E need the information. A separate transmission channel needs to be established from the information source to each of these hosts.

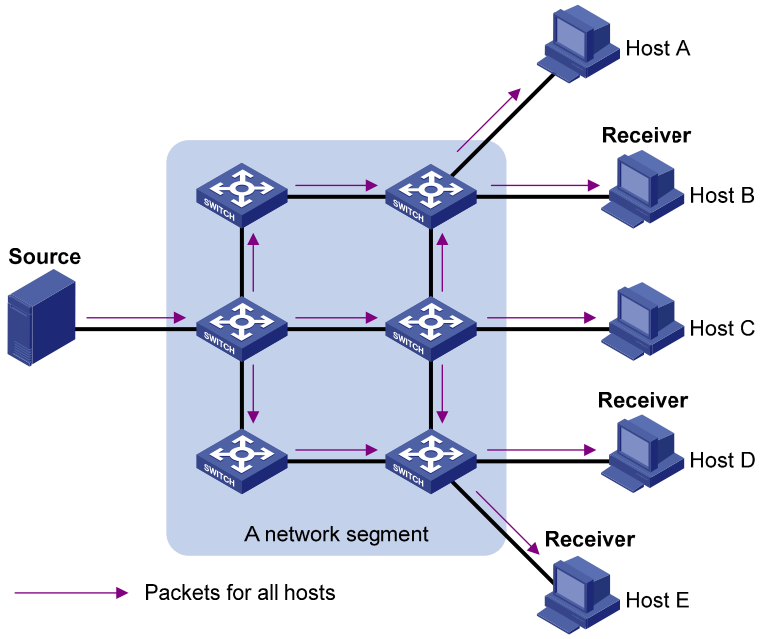
In unicast transmission, the traffic transmitted over the network is proportional to the number of hosts that need the information. If a large number of users need the information, the information source needs to send a copy of the same information to each of these users. This means a tremendous pressure on the information source and the network bandwidth.

As we can see from the information transmission process, unicast is not suitable for batch transmission of information.

Broadcast

In broadcast, the information source sends information to all hosts on the subnet, even if some hosts do not need the information, as shown in Figure 1-2.

Figure 1-2 Broadcast transmission



Assume that only Host B, Host D, and Host E need the information. If the information is broadcast to the subnet, Host A and Host C also receive it. In addition to information security issues, this also causes traffic flooding on the same subnet.

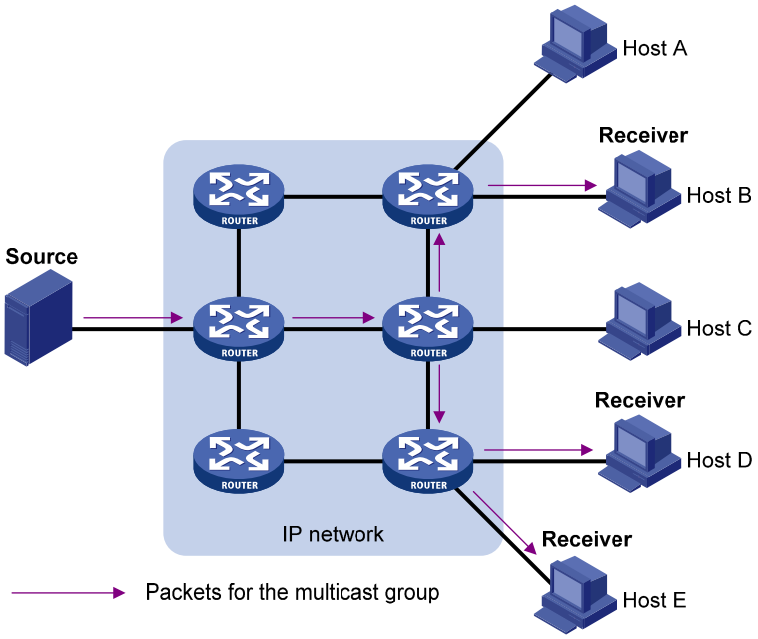
Therefore, broadcast is disadvantageous in transmitting data to specific hosts; moreover, broadcast transmission is a significant waste of network resources.

Multicast

As discussed above, unicast and broadcast techniques are unable to provide point-to-multipoint data transmissions with the minimum network consumption.

Multicast can well solve this problem. When some hosts on the network need multicast information, the information sender, or multicast source, sends only one copy of the information. Multicast distribution trees are built through multicast routing protocols, and the packets are replicated only on nodes where the trees branch. Figure 1-3 shows the delivery of a data stream to receiver hosts through multicast.

Figure 1-3 Multicast transmission



The multicast source (Source in the figure) sends only one copy of the information to a multicast group. Host B, Host D and Host E, which are receivers of the information, need to join the multicast group. The routers on the network duplicate and forward the information based on the distribution of the group members. Finally, the information is correctly delivered to Host B, Host D, and Host E.

To sum up, the advantages of multicast are summarized as follows:

- Over unicast: As multicast traffic flows to the node the farthest possible from the source before it is replicated and distributed, an increase of the number of hosts will not increase the load of the source and will not remarkably add to network resource usage.
- Over broadcast: As multicast data is sent only to the receivers that need it, multicast uses the network bandwidth reasonably and enhances network security. In addition, data broadcast is confined to the same subnet, while multicast is not.

Features of Multicast

Multicast has the following features:

- A multicast group is a multicast receiver set identified by an IP multicast address. Hosts join a multicast group to become members of the multicast group, before they can receive the multicast data addressed to that multicast group. Typically, a multicast source does not need to join a multicast group.
- An information sender is referred to as a multicast source (Source in Figure 1-3). A multicast source can send data to multiple multicast groups at the same time, and multiple multicast sources can send data to the same multicast group at the same time.
- All hosts that have joined a multicast group become members of the multicast group (Receiver in Figure 1-3). The group memberships are dynamic. Hosts can join or leave multicast groups at any time. Multicast groups are not subject to geographic restrictions.
- Routers or Layer 3 switches that support Layer 3 multicast are called multicast routers or Layer 3 multicast devices. In addition to providing the multicast routing function, a multicast router can also manage multicast group memberships on stub subnets with attached group members. A multicast router itself can be a multicast group member.

For a better understanding of the multicast concept, you can assimilate multicast transmission to the transmission of TV programs, as shown in Table 1-1.

Table 1-1 An analogy between TV transmission and multicast transmission

TV transmission	Multicast transmission
A TV station transmits a TV program through a channel.	A multicast source sends multicast data to a multicast group.
A user tunes the TV set to the channel.	A receiver joins the multicast group.
The user starts to watch the TV program transmitted by the TV station via the channel.	The receiver starts to receive the multicast data that the source is sending to the multicast group.
The user turns off the TV set or tunes to another channel.	The receiver leaves the multicast group or joins another group.

Common Notations in Multicast

Two notations are commonly used in multicast:

- (*, G): Indicates a rendezvous point tree (RPT), or a multicast packet that any multicast source sends to multicast group G. Here "*" represents any multicast source, while "G" represents a specific multicast group.
- (S, G): Indicates a shortest path tree (SPT), or a multicast packet that multicast source S sends to multicast group G. Here "S" represents a specific multicast source, while "G" represents a specific multicast group.

Note

For details about the concepts RPT and SPT, see *PIM Configuration* or *IPv6 PIM Configuration* in the *IP Multicast Volume*.

Advantages and Applications of Multicast

Advantages of multicast

Advantages of the multicast technique include:

- Enhanced efficiency: reduces the CPU load of information source servers and network devices.
- Optimal performance: reduces redundant traffic.
- Distributive application: enables point-to-multipoint applications at the price of minimum network resources.

Applications of multicast

Applications of the multicast technique include:

- Multimedia and streaming applications, such as Web TV, Web radio, and real-time video/audio conferencing.
- Communication for training and cooperative operations, such as distance learning and telemedicine.

- Data warehouse and financial applications (stock quotes).
- Any other point-to-multipoint data distribution application.

Multicast Models

Based on how the receivers treat the multicast sources, there are three multicast models: any-source multicast (ASM), source-filtered multicast (SFM), and source-specific multicast (SSM).

ASM model

In the ASM model, any sender can send information to a multicast group as a multicast source, and numbers of receivers can join a multicast group identified by a group address and obtain multicast information addressed to that multicast group. In this model, receivers are not aware of the position of multicast sources in advance. However, they can join or leave the multicast group at any time.

SFM model

The SFM model is derived from the ASM. From the view of a sender, the two models have the same multicast membership architecture.

The SFM model functionally extends the ASM model: In the SFM model, the upper layer software checks the source address of received multicast packets and permits or denies multicast traffic from specific sources. Therefore, receivers can receive the multicast data from only part of the multicast sources. From the view of a receiver, multicast sources are not all valid: they are filtered.

SSM model

In the practical life, users may be interested in the multicast data from only certain multicast sources. The SSM model provides a transmission service that allows users to specify the multicast sources they are interested in at the client side.

The radical difference between the SSM model and the ASM model is that in the SSM model, receivers already know the locations of the multicast sources by some other means. In addition, the SSM model uses a multicast address range that is different from that of the ASM/SFM model, and dedicated multicast forwarding paths are established between receivers and the specified multicast sources.

Multicast Architecture

IP multicast addresses the following questions:

- Where should the multicast source transmit information to? (multicast addressing)
- What receivers exist on the network? (host registration)
- Where is the multicast source the receivers need to receive multicast data from? (multicast source discovery)
- How should information be transmitted to the receivers? (multicast routing)

IP multicast falls in the scope of end-to-end service. The multicast architecture involves the following four parts:

1) Addressing mechanism: Information is sent from a multicast source to a group of receivers through a multicast address.
2) Host registration: Receiver hosts are allowed to join and leave multicast groups dynamically. This mechanism is the basis for group membership management.
3) Multicast routing: A multicast distribution tree (namely a forwarding path tree for multicast data on the network) is constructed for delivering multicast data from a multicast source to receivers.

4) Multicast applications: A software system that supports multicast applications, such as video conferencing, must be installed on multicast sources and receiver hosts, and the TCP/IP stack must support reception and transmission of multicast data.

Multicast Addresses

To allow communication between multicast sources and multicast group members, network-layer multicast addresses, namely, multicast IP addresses must be provided. In addition, a technique must be available to map multicast IP addresses to link-layer multicast MAC addresses.

IP multicast addresses

1) IPv4 multicast addresses

Internet Assigned Numbers Authority (IANA) assigned the Class D address space (224.0.0.0 to 239.255.255.255) for IPv4 multicast. The specific address blocks and usages are shown in Table 1-2.

Table 1-2 Class D IP address blocks and description

Address block	Description
224.0.0.0 to 224.0.0.255	Reserved permanent group addresses. The IP address 224.0.0.0 is reserved, and other IP addresses can be used by routing protocols and for topology searching, protocol maintenance, and so on. Common permanent group addresses are listed in Table 1-3. A packet destined for an address in this block will not be forwarded beyond the local subnet regardless of the Time to Live (TTL) value in the IP header.
224.0.1.0 to 238.255.255.255	Globally scoped group addresses. This block includes two types of designated group addresses: • 232.0.0.0/8: SSM group addresses, and • 233.0.0.0/8: Glop group addresses.
239.0.0.0 to 239.255.255.255	Administratively scoped multicast addresses. These addresses are considered to be locally rather than globally unique, and can be reused in domains administered by different organizations without causing conflicts. For details, refer to RFC 2365.

Note

- The membership of a group is dynamic. Hosts can join or leave multicast groups at any time.
- “Glop” is a mechanism for assigning multicast addresses between different autonomous systems (ASs). By filling an AS number into the middle two bytes of 233.0.0.0, you get 255 multicast addresses for that AS. For more information, refer to RFC 2770.

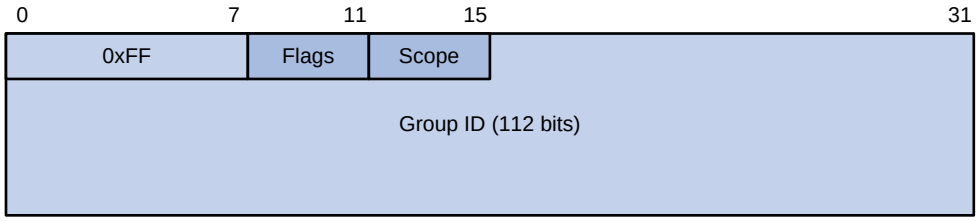
Table 1-3 Some reserved multicast addresses

Address	Description
224.0.0.1	All systems on this subnet, including hosts and routers
224.0.0.2	All multicast routers on this subnet
224.0.0.3	Unassigned
224.0.0.4	Distance Vector Multicast Routing Protocol (DVMRP) routers

Address	Description
224.0.0.5	Open Shortest Path First (OSPF) routers
224.0.0.6	OSPF designated routers/backup designated routers
224.0.0.7	Shared Tree (ST) routers
224.0.0.8	ST hosts
224.0.0.9	Routing Information Protocol version 2 (RIPv2) routers
224.0.0.11	Mobile agents
224.0.0.12	Dynamic Host Configuration Protocol (DHCP) server/relay agent
224.0.0.13	All Protocol Independent Multicast (PIM) routers
224.0.0.14	Resource Reservation Protocol (RSVP) encapsulation
224.0.0.15	All Core-Based Tree (CBT) routers
224.0.0.16	Designated Subnetwork Bandwidth Management (SBM)
224.0.0.17	All SBMs
224.0.0.18	Virtual Router Redundancy Protocol (VRRP)

2) IPv6 multicast addresses

Figure 1-4 IPv6 multicast format



Referring to Figure 1-4, the meanings of the fields of an IPv6 multicast address are as follows:

- 0xFF: The most significant 8 bits are 11111111, indicating that this address is an IPv6 multicast address.

Figure 1-5 Format of the Flags field

- Flags: Referring to Figure 1-5, the following table describes the four bits of the Flags field.

Table 1-4 Description on the bits of the Flags field

Bit	Description
0	Reserved, set to 0
R	• When set to 0, it indicates that this address is an IPv6 multicast address without an embedded RP address • When set to 1, it indicates that this address is an IPv6 multicast address with an embedded RP address (The P and T bits must also be set to 1)

Bit	Description
P	• When set to 0, it indicates that this address is an IPv6 multicast address not based on a unicast prefix • When set to 1, it indicates that this address is an IPv6 multicast address based on a unicast prefix (the T bit must also be set to 1)
T	• When set to 0, it indicates that this address is an IPv6 multicast address permanently-assigned by IANA • When set to 1, it indicates that this address is a transient, or dynamically assigned IPv6 multicast address

- Scope: 4 bits, indicating the scope of the IPv6 internetwork for which the multicast traffic is intended. Possible values of this field are given in Table 1-5.

Table 1-5 Values of the Scope field

Value	Meaning
0, 3, F	Reserved
1	Interface-local scope
2	Link-local scope
4	Admin-local scope
5	Site-local scope
6, 7, 9 through D	Unassigned
8	Organization-local scope
E	Global scope

- Group ID: 112 bits, IPv6 multicast group identifier that uniquely identifies an IPv6 multicast group in the scope defined by the Scope field.

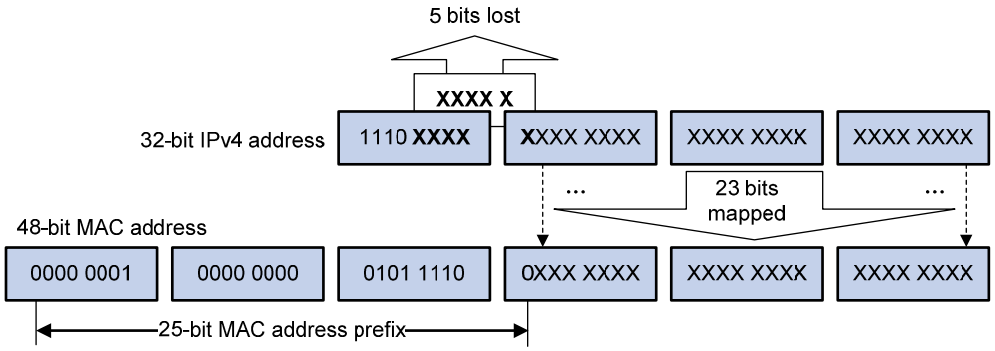
Ethernet multicast MAC addresses

When a unicast IP packet is transmitted over Ethernet, the destination MAC address is the MAC address of the receiver. When a multicast packet is transmitted over Ethernet, however, the destination address is a multicast MAC address because the packet is directed to a group formed by a number of receivers, rather than to one specific receiver.

1) IPv4 multicast MAC addresses

As defined by IANA, the high-order 24 bits of an IPv4 multicast MAC address are 0x01005E, bit 25 is 0, and the low-order 23 bits are the low-order 23 bits of a multicast IPv4 address. The IPv4-to-MAC mapping relation is shown in Figure 1-6.

Figure 1-6 IPv4-to-MAC address mapping

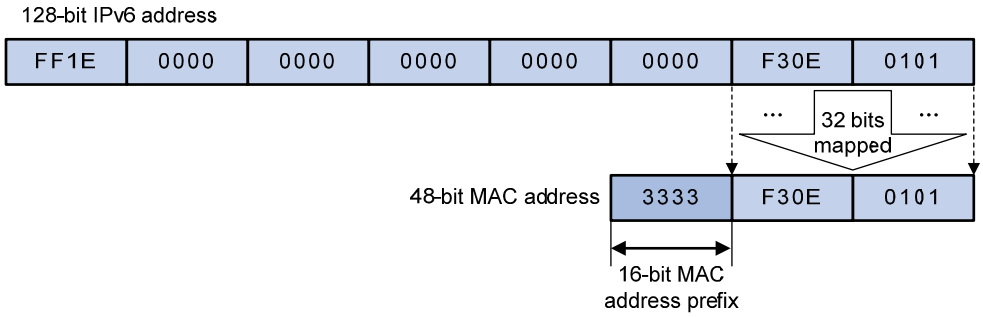


The high-order four bits of a multicast IPv4 address are 1110, indicating that this address is a multicast address, and only 23 bits of the remaining 28 bits are mapped to a MAC address, so five bits of the multicast IPv4 address are lost. As a result, 32 multicast IPv4 addresses map to the same MAC address. Therefore, in Layer 2 multicast forwarding, a device may receive some multicast data addressed for other IPv4 multicast groups, and such redundant data needs to be filtered by the upper layer.

2) IPv6 multicast MAC addresses

The high-order 16 bits of an IPv6 multicast MAC address are 0x3333, and the low-order 32 bits are the low-order 32 bits of a multicast IPv6 address. Figure 1-7 shows an example of mapping an IPv6 multicast address, FF1E::F30E:101, to a MAC address.

Figure 1-7 An example of IPv6-to-MAC address mapping



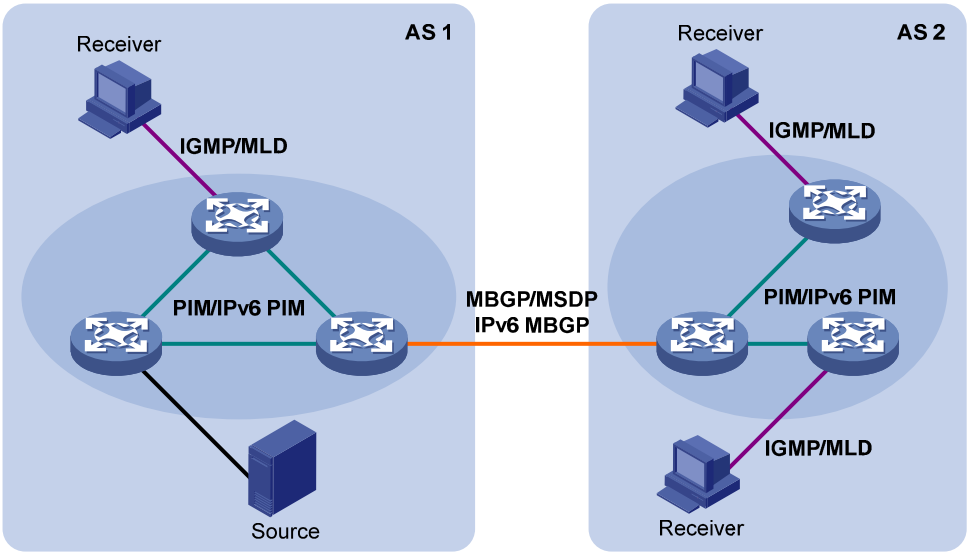
Note

- Generally, we refer to IP multicast working at the network layer as Layer 3 multicast and the corresponding multicast protocols as Layer 3 multicast protocols, which include IGMP/MLD, PIM/IPv6 PIM, MSDP, and MBGP/IPv6 MBGP; we refer to IP multicast working at the data link layer as Layer 2 multicast and the corresponding multicast protocols as Layer 2 multicast protocols, which include IGMP Snooping/MLD Snooping, and multicast VLAN/IPv6 multicast VLAN.
- IGMP Snooping, IGMP, multicast VLAN, PIM, MSDP, and MBGP are for IPv4, MLD Snooping, MLD, IPv6 multicast VLAN, IPv6 PIM, and IPv6 MBGP are for IPv6.
- This section provides only general descriptions about applications and functions of the Layer 2 and Layer 3 multicast protocols in a network. For details of these protocols, refer to the related configuration manuals in the *IP Multicast Volume*.

Layer 3 multicast protocols

Layer 3 multicast protocols include multicast group management protocols and multicast routing protocols. Figure 1-8 describes where these multicast protocols are in a network.

Figure 1-8 Positions of Layer 3 multicast protocols



2) Multicast management protocols

Typically, the internet group management protocol (IGMP) or multicast listener discovery protocol (MLD) is used between hosts and Layer 3 multicast devices directly connected with the hosts. These protocols define the mechanism of establishing and maintaining group memberships between hosts and Layer 3 multicast devices.

3) Multicast routing protocols

A multicast routing protocol runs on Layer 3 multicast devices to establish and maintain multicast routes and forward multicast packets correctly and efficiently. Multicast routes constitute a loop-free data transmission path from a data source to multiple receivers, namely, a multicast distribution tree.

In the ASM model, multicast routes come in intra-domain routes and inter-domain routes.

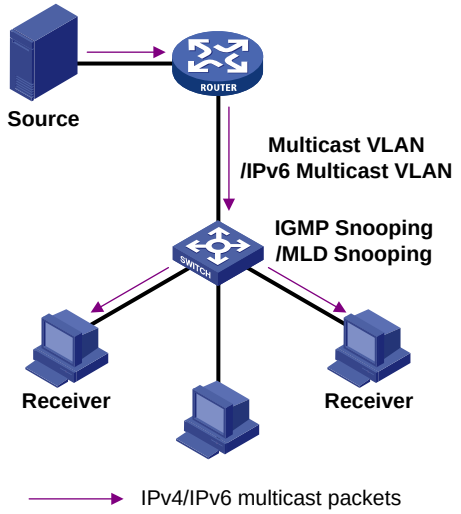
- An intra-domain multicast routing protocol is used to discover multicast sources and build multicast distribution trees within an AS so as to deliver multicast data to receivers. Among a variety of mature intra-domain multicast routing protocols, protocol independent multicast (PIM) is a popular one. Based on the forwarding mechanism, PIM comes in two modes – dense mode (often referred to as PIM-DM) and sparse mode (often referred to as PIM-SM).
- An inter-domain multicast routing protocol is used for delivery of multicast information between two ASs. So far, mature solutions include multicast source discovery protocol (MSDP) and multicast border gateway protocol (MBGP). MSDP is used to propagate multicast source information among different ASs, while MBGP, an extension of the Multi-protocol Border Gateway Protocol (MP-BGP), is used for exchanging multicast routing information among different ASs.

For the SSM model, multicast routes are not divided into inter-domain routes and intra-domain routes. Since receivers know the position of the multicast source, channels established through PIM-SM are sufficient for multicast information transport.

Layer 2 multicast protocols

Layer 2 multicast protocols include IGMP Snooping/MLD Snooping and multicast VLAN/IPv6 multicast VLAN. Figure 1-9 shows where these protocols are in the network.

Figure 1-9 Position of Layer 2 multicast protocols



1) IGMP Snooping/MLD Snooping

Running on Layer 2 devices, Internet Group Management Protocol Snooping (IGMP Snooping) and Multicast Listener Discovery Snooping (MLD Snooping) are multicast constraining mechanisms that manage and control multicast groups by listening to and analyzing IGMP or MLD messages exchanged between the hosts and Layer 3 multicast devices, thus effectively controlling the flooding of multicast data in a Layer 2 network.

2) Multicast VLAN/IPv6 multicast VLAN

In the traditional multicast-on-demand mode, when users in different VLANs on a Layer 2 device need multicast information, the upstream Layer 3 device needs to forward a separate copy of the multicast

data to each VLAN of the Layer 2 device. With the multicast VLAN or IPv6 multicast VLAN feature enabled on the Layer 2 device, the Layer 3 multicast device needs to send only one copy of multicast to the multicast VLAN or IPv6 multicast VLAN on the Layer 2 device. This avoids waste of network bandwidth and extra burden on the Layer 3 device.

Multicast Packet Forwarding Mechanism

In a multicast model, a multicast source sends information to the host group identified by the multicast group address in the destination address field of IP multicast packets. Therefore, to deliver multicast packets to receivers located in different parts of the network, multicast routers on the forwarding path usually need to forward multicast packets received on one incoming interface to multiple outgoing interfaces. Compared with a unicast model, a multicast model is more complex in the following aspects.

- To ensure multicast packet transmission in the network, unicast routing tables or multicast routing tables (for example, the MBGP routing table) specially provided for multicast must be used as guidance for multicast forwarding.
- To process the same multicast information from different peers received on different interfaces of the same device, every multicast packet is subject to a reverse path forwarding (RPF) check on the incoming interface. The result of the RPF check determines whether the packet will be forwarded or discarded. The RPF check mechanism is the basis for most multicast routing protocols to implement multicast forwarding.

Note

For details about the RPF mechanism, refer to *Multicast Routing and Forwarding Configuration* or *IPv6 Multicast Routing and Forwarding Configuration* in the *IP Multicast Volume*.

Table of Contents

1 Multicast Routing and Forwarding Configuration

When configuring multicast routing and forwarding, go to these sections for information you are interested in:

Note

The term "router" in this document refers to a router in a generic sense or a Layer 3 switch running an IP routing protocol.

Multicast Routing and Forwarding Overview

Introduction to Multicast Routing and Forwarding

In multicast implementations, multicast routing and forwarding are implemented by three types of tables:

- Each multicast routing protocol has its own multicast routing table, such as PIM routing table.
- The information of different multicast routing protocols forms a general multicast routing table.
- The multicast forwarding table is directly used to control the forwarding of multicast packets.

A multicast forwarding table consists of a set of (S, G) entries, each indicating the routing information for delivering multicast data from a multicast source to a multicast group. If a router supports multiple multicast protocols, its multicast routing table will include routes generated by multiple protocols. The router chooses the optimal route from the multicast routing table based on the configured multicast routing and forwarding policy and adds the route entry into its multicast forwarding table.

RPF Check Mechanism

A multicast routing protocol relies on the existing unicast routing information, MBGP routes, or multicast static routes in creating multicast routing entries. When creating multicast routing table entries, a multicast routing protocol uses the reverse path forwarding (RPF) check mechanism to ensure multicast data delivery along the correct path. In addition, the RPF check mechanism also helps avoid data loops caused by various reasons.

RPF check process

The basis for an RPF check is a unicast route, an MBGP route, or a multicast static route.

- A unicast routing table contains the shortest path to each destination subnet,
- An MBGP routing table contains multicast routing information, and
- A multicast static routing table contains the RPF routing information defined by the user through static configuration.

When performing an RPF check, a router searches its unicast routing table and multicast static routing table at the same time. The specific process is as follows:

1) The router first chooses an optimal route from the unicast routing table, MBGP routing table, and multicast static routing table:

- The router automatically chooses an optimal unicast route by searching its unicast routing table, using the IP address of the “packet source” as the destination address. The outgoing interface in the corresponding routing entry is the RPF interface and the next hop is the RPF neighbor. The router considers the path along which the packet from the RPF neighbor arrived on the RPF interface to be the shortest path that leads back to the source.
- The router automatically chooses an optimal MBGP route by searching its MBGP routing table, using the IP address of the “packet source” as the destination address. The outgoing interface in the corresponding routing entry is the RPF interface and the next hop is the RPF neighbor.
- The router automatically chooses an optimal multicast static route by searching its multicast static routing table, using the IP address of the “packet source” as the destination address. The corresponding routing entry explicitly defines the RPF interface and the RPF neighbor.

2) Then, the router selects one from these three optimal routes as the RPF route. The selection process is as follows:

- If configured to use the longest match principle, the router selects the longest match route from the three; if these three routes have the same mask, the router selects the route with the highest priority; if the three routes have the same priority, the router selects the RPF route according to the sequence of multicast static route, MBGP route, and unicast route.
- If not configured to use the longest match principle, the router selects the route with the highest priority; if the three routes have the same priority, the router selects the RPF route according to the sequence of multicast static route, MBGP route, and unicast route.

Note

The above-mentioned “packet source” can mean different things in different situations:

- For a packet traveling along the shortest path tree (SPT) from the multicast source to the receivers or the rendezvous point (RP), the “packet source” for RPF check is the multicast source.
- For a packet traveling along the rendezvous point tree (RPT) from the RP to the receivers, the “packet source” for RPF check is the RP.
- For a bootstrap message from the bootstrap router (BSR), the “packet source” for RPF check is the BSR.

For details about the concepts of SPT, RPT and BSR, refer to *PIM Configuration* in the *IP Multicast Volume*.

Implementation of RPF check in multicast

Implementing an RPF check on each received multicast data packet would bring a big burden to the router. The use of a multicast forwarding table is the solution to this issue. When creating a multicast

routing entry and a multicast forwarding entry for a multicast packet, the router sets the RPF interface of the packet as the incoming interface of the (S, G) entry. Upon receiving an (S, G) multicast packet, the router first searches its multicast forwarding table:

1) If the corresponding (S, G) entry does not exist in the multicast forwarding table, the packet is subject to an RPF check. The router creates a multicast routing entry based on the relevant routing information and adds the entry into the multicast forwarding table, with the RPF interface as the incoming interface.

- If the interface on which the packet actually arrived is the RPF interface, the RPF check succeeds and the router forwards the packet to all the outgoing interfaces.
- If the interface on which the packet actually arrived is not the RPF interface, the RPF check fails and the router discards the packet.

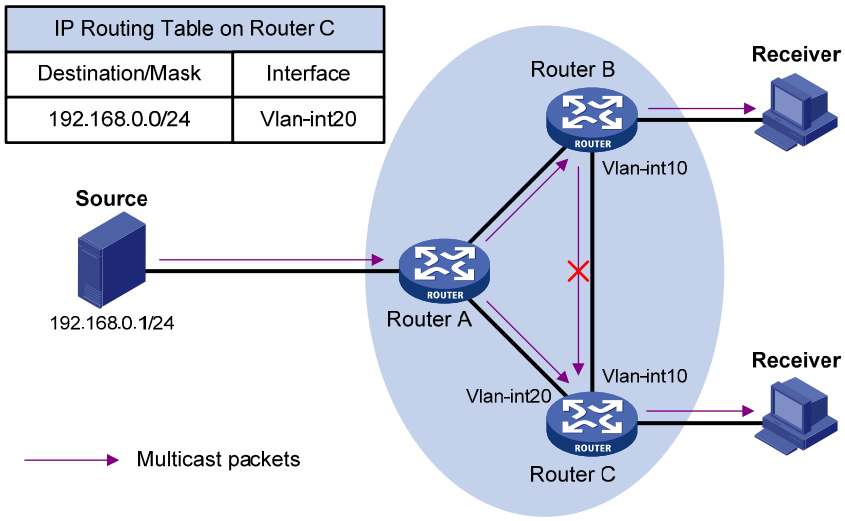
2) If the corresponding (S, G) entry exists, and the interface on which the packet actually arrived is the incoming interface, the router forwards the packet to all the outgoing interfaces.

3) If the corresponding (S, G) entry exists, but the interface on which the packet actually arrived is not the incoming interface in the multicast forwarding table, the multicast packet is subject to an RPF check.

- If the RPF interface is the incoming interface of the (S, G) entry, this means the (S, G) entry is correct but the packet arrived from a wrong path. The packet is to be discarded.
- If the RPF interface is not the incoming interface, this means the (S, G) entry has expired, and router replaces the incoming interface with the RPF interface. If the interface on which the packet arrived in the RPF interface, the router forwards the packet to all the outgoing interfaces; otherwise it discards the packet.

Assume that unicast routes are available in the network, MBGP is not configured, and no multicast static routes have been configured on Router C, as shown in Figure 1-1. Multicast packets travel along the SPT from the multicast source to the receivers. The multicast forwarding table on Router C contains the (S, G) entry, with Vlan-interface 20 as the RPF interface.

Figure 1-1 RPF check process



- When a multicast packet arrives on Vlan-interface 20 of Router C, as the interface is the incoming interface of the (S, G) entry, the router forwards the packet to all outgoing interfaces.
- When a multicast packet arrives on Vlan-interface 10 of Router C, as the interface is not the incoming interface of the (S, G) entry, the router performs an RPF check on the packet: The router searches its unicast routing table and finds that the outgoing interface to Source (the RPF interface)

is Vlan-interface 20. This means the (S, G) entry is correct and packet arrived along a wrong path. The RPF check fails and the packet is discarded.

Multicast Static Routes

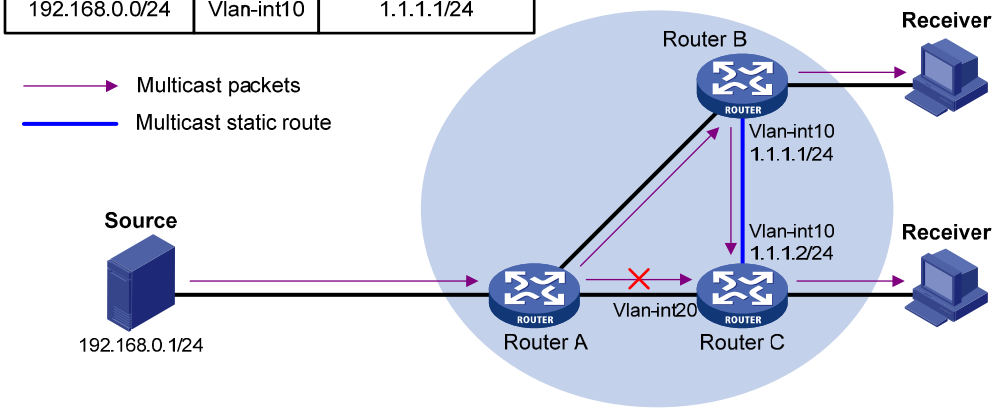
A multicast static route is an important basis for RPF check. Depending on the application environment, a multicast static route has the following two functions:

Changing an RPF route

Typically, the topology structure of a multicast network is the same as that of a unicast network, and multicast traffic follows the same transmission path as unicast traffic does. By configuring a multicast static route for a given multicast source, you can change the RPF route so as to create a transmission path for multicast traffic different from that for unicast traffic.

Figure 1-2 Changing an RPF route

Multicast Routing Table Static on Router C		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Vlan-int10	1.1.1.1/24



As shown in Figure 1-2, when no multicast static route is configured, Router C's RPF neighbor on the path back to Source is Router A and the multicast information from Source travels along the path from Router A to Router C, which is the unicast route between the two routers; with a static route configured on Router C and with Router B as Router C's RPF neighbor on the path back to Source, the multicast information from Source travels from Router A to Router B and then to Router C.

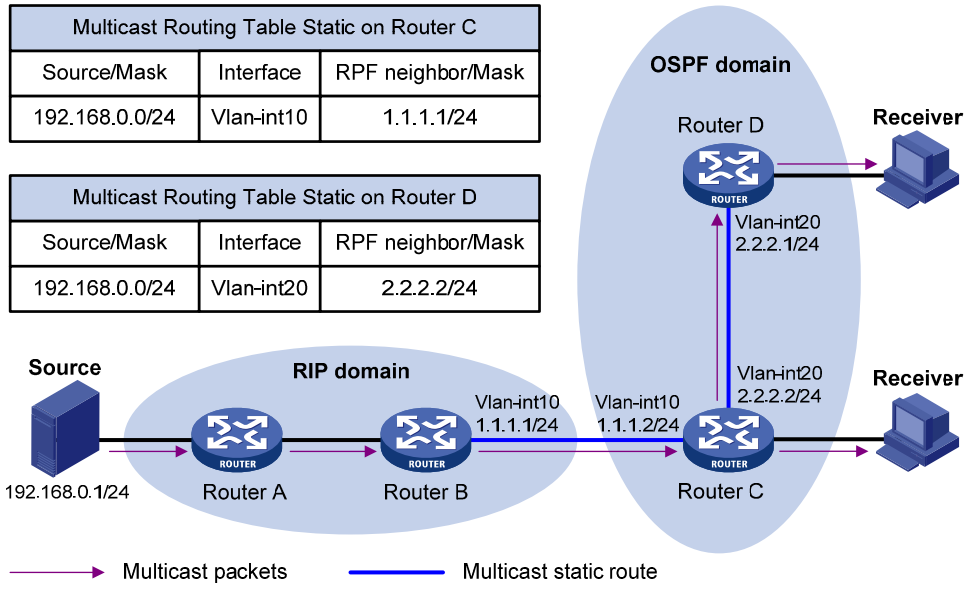
Creating an RPF route

When a unicast route is blocked, multicast traffic forwarding is stopped due to lack of an RPF route. By configuring a multicast static route for a given multicast source, you can create an RPF route so that a multicast routing entry is created to guide multicast traffic forwarding.

Figure 1-3 Creating an RPF route

Multicast Routing Table Static on Router C		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Vlan-int10	1.1.1.1/24

Multicast Routing Table Static on Router D		
Source/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	Vlan-int20	2.2.2.2/24



As shown in Figure 1-3, the RIP domain and the OSPF domain are unicast isolated from each other. When no multicast static route is configured, the hosts (Receivers) in the OSPF domain cannot receive the multicast packets sent by the multicast source (Source) in the RIP domain. After you configure a multicast static route on Router C and Router D, specifying Router B as the RPF neighbor of Router C and specifying Router C as the RPF neighbor of Router D, the receivers can receive multicast data sent by the multicast source.

Note

- A multicast static route only affects RPF check; it cannot guide multicast forwarding. Therefore, a multicast static route is also called an RPF static route.
- A multicast static route is effective only on the multicast router on which it is configured, and will not be advertised throughout the network or redistributed to other routers.

Multicast Traceroute

The multicast traceroute utility is used to trace the path that a multicast stream flows down from the first-hop router to the last-hop router.

Concepts in multicast traceroute

1) Last-hop router: If a router has one of its interfaces connecting to the subnet the given destination address is on, and if the router is able to forward multicast streams from the given multicast source onto that subnet, that router is called last-hop router.
2) First-hop router: the router that directly connects to the multicast source.
3) Querier: the router requesting the multicast traceroute.

Introduction to multicast traceroute packets

A multicast traceroute packet is a special IGMP packet, which differs from common IGMP packets in that its IGMP Type field is set to 0x1F or 0x1E and that its destination IP address is a unicast address. There are three types of multicast traceroute packets:

- Query, with the IGMP Type field set to 0x1F,
- Request, with the IGMP Type field set to 0x1F, and
- Response, with the IGMP Type field set to 0x1E.

Process of multicast traceroute

1) The querier sends a query to the last-hop router.
2) Upon receiving the query, the last-hop router turns the query packet into a request packet by adding a response data block containing its interface addresses and packet statistics to the end of the packet, and forwards the request packet via unicast to the previous hop for the given multicast source and group.
3) From the last-hop router to the multicast source, each hop adds a response data block to the end of the request packet and unicasts it to the previous hop.
4) When the first-hop router receives the request packet, it changes the packet type to indicate a response packet, and then sends the completed packet via unicast to the multicast traceroute querier.

Configuration Task List

Complete these tasks to configure multicast routing and forwarding:

Task		Remarks
Enabling IP Multicast Routing		Required
Configuring Multicast Routing and Forwarding	Configuring Multicast Static Routes	Optional
	Configuring a Multicast Routing Policy	Optional
	Configuring a Multicast Forwarding Range	Optional
	Configuring the Multicast Forwarding Table Size	Optional
	Tracing a Multicast Path	Optional

Caution

IP multicast does not support the use of secondary IP address segments. Namely, multicast can be routed and forwarded only through primary IP addresses, rather than secondary addresses, even if configured on interfaces.

For details about primary and secondary IP addresses, refer to *IP Addressing Configuration* in the *IP Services Volume*.

Enabling IP Multicast Routing

Before configuring any Layer 3 multicast functionality, you must enable IP multicast routing.

Enabling IP multicast routing in the public instance

Follow these steps to enable IP multicast routing in the public instance:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disabled by default

Configuring Multicast Routing and Forwarding

Configuration Prerequisites

Before configuring multicast routing and forwarding, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Enable PIM (PIM-DM or PIM-SM).

Before configuring multicast routing and forwarding, prepare the following data:

- The minimum TTL value required for a multicast packet to be forwarded
- The maximum number of downstream nodes for a single multicast forwarding table entry
- The maximum number of entries in the multicast forwarding table

Configuring Multicast Static Routes

By configuring a multicast static route for a given multicast source, you can specify an RPF interface or an RPF neighbor for multicast traffic from that source.

Follow these steps to configure a multicast static route:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure a multicast static route	**ip rpf-route-static** *source-address* { *mask* | *mask-length* } [*protocol* [*process-id*]] [**route-policy** *policy-name*] { *rpf-nbr-address* | *interface-type interface-number* } [**preference** *preference*] [**order** *order-number*]	Required No multicast static route configured by default.

Caution

When configuring a multicast static route, you cannot specify an RPF neighbor by its interface type and number (*interface-type interface-number*) if the interface type of the RPF neighbor is Ethernet, Layer 3 aggregate, Loopback, RPR, or VLAN-interface; instead, you can such an RPF neighbor only by its address (*rpf-nbr-address*).

Configuring a Multicast Routing Policy

You can configure the router to determine the RPF route based on the longest match principle. For details about RPF route selection, refer to RPF check process.

By configuring per-source or per-source-and-group load splitting, you can optimize the traffic delivery when multiple data flows are handled.

Configuring a multicast routing policy in the public instance

Follow these steps to configure a multicast routing policy in the public instance:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the device to select the RPF route based on the longest match	**multicast longest-match**	Required The route with the highest priority is selected as the RPF route by default
Configure multicast load splitting	**multicast load-splitting** { **source** | **source-group** }	Optional Disabled by default

Configuring a Multicast Forwarding Range

Multicast packets do not travel without a boundary in a network. The multicast data corresponding to each multicast group must be transmitted within a definite scope. Presently, you can define a multicast forwarding range by:

- Specifying boundary interfaces, which form a closed multicast forwarding area, or
- Setting the minimum time to live (TTL) value required for a multicast packet to be forwarded.

Note

Setting the minimum TTL is not supported on 3Com Switch 4800G.

You can configure a forwarding boundary specific to a particular multicast group on all interfaces that support multicast forwarding. A multicast forwarding boundary sets the boundary condition for the multicast groups in the specified range. If the destination address of a multicast packet matches the set boundary condition, the packet will not be forwarded. Once a multicast boundary is configured on an interface, this interface can no longer forward multicast packets (including packets sent from the local device) or receive multicast packets.

Follow these steps to configure a multicast forwarding range:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do...	Use the command...	Remarks
Configure a multicast forwarding boundary	**multicast boundary** *group-address* { *mask* | *mask-length* }	Required No forwarding boundary by default

Configuring the Multicast Forwarding Table Size

The router maintains the corresponding forwarding entry for each multicast packet it receives. Excessive multicast routing entries, however, can exhaust the router's memory and thus result in lower router performance. You can set a limit on the number of entries in the multicast forwarding table based on the actual networking situation and the performance requirements. If the configured maximum number of multicast forwarding table entries is smaller than the current value, the forwarding entries in excess will not be immediately deleted; instead they will be deleted by the multicast routing protocol running on the router. The router will no longer add new multicast forwarding entries until the number of existing multicast forwarding entries comes down below the configured value.

When forwarding multicast traffic, the router replicates a copy of the multicast traffic for each downstream node and forwards the traffic, and thus each of these downstream nodes forms a branch of the multicast distribution tree. You can configure the maximum number of downstream nodes (namely, the maximum number of outgoing interfaces) for a single entry in the multicast forwarding table to lessen burden on the router for replicating multicast traffic. If the configured maximum number of downstream nodes for a single multicast forwarding entry is smaller than the current number, the downstream nodes in excess will not be deleted immediately; instead they must be deleted by the multicast routing protocol. The router will no longer add new multicast forwarding entries for newly added downstream nodes until the number of existing downstream nodes comes down below the configured value.

Follow these steps to configure the multicast forwarding table size:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the maximum number of entries in the multicast forwarding table	**multicast forwarding-table route-limit** *limit*	Optional 1024 by default.
Configure the maximum number of downstream nodes for a single multicast forwarding entry	**multicast forwarding-table downstream-limit** *limit*	Optional 128 by default.

Tracing a Multicast Path

You can run the **mtracert** command to trace the path down which the multicast traffic flows from a given first-hop router to the last-hop router.

To do…	Use the command…	Remarks
Trace a multicast path	**mtracert** *source-address* [[*last-hop-router-address*] *group-address*]	Required Available in any view

Displaying and Maintaining Multicast Routing and Forwarding

To do...	Use the command...	Remarks
View the multicast boundary information	**display multicast boundary** [*group-address* [*mask* | *mask-length*]] [**interface** *interface-type interface-number*]	Available in any view
View the multicast forwarding table information	**display multicast forwarding-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics** | **slot** *slot-number*] * [**port-info**]	Available in any view
View the multicast routing table information	**display multicast routing-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } }] *	Available in any view
View the information of the multicast static routing table	**display multicast routing-table static** [**config**] [*source-address* { *mask-length* | *mask* }]	Available in any view
View the RPF route information of the specified multicast source	**display multicast rpf-info** *source-address* [*group-address*]	Available in any view
Clear forwarding entries from the multicast forwarding table	**reset multicast forwarding-table** { { *source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }	Available in user view
Clear routing entries from the multicast routing table	**reset multicast routing-table** { { *source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }	Available in user view

Caution

- The reset command clears the information in the multicast routing table or the multicast forwarding table, and thus may cause failure of multicast transmission.
- When a routing entry is deleted from the multicast routing table, the corresponding forwarding entry will also be deleted from the multicast forwarding table.
- When a forwarding entry is deleted from the multicast forwarding table, the corresponding route entry will also be deleted from the multicast routing table.

Configuration Examples

Changing an RPF Route

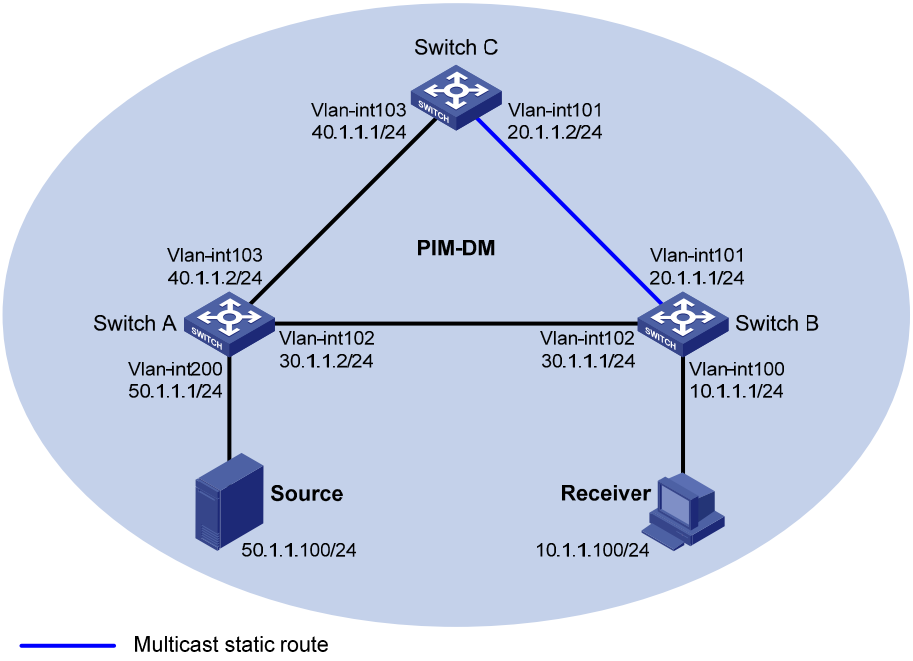
Network requirements

- PIM-DM runs in the network. All switches in the network support multicast.

- Switch A, Switch B and Switch C run OSPF.
- Typically, Receiver can receive the multicast data from Source through the path Switch A – Switch B, which is the same as the unicast route.
- Perform the following configuration so that Receiver can receive the multicast data from Source through the path Switch A – Switch C – Switch B, which is different from the unicast route.

Network diagram

Figure 1-4 Network diagram for RPF route alternation configuration



Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-4. The detailed configuration steps are omitted here.

Enable OSPF on the switches in the PIM-DM domain. Ensure the network-layer interoperation among the switches in the PIM-DM domain. Ensure that the switches can dynamically update their routing information by leveraging the unicast routing protocol. The specific configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-DM and IGMP

Enable IP multicast routing on Switch B, enable PIM-DM on each interface, and enable IGMP on the host-side interface VLAN-interface 100.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] pim dm
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim dm
[SwitchB-Vlan-interface101] quit
```

```
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim dm
[SwitchB-Vlan-interface102] quit
```

Enable IP multicast routing on Switch A, and enable PIM-DM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] pim dm
[SwitchA-Vlan-interface200] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim dm
[SwitchA-Vlan-interface102] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
[SwitchA-Vlan-interface103] quit
```

The configuration on Switch C is similar to the configuration on Switch A. The specific configuration steps are omitted here.

Use the **display multicast rpf-info** command to view the RPF route to Source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: igp
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

As shown above, the current RPF route on Switch B is contributed by a unicast routing protocol and the RPF neighbor is Switch A.

3) Configure a multicast static route

Configure a multicast static route on Switch B, specifying Switch C as its RPF neighbor on the route to Source.

```
[SwitchB] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

4) Verify the configuration

Use the **display multicast rpf-info** command to view the information about the RPF route to Source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

As shown above, the RPF route on Switch B has changed. It is now the configured multicast static route, and the RPF neighbor is now Switch C.

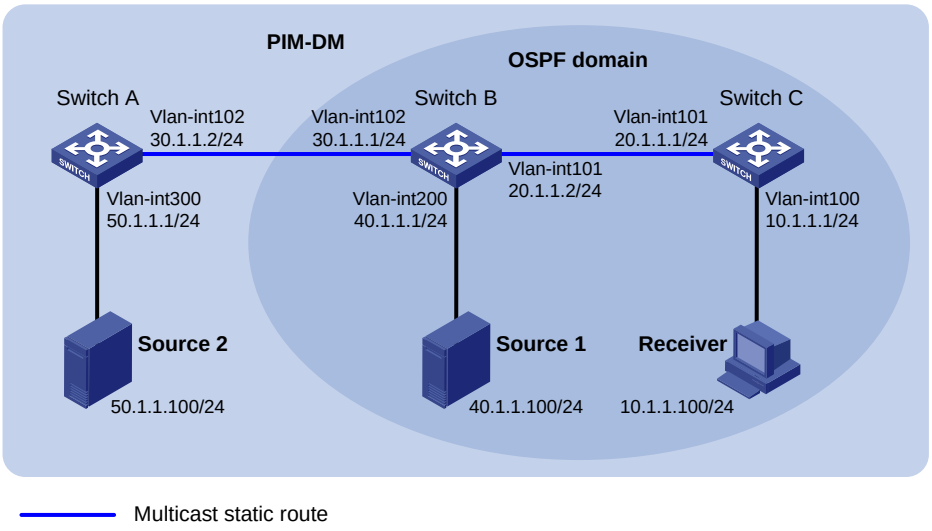
Creating an RPF Route

Network requirements

- PIM-DM runs in the network and all switches in the network support IP multicast.
- Switch B and Switch C run OSPF, and have no unicast routes to Switch A.
- Typically, Receiver can receive the multicast data from Source 1 in the OSPF domain.
- Perform the following configuration so that Receiver can receive multicast data from Source 2, which is outside the OSPF domain.

Network diagram

Figure 1-5 Network diagram for creating an RPF route



Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-5. The detailed configuration steps are omitted here.

Enable OSPF on Switch B and Switch C. Ensure the network-layer interoperation among Switch B and Switch C. Ensure that the switches can dynamically update their routing information by leveraging the unicast routing protocol. The specific configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-DM and IGMP

Enable IP multicast routing on Switch C, enable PIM-DM on each interface, and enable IGMP on the host-side interface VLAN-interface 100.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] igmp enable
[SwitchC-Vlan-interface100] pim dm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim dm
[SwitchC-Vlan-interface101] quit
```

Enable IP multicast routing on Switch A and enable PIM-DM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] pim dm
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim dm
[SwitchC-Vlan-interface102] quit
```

The configuration on Switch B is similar to that on Switch A. The specific configuration steps are omitted here.

Use the **display multicast rpf-info** command to view the RPF routes to Source 2 on Switch B and Switch C.

```
[SwitchB] display multicast rpf-info 50.1.1.100
[SwitchC] display multicast rpf-info 50.1.1.100
```

No information is displayed. This means that no RPF route to Source 2 exists on Switch B and Switch C.

3) Configure a multicast static route

Configure a multicast static route on Switch B, specifying Switch A as its RPF neighbor on the route to Source 2.

```
[SwitchB] ip rpf-route-static 50.1.1.100 24 30.1.1.2
```

Configure a multicast static route on Switch C, specifying Switch B as its RPF neighbor on the route to Source 2.

```
[SwitchC] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

4) Verify the configuration

Use the **display multicast rpf-info** command to view the RPF routes to Source 2 on Switch B and Switch C.

```
[SwitchB] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
[SwitchC] display multicast rpf-info 50.1.1.100
 RPF information about source 50.1.1.100:
     RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
     Referenced route/mask: 50.1.1.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

As shown above, the RPF routes to Source 2 exist on Switch B and Switch C. The source is the configured static route.

Troubleshooting Multicast Routing and Forwarding

Multicast Static Route Failure

Symptom

No dynamic routing protocol is enabled on the routers, and the physic status and link layer status of interfaces are both up, but the multicast static route fails.

Analysis

- If the multicast static route is not configured or updated correctly to match the current network conditions, the route entry and the configuration information of multicast static routes do not exist in the multicast routing table.
- If the optimal route is found, the multicast static route may also fail.

Solution

1) In the configuration, you can use the **display multicast routing-table static config** command to view the detailed configuration information of multicast static routes to verify that the multicast static route has been correctly configured and the route entry exists.
2) In the configuration, you can use the **display multicast routing-table static** command to view the information of multicast static routes to verify that the multicast static route has been correctly configured and the route entry exists in the multicast routing table.
3) Check the next hop interface type of the multicast static route. If the interface is not a point-to-point interface, be sure to specify the next hop address to configure the outgoing interface when you configure the multicast static route.
4) Check that the multicast static route matches the specified routing protocol. If a protocol was specified in multicast static route configuration, enter the **display ip routing-table** command to check if an identical route was added by the protocol.
5) Check that the multicast static route matches the specified routing policy. If a routing policy was specified when the multicast static route was configured, enter the **display route-policy** command to check the configured routing policy.

Multicast Data Fails to Reach Receivers

Symptom

The multicast data can reach some routers but fails to reach the last hop router.

Analysis

If a multicast forwarding boundary has been configured through the **multicast boundary** command, any multicast packet will be kept from crossing the boundary.

Solution

1) Use the **display pim routing-table** command to check whether the corresponding (S, G) entries exist on the router. If so, the router has received the multicast data; otherwise, the router has not received the data.
2) Use the **display multicast boundary** command to view the multicast boundary information on the interfaces. Use the **multicast boundary** command to change the multicast forwarding boundary setting.

3) In the case of PIM-SM, use the **display current-configuration** command to check the BSR and RP information.

Table of Contents

1 IGMP Configuration

- When configuring IGMP, go to the following sections for the information you are interested in:
- IGMP Overview
- IGMP Configuration Task List
- IGMP Configuration Examples
- Troubleshooting IGMP

Note

The term "router" in this document refers to a router in a generic sense or a Layer 3 switch running an IP routing protocol.

IGMP Overview

As a TCP/IP protocol responsible for IP multicast group member management, the Internet Group Management Protocol (IGMP) is used by IP hosts to establish and maintain their multicast group memberships to immediately neighboring multicast routers.

IGMP Versions

So far, there are three IGMP versions:

- IGMPv1 (documented in RFC 1112)
- IGMPv2 (documented in RFC 2236)
- IGMPv3 (documented in RFC 3376)

All IGMP versions support the Any-Source Multicast (ASM) model. In addition to support of the ASM model, IGMPv3 can be directly deployed to implement the Source-Specific Multicast (SSM) model, while IGMPv1 and IGMPv2 need to work with the IGMP SSM mapping function to implement the SSM model.

Note

For more information about the ASM and SSM models, see *Multicast Overview* of the *IP Multicast Volume*.

Introduction to IGMPv1

IGMPv1 manages multicast group memberships mainly based on the query and response mechanism.

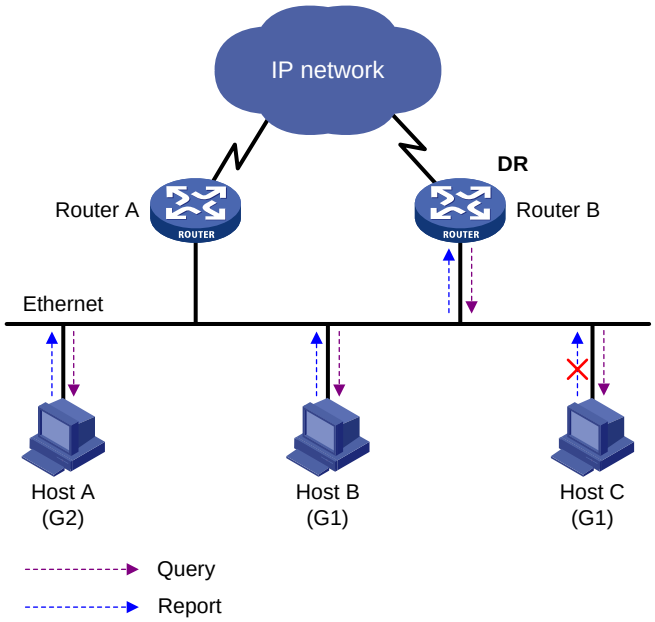
Of multiple multicast routers on the same subnet, all the routers can hear IGMP membership report messages (often referred to as reports) from hosts, but only one router is needed for sending IGMP query messages (often referred to as queries). So, a querier election mechanism is required to determine which router will act as the IGMP querier on the subnet.

In IGMPv1, the designated router (DR) elected by the working multicast routing protocol (such as PIM) serves as the IGMP querier.

Note

For more information about DR, refer to *PIM Configuration* in the *IP Multicast Volume*.

Figure 1-1 IGMP queries and reports



Assume that Host B and Host C are expected to receive multicast data addressed to multicast group G1, while Host A is expected to receive multicast data addressed to G2, as shown in Figure 1-1. The following describes how the hosts join the multicast groups and the IGMP querier (Router B in the figure) maintains the multicast group memberships:

1) The hosts send unsolicited IGMP reports to the addresses of the multicast groups that they want to join, without having to wait for the IGMP queries from the IGMP querier.
2) The IGMP querier periodically multicasts IGMP queries (with the destination address of 224.0.0.1) to all hosts and routers on the local subnet.
3) Upon receiving a query message, Host B or Host C (the delay timer of whichever expires first) sends an IGMP report to the multicast group address of G1, to announce its membership for G1. Assume it is Host B that sends the report message. Upon hearing the report from Host B, Host C, which is on the same subnet with Host B, suppresses its own report for G1, because the IGMP routers (Router A and Router B) already know that at least one host on the local subnet is interested in G1. This mechanism, known as IGMP report suppression, helps reduce traffic on the local subnet.

4) At the same time, because Host A is interested in G2, it sends a report to the multicast group address of G2.
5) Through the above-mentioned query/report process, the IGMP routers learn that members of G1 and G2 are attached to the local subnet, and the multicast routing protocol (PIM for example) running on the routers generates (*, G1) and (*, G2) multicast forwarding entries, which will be the basis for subsequent multicast forwarding, where * represents any multicast source.
6) When the multicast data addressed to G1 or G2 reaches an IGMP router, because the (*, G1) and (*, G2) multicast forwarding entries exist on the IGMP router, the router forwards the multicast data to the local subnet, and then the receivers on the subnet receive the data.

As IGMPv1 does not specifically define a Leave Group message, upon leaving a multicast group, an IGMPv1 host stops sending reports to the address of the multicast group it listened to. If no member of a multicast group exists on the subnet, the IGMP router will not receive any report addressed to that multicast group, so the routers will delete the multicast forwarding entries for that multicast group after a period of time.

Enhancements in IGMPv2

Compared with IGMPv1, IGMPv2 has introduced a querier election mechanism and a leave-group mechanism.

Querier election mechanism

In IGMPv1, the DR elected by the Layer 3 multicast routing protocol (such as PIM) serves as the querier among multiple routers on the same subnet.

In IGMPv2, an independent querier election mechanism is introduced. The querier election process is as follows:

1) Initially, every IGMPv2 router assumes itself as the querier and sends IGMP general query messages (often referred to as general queries) to all hosts and routers on the local subnet (the destination address is 224.0.0.1).
2) Upon hearing a general query, every IGMPv2 router compares the source IP address of the query message with its own interface address. After comparison, the router with the lowest IP address wins the querier election and all other IGMPv2 routers become non-queriers.
3) All the non-queriers start a timer, known as “other querier present timer”. If a router receives an IGMP query from the querier before the timer expires, it resets this timer; otherwise, it assumes the querier to have timed out and initiates a new querier election process.

“Leave group” mechanism

In IGMPv1, when a host leaves a multicast group, it does not send any notification to the multicast router. The multicast router relies on host response timeout to know whether a group no longer has members. This adds to the leave latency.

In IGMPv2, on the other hand, when a host leaves a multicast group:

1) This host sends a Leave Group message (often referred to as leave message) to all routers (the destination address is 224.0.0.2) on the local subnet.
2) Upon receiving the leave message, the querier sends a configurable number of group-specific queries to the group being left. The destination address field and group address field of the message are both filled with the address of the multicast group being queried.
3) One of the remaining members, if any on the subnet, of the group being queried should send a membership report within the maximum response time set in the query messages.

4) If the querier receives a membership report for the group within the maximum response time, it will maintain the memberships of the group; otherwise, the querier will assume that no hosts on the subnet are still interested in multicast traffic to that group and will stop maintaining the memberships of the group.

Enhancements in IGMPv3

Built upon and being compatible with IGMPv1 and IGMPv2, IGMPv3 provides hosts with enhanced control capabilities and provides enhancements of query and report messages.

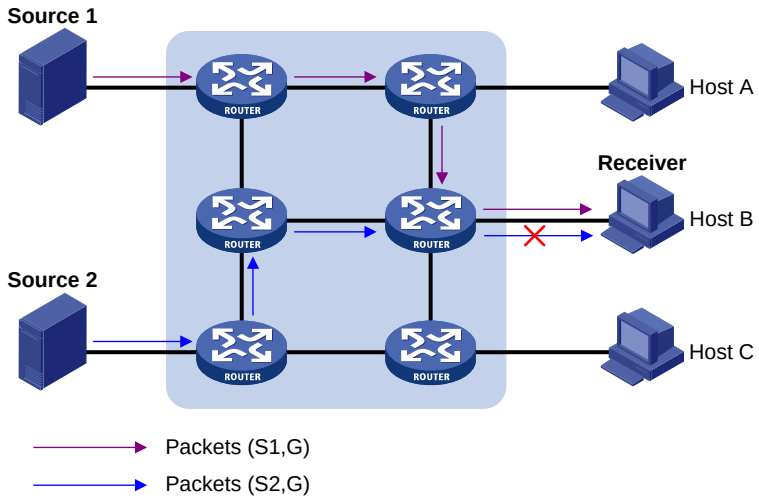
Enhancements in control capability of hosts

IGMPv3 has introduced source filtering modes (Include and Exclude), so that a host can specify a list of sources it expect or does not expect multicast data from when it joins a multicast group:

- If it expects multicast data from specific sources like S1, S2, …, it sends a report with the Filter-Mode denoted as “Include Sources (S1, S2, …).
- If it does not expect multicast data from specific sources like S1, S2, …, it sends a report with the Filter-Mode denoted as “Exclude Sources (S1, S2, …).

As shown in Figure 1-2, the network comprises two multicast sources, Source 1 (S1) and Source 2 (S2), both of which can send multicast data to multicast group G. Host B is interested only in the multicast data that Source 1 sends to G but not in the data from Source 2.

Figure 1-2 Flow paths of source-and-group-specific multicast traffic



In the case of IGMPv1 or IGMPv2, Host B cannot select multicast sources when it joins multicast group G. Therefore, multicast streams from both Source 1 and Source 2 will flow to Host B whether it needs them or not.

When IGMPv3 is running between the hosts and routers, Host B can explicitly express its interest in the multicast data Source 1 sends to multicast group G (denoted as (S1, G)), rather than the multicast data Source 2 sends to multicast group G (denoted as (S2, G)). Thus, only multicast data from Source 1 will be delivered to Host B.

Enhancements in query and report capabilities

1) Query message carrying the source addresses

IGMPv3 supports not only general queries (feature of IGMPv1) and group-specific queries (feature of IGMPv2), but also group-and-source-specific queries.

- A general query does not carry a group address, nor a source address;
- A group-specific query carries a group address, but no source address;
- A group-and-source-specific query carries a group address and one or more source addresses.

2) Reports containing multiple group records

Unlike an IGMPv1 or IGMPv2 report message, an IGMPv3 report message is destined to 224.0.0.22 and contains one or more group records. Each group record contains a multicast group address and a multicast source address list.

Group record types include:

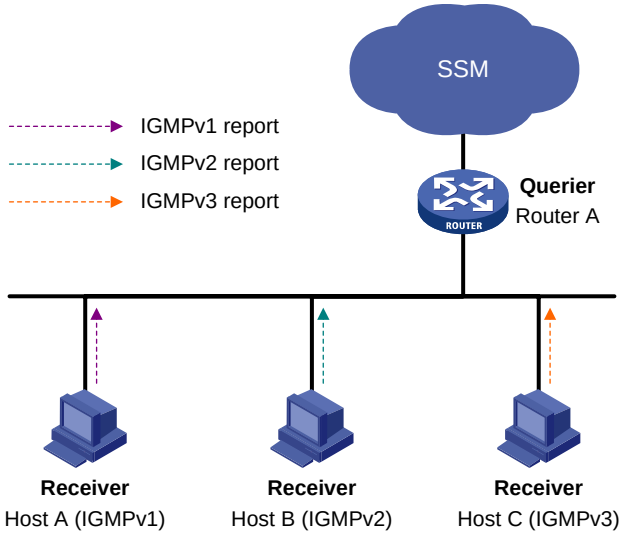
- IS_IN: The source filtering mode is Include, namely, the report sender requests the multicast data from only the sources defined in the specified multicast source list.
- IS_EX: The source filtering mode is Exclude, namely, the report sender requests the multicast data from any sources but those defined in the specified multicast source list.
- TO_IN: The filtering mode has changed from Exclude to Include.
- TO_EX: The filtering mode has changed from Include to Exclude.
- ALLOW: The Source Address fields in this Group Record contain a list of the additional sources that the system wishes to hear from, for packets sent to the specified multicast address. If the change was to an Include source list, these are the addresses that were added to the list; if the change was to an Exclude source list, these are the addresses that were deleted from the list.
- BLOCK: indicates that the Source Address fields in this Group Record contain a list of the sources that the system no longer wishes to hear from, for packets sent to the specified multicast address. If the change was to an Include source list, these are the addresses that were deleted from the list; if the change was to an Exclude source list, these are the addresses that were added to the list.

IGMP SSM Mapping

The IGMP SSM mapping feature allows you to configure static IGMP SSM mappings on the last hop router to provide SSM support for receiver hosts running IGMPv1 or IGMPv2. The SSM model assumes that the last hop router is aware of the desired multicast sources when receivers join multicast groups.

- When a host running IGMPv3 joins a multicast group, it can explicitly specify one or more multicast sources in its IGMPv3 report.
- A host running IGMPv1 or IGMPv2, however, cannot specify multicast source addresses in its report. In this case, you need to configure the IGMP SSM mapping feature to translate the (*, G) information in the IGMPv1 or IGMPv2 report into (G, INCLUDE, (S1, S2...)) information.

Figure 1-3 Network diagram for IGMP SSM mapping



As shown in Figure 1-3, on an SSM network, Host A, Host B and Host C are running IGMPv1, IGMPv2 and IGMPv3 respectively. To provide SSM service for all the hosts while it is infeasible to run IGMPv3 on Host A and Host B, you need to configure the IGMP SSM mapping feature on Router A.

With the IGMP SSM mapping feature configured, when Router A receives an IGMPv1 or IGMPv2 report, it checks the multicast group address G carried in the message:

- If G is not in the SSM group range, Router A cannot provide the SSM service but the ASM service.
- If G is in the SSM group range but no IGMP SSM mappings corresponding to the multicast group G have been configured on Router A, Router A cannot provide SSM service and drops the message.
- If G is in the SSM group range and the IGMP SSM mappings have been configured on Router A for multicast group G, Router A translates the (*, G) information in the IGMP report into (G, INCLUDE, (S1, S2...)) information based on the configured IGMP SSM mappings and provides SSM service accordingly.

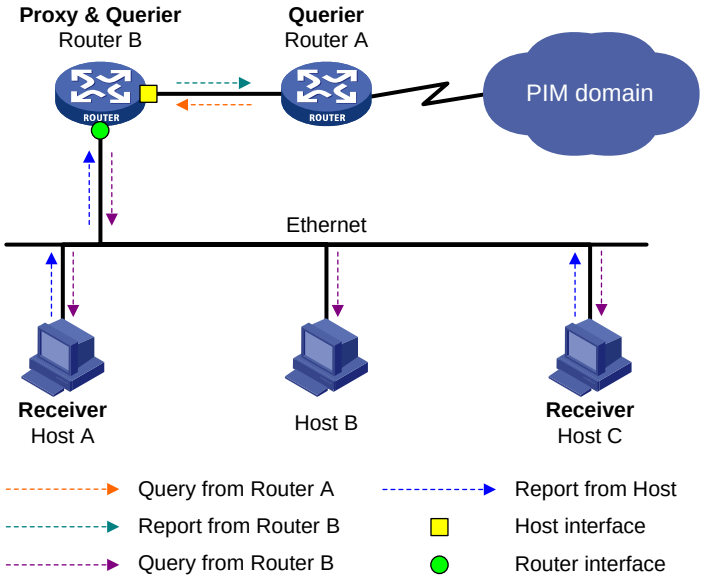
Note

- The IGMP SSM mapping feature does not process IGMPv3 reports.
- For more information about the SSM group range, refer to *PIM Configuration* in the *IP Multicast Volume*.

IGMP Proxying

In some simple tree-shaped topologies, it is not necessary to configure complex multicast routing protocols, such as PIM, on the boundary device. Instead, you can configure IGMP proxying on the boundary device. With IGMP proxying configured, the device serves as a proxy for the downstream hosts to send IGMP messages, maintain group memberships, and implement multicast forwarding based on the memberships. In this case, the boundary device is a host but no longer a PIM neighbor to the upstream device.

Figure 1-4 Network diagram for IGMP proxying



As shown in Figure 1-4, two types of interfaces are defined on an IGMP proxy device:

- Upstream interface: Also referred to as the proxy interface. A proxy interface is an interface on which IGMP proxying is configured. It is in the direction toward the root of the multicast forwarding tree. An upstream interface acts as a host running IGMP; therefore, it is also called host interface.
- Downstream interface: An interface that is running IGMP and not in the direction toward the root of the multicast forwarding tree. A downstream interface acts as a router running IGMP; therefore, it is also called router interface.

A device with IGMP proxying configured maintains a group membership database, which storesthe group memberships on all the downstream interfaces. Each entry comprises the multicast address, filter mode, and source list. Such an entry is a collection of members in the same multicast group on each downstream interface.

A proxy device performs host functions on the upstream interface based on the database. It responds to queries according to the information in the database or sends join/leave messages when the database changes. On the other hand, the proxy device performs router functions on the downstream interfaces by participating in the querier election, sending queries, and maintaining memberships based on the reports.

Protocols and Standards

The following documents describe different IGMP versions:

- RFC 1112: Host Extensions for IP Multicasting
- RFC 2236: Internet Group Management Protocol, Version 2
- RFC 3376: Internet Group Management Protocol, Version 3
- RFC 4605: Internet Group Management Protocol (IGMP)/Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")

IGMP Configuration Task List

Complete these tasks to configure IGMP:

Task		Remarks
Configuring Basic Functions of IGMP	Enabling IGMP	Required
	Configuring IGMP Versions	Optional
	Configuring Static Joining	Optional
	Configuring a Multicast Group Filter	Optional
	Configuring the Maximum Number of Multicast Groups on an Interface	Optional
Adjusting IGMP Performance	Configuring IGMP Message Options	Optional
	Configuring IGMP Query and Response Parameters	Optional
	Configuring IGMP Fast Leave Processing	Optional
Configuring IGMP SSM Mapping	Enabling SSM Mapping	Optional
	Configuring SSM Mappings	Optional
Configuring IGMP Proxying	Enabling IGMP Proxying	Optional
	Configuring Multicast Forwarding on a Downstream Interface	Optional

Note

- Configurations performed in IGMP view are effective on all interfaces, while configurations performed in interface view are effective on the current interface only.
- If a feature is not configured for an interface in interface view, the global configuration performed in IGMP view will apply to that interface. If a feature is configured in both IGMP view and interface view, the configuration performed in interface view will be given priority.

Configuring Basic Functions of IGMP

Configuration Prerequisites

Before configuring the basic functions of IGMP, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure PIM-DM or PIM-SM

Before configuring the basic functions of IGMP, prepare the following data:

- IGMP version
- Multicast group and multicast source addresses for static group member configuration
- ACL rule for multicast group filtering
- The maximum number of multicast groups that can be joined on an interface

Enabling IGMP

First, IGMP must be enabled on the interface on which the multicast group memberships are to be established and maintained.

Enabling IGMP

Follow these steps to enable IGMP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disabled by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable IGMP	**igmp enable**	Required Disabled by default

Note

For details about the **multicast routing-table** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring IGMP Versions

Because the protocol packets of different IGMP versions vary in structure and type, the same IGMP version should be configured for all routers on the same subnet before IGMP can work properly.

Configuring an IGMP version globally

Follow these steps to configure an IGMP version globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP view	**igmp**	—
Configure an IGMP version globally	**version** *version-number*	Optional IGMPv2 by default

Configuring an IGMP version on an interface

Follow these steps to configure an IGMP version on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do...	Use the command...	Remarks
Configure an IGMP version on the interface	**igmp version** *version-number*	Optional IGMPv2 by default

Configuring Static Joining

After an interface is configured as a static member of a multicast group or a multicast source and group, it will act as a virtual member of the multicast group to receive multicast data addressed to that multicast group for the purpose of testing multicast data forwarding.

Follow these steps to configure an interface as a statically connected member of a multicast group or a multicast source and group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the interface as a static member of a multicast group or a multicast source and group	**igmp static-group** *group-address* [**source** *source-address*]	Required An interface is not a static member of any multicast group or multicast source and group by default.

Note

- Before you can configure an interface of a PIM-SM device as a static member of a multicast group or a multicast source and group, if the interface is PIM-SM enabled, it must be a PIM-SM DR; if this interface is IGMP enabled but not PIM-SM enabled, it must be an IGMP querier. For more information about PIM-SM and a DR, refer to *PIM Configuration* in the *IP Multicast Volume*.
- As a static member of a multicast group or a multicast source and group, the interface does not respond to the queries from the IGMP querier, nor does it send an unsolicited IGMP membership report or an IGMP leave group message when it joins or leaves a multicast group or a multicast source and group. In other words, the interface will not become a real member of the multicast group or the multicast source and group.

Configuring a Multicast Group Filter

To restrict the hosts on the network attached to an interface from joining certain multicast groups, you can set an ACL rule on the interface as a packet filter that limits the range of multicast groups the interface serves.

Follow these steps to configure a multicast group filter:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a multicast group filter	**igmp group-policy** *acl-number* [*version-number*]	Required No multicast group filter configured by default

Configuring the Maximum Number of Multicast Groups on an Interface

You can configure the allowed maximum number of multicast groups on an interface to flexibly control the number of multicast groups the interface can join.

Follow these steps to configure the maximum number of multicast groups an interface can join:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the maximum number of multicast groups that can be joined on the current interface	**igmp group-limit** *limit*	Required 1024 by default

Adjusting IGMP Performance

Note

For the configuration tasks described in this section:

- Configurations performed in IGMP view are effective on all interfaces, while configurations performed in interface view are effective on the current interface only.
- If the same feature is configured in both IGMP view and interface view, the configuration performed in interface view is given priority, regardless of the configuration sequence.

Configuration Prerequisites

Before adjusting IGMP performance, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure basic functions of IGMP

Before adjusting IGMP performance, prepare the following data:

- Startup query interval

- Startup query count
- IGMP general query interval
- IGMP querier's robustness variable
- Maximum response time for IGMP general queries
- IGMP last-member query interval
- Other querier present interval

Configuring IGMP Message Options

IGMP queries include group-specific queries and group-and-source-specific queries, and multicast groups change dynamically, so a device cannot maintain the information for all multicast sources and groups, For this reason, when receiving a multicast packet but unable to locate the outgoing interface for the destination multicast group, an IGMP router needs to leverage the Router-Alert option to pass the multicast packet to the upper-layer protocol for processing. For details about the Router-Alert option, refer to RFC 2113.

An IGMP message is processed differently depending on whether it carries the Router-Alert option in the IP header:

- By default, for the consideration of compatibility, the device does not check the Router-Alert option, namely it processes all the IGMP messages it received. In this case, IGMP messages are directly passed to the upper layer protocol, no matter whether the IGMP messages carry the Router-Alert option or not.
- To enhance the device performance and avoid unnecessary costs, and also for the consideration of protocol security, you can configure the device to discard IGMP messages that do not carry the Router-Alert option.

Configuring IGMP packet options globally

Follow these steps to configure IGMP packet options globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP view	**igmp**	—
Configure the router to discard any IGMP message that does not carry the Router-Alert option	**require-router-alert**	Optional By default, the device does not check the Router-Alert option.
Enable insertion of the Router-Alert option into IGMP messages	**send-router-alert**	Optional By default, IGMP messages carry the Router-Alert option.

Configuring IGMP packet options on an interface

Follow these steps to configure IGMP packet options on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do...	Use the command...	Remarks
Configure the interface to discard any IGMP message that does not carry the Router-Alert option	**igmp require-router-alert**	Optional By default, the device does not check the Router-Alert option.
Enable insertion of the Router-Alert option into IGMP messages	**igmp send-router-alert**	Optional By default, IGMP messages carry the Router-Alert option.

Configuring IGMP Query and Response Parameters

On startup, the IGMP querier sends “startup query count” IGMP general queries at the “startup query interval”, which is 1/4 of the “IGMP query interval”.

The IGMP querier periodically sends IGMP general queries at the “IGMP query interval” to determine whether any multicast group member exists on the network. You can tune the IGMP general query interval based on actual condition of the network.

Upon receiving an IGMP leave message, the IGMP querier sends “last member query count” IGMP group-specific queries at the “IGMP last member query interval”. IGMP is robust to “robustness variable minus 1” packet losses on a network. Therefore, a greater value of the robustness variable makes the IGMP querier “more robust”, but results in a longer multicast group timeout time.

Upon receiving an IGMP query (general query or group-specific query), a host starts a delay timer for each multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time, which is derived from the Max Response Time field in the IGMP query. When the timer value comes down to 0, the host sends an IGMP report to the corresponding multicast group.

An appropriate setting of the maximum response time for IGMP queries allows hosts to respond to queries quickly and avoids bursts of IGMP traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expires simultaneously.

- For IGMP general queries, you can configure the maximum response time to fill their Max Response time field.
- For IGMP group-specific queries, you can configure the IGMP last member query interval to fill their Max Response time field. Namely, for IGMP group-specific queries, the maximum response time equals the IGMP last member query interval.

When multiple multicast routers exist on the same subnet, the IGMP querier is responsible for sending IGMP queries. If a non-querier router receives no IGMP query from the querier within the “other querier present interval”, it will assume the querier to have expired and a new querier election process is launched; otherwise, the non-querier router will reset its “other querier present timer”.

Configuring IGMP query and response parameters globally

Follow these steps to configure IGMP query and response parameters globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP view	**igmp**	—

To do...	Use the command...	Remarks
Configure the startup query interval	**startup-query-interval** *interval*	Optional For the system default, see “Note” below.
Configure the startup query count	**startup-query-count** *value*	Optional For the system default, see “Note” below.
Configure the IGMP query interval	**timer query** *interval*	Optional 60 seconds by default
Configure the IGMP querier robustness variable	**robust-count** *robust-value*	Optional 2 by default
Configure the maximum response time for IGMP general queries	**max-response-time** *interval*	Optional 10 seconds by default
Configure the IGMP last member query interval	**last-member-query-interval** *interval*	Optional 1 second by default
Configure the other querier present interval	**timer other-querier-present** *interval*	Optional For the system default, see “Note” below.

Configuring IGMP query and response parameters on an interface

Follow these steps to configure IGMP query and response parameters on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the startup query interval	**igmp startup-query-interval** *interval*	Optional For the system default, see “Note” below.
Configure the startup query count	**igmp startup-query-count** *value*	Optional For the system default, see “Note” below.
Configure the IGMP query interval	**igmp timer query** *interval*	Optional 60 seconds by default
Configure the IGMP querier robustness variable	**igmp robust-count** *robust-value*	Optional 2 by default
Configure the maximum response time for IGMP general queries	**igmp max-response-time** *interval*	Optional 10 seconds by default
Configure the IGMP last member query interval	**igmp last-member-query-interval** *interval*	Optional 1 second by default

To do...	Use the command...	Remarks
Configure the other querier present interval	**igmp timer other-querier-present** *interval*	Optional For the system default, see “Note” below.

Note

- If not statically configured, the startup query interval is 1/4 of the “IGMP query interval”. By default, the IGMP query interval is 60 seconds, so the startup query interval = 60 / 4 = 15 (seconds).
- If not statically configured, the startup query count is set to the IGMP querier robustness variable. By default, the IGMP querier robustness variable is 2, so the startup query count is also 2.
- If not statically configured, the other querier present interval is [IGMP query interval] times [IGMP robustness variable] plus [maximum response time for IGMP general queries] divided by two. The default values of these three parameters are 60 (seconds), 2 and 10 (seconds) respectively, so the other querier present interval = 60 × 2 + 10 / 2 = 125 (seconds).
- If statically configured, the startup query interval, the startup query count, and the other querier present interval take the configured values.

Caution

- Make sure that the other querier present interval is greater than the IGMP query interval; otherwise the IGMP querier may change frequently on the network.
- Make sure that the IGMP query interval is greater than the maximum response time for IGMP general queries; otherwise, multicast group members may be wrongly removed.
- The configurations of the maximum response time for IGMP general queries, the IGMP last member query interval and the IGMP other querier present interval are effective only for IGMPv2 or IGMPv3.

Configuring IGMP Fast Leave Processing

In some applications, such as ADSL dial-up networking, only one multicast receiver host is attached to a port of the IGMP querier. To allow fast response to the leave messages of the host when it switches frequently from one multicast group to another, you can enable IGMP fast leave processing on the IGMP querier.

With fast leave processing enabled, after receiving an IGMP leave message from a host, the IGMP querier directly sends a leave notification to the upstream without sending IGMP group-specific queries. Thus, the leave latency is reduced on one hand, and the network bandwidth is saved on the other hand.

Follow these steps to configure IGMP fast leave processing

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP view	**igmp**	—

To do...	Use the command...	Remarks
Configure IGMP fast leave processing	**fast-leave** [**group-policy** *acl-number*]	Required Disabled by default

Caution

The IGMP fast leave processing configuration is effective only if the device is running IGMPv2 or IGMPv3.

Configuring IGMP SSM Mapping

Due to some possible restrictions, some receiver hosts on an SSM network may run IGMPv1 or IGMPv2. To provide SSM service support for these receiver hosts, you need to configure the IGMP mapping feature on the last hop router.

Configuration Prerequisites

Before configuring the IGMP SSM mapping feature, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure basic functions of IGMP.

Enabling SSM Mapping

Follow these steps to enable the IGMP SSM mapping feature:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the IGMP SSM mapping feature	**igmp ssm-mapping enable**	Required Disabled by default

Note

To ensure SSM service for all hosts on a subnet, regardless of the IGMP version running on the hosts, enable IGMPv3 on the interface that forwards multicast traffic onto the subnet.

Configuring SSM Mappings

By performing this configuration multiple times, you can map a multicast group to different multicast sources.

Follow these steps to configure an IGMP SSM mapping:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IGMP view	**igmp**	—
Configure an IGMP SSM mapping	**ssm-mapping** *group-address* { *mask* | *mask-length* } *source-address*	Required No IGMP mappings are configured by default.

Caution

If IGMPv3 is enabled on a VLAN interface of a switch that supports both IGMP Snooping and IGMP, and if a port in that VLAN is configured as a simulated host, the simulated host will send IGMPv3 reports even if you did not specify a multicast source when configuring simulated joining with the **igmp-snooping host-join** command. In this case, the corresponding multicast group will not be created based on the configured IGMP SSM mappings. For details about the **igmp-snooping host-join** command, refer to *IGMP Snooping Commands* in the *IP Multicast Volume*.

Configuring IGMP Proxying

Configuration Prerequisites

Before configuring the IGMP proxying feature, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Enable IP multicast routing.

Enabling IGMP Proxying

You can enable IGMP proxying on the interface in the direction toward the root of the multicast forwarding tree to make the device serve as an IGMP proxy.

Follow these steps to enable IGMP proxying:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the IGMP proxying feature	**igmp proxying enable**	Required Disabled by default.

Note

- Each device can have only one interface serving as the proxy interface. In scenarios with multiple instances, IGMP proxying is configured on only one interface per instance.
- You cannot enable IGMP on interfaces with IGMP proxying enabled. Moreover, only the **igmp require-router-alert**, **igmp send-router-alert**, and **igmp version** commands can take effect on such interfaces.
- You cannot enable other multicast routing protocols (such as PIM-DM or PIM-SM) on interfaces with IGMP proxying enabled, or vice versa. However, the **source-lifetime**, **source-policy**, and **ssm-policy** commands configured in PIM view can still take effect. In addition, in IGMPv1, the designated router (DR) is elected by the working multicast routing protocol (such as PIM) to serve as the IGMP querier. Therefore, a downstream interface running IGMPv1 cannot be elected as the DR and thus cannot serve as the IGMP querier.
- You cannot enable IGMP proxying on a VLAN interface with IGMP Snooping enabled, or vice versa.

Configuring Multicast Forwarding on a Downstream Interface

Typically, only queriers are able to forward multicast traffic while non-queriers have no multicast forwarding capabilities, to avoid duplicate multicast flows. It is the same on IGMP proxy devices. Only the downstream interfaces acting as a querier can forward multicast traffic to downstream hosts.

However, when a downstream interface of a proxy device fails to win the querier election, you need to enable multicast forwarding on this interface.

Follow these steps to enable multicast forwarding on a downstream interface

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable multicast forwarding on a non-querier downstream interface	**igmp proxying forwarding**	Required Disabled by default.

Caution

On a multi-access network with more than one IGMP proxy device, you cannot enable multicast forwarding on any other non-querier downstream interface after one of the downstream interfaces of these IGMP proxy devices has been elected as the querier. Otherwise, duplicate multicast flows may be received on the multi-access network.

Displaying and Maintaining IGMP

To do...	Use the command...	Remarks
View IGMP multicast group information	**display igmp group** [*group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]	Available in any view
View layer 2 port information about IGMP multicast groups	**display igmp group port-info** [**vlan** *vlan-id*] [**slot** *slot-number*] [**verbose**]	Available in any view
View IGMP configuration and operation information	**display igmp interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View the information of IGMP proxying groups	**display igmp proxying group** [*group-address*] [**verbose**]	Available in any view
View information in the IGMP routing table	**display igmp routing-table** [*source-address* [**mask** { *mask* | *mask-length* }] | *group-address* [**mask** { *mask* | *mask-length* }]] * [**flags** { **act** | **suc** }]	Available in any view
View IGMP SSM mappings	**display igmp ssm-mapping** *group-address*	Available in any view
View the multicast group information created from IGMPv1 and IGMPv2 reports based on the configured IGMP SSM mappings	**display igmp ssm-mapping group** [*group-address* | **interface** *interface-type interface-number*] [**verbose**]	Available in any view
Clear IGMP multicast group information	**reset igmp group** { **all** | **interface** *interface-type interface-number* { **all** | *group-address* [**mask** { *mask* | *mask-length* }] [*source-address* [**mask** { *mask* | *mask-length* }]] } }	Available in user view
Clear layer 2 port information about IGMP multicast groups	**reset igmp group port-info** { **all** | *group-address* } [**vlan** *vlan-id*]	Available in user view
Clear IGMP SSM mappings	**reset igmp ssm-mapping group** { **all** | **interface** *interface-type interface-number* { **all** | *group-address* [**mask** { *mask* | *mask-length* }] [*source-address* [**mask** { *mask* | *mask-length* }]] } }	Available in user view

Note

- The **reset igmp group** command cannot clear the IGMP multicast group information of static joins.
- The **reset igmp group port-info** command cannot clear Layer 2 port information about IGMP multicast groups of static joins.

Caution

The **reset igmp group** command may cause an interruption of receivers' reception of multicast data.

IGMP Configuration Examples

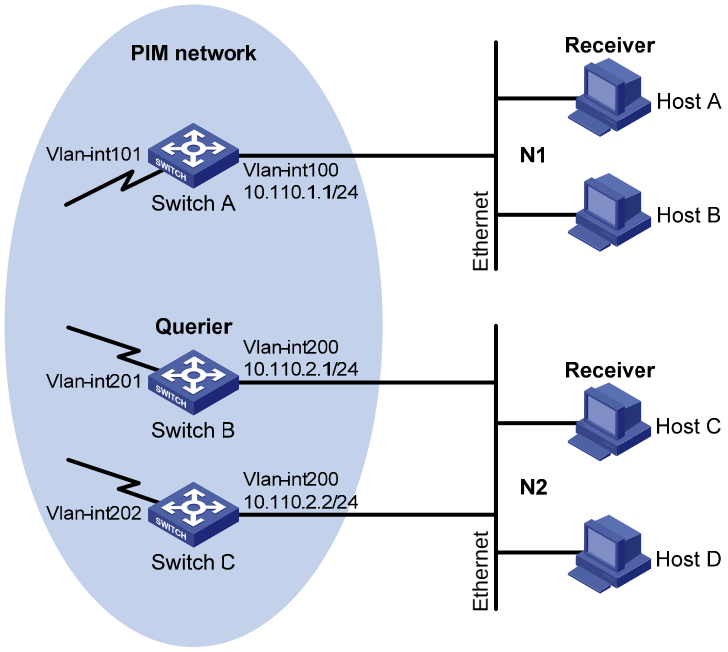
Basic IGMP Functions Configuration Example

Network requirements

- Receivers receive VOD information through multicast. Receivers of different organizations form stub networks N1 and N2, and Host A and Host C are receivers in N1 and N2 respectively.
- Switch A in the PIM network connects to N1, and both Switch B and Switch C connect to N2.
- Switch A connects to N1 through VLAN-interface 100, and to other devices in the PIM network through VLAN-interface 101.
- Switch B and Switch C connect to N2 through their respective VLAN-interface 200, and to other devices in the PIM network through VLAN-interface 201 and VLAN-interface 202 respectively.
- IGMPv2 is required between Switch A and N1. IGMPv2 is also required between the other two switches and N2, Switch B serves as the IGMP querier in N2 because its IP address is lower.

Network diagram

Figure 1-5 Network diagram for basic IGMP functions configuration



Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask of each interface as per Figure 1-5. The detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation on the PIM network. Ensure the network-layer interoperation on the PIM network and dynamic update of routing information among the switches through a unicast routing protocol. The detailed configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-DM and IGMP

Enable IP multicast routing on Switch A, enable PIM-DM on each interface, and enable IGMP on VLAN-interface 100.

```
<SwitchA> system-view
```

```
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim dm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim dm
[SwitchA-Vlan-interface101] quit
```

Enable IP multicast routing on Switch B, enable PIM-DM on each interface, and enable IGMP on VLAN-interface 200.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] igmp enable
[SwitchB-Vlan-interface200] pim dm
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 201
[SwitchB-Vlan-interface201] pim dm
[SwitchB-Vlan-interface201] quit
```

Enable IP multicast routing on Switch C, enable PIM-DM on each interface, and enable IGMP on VLAN-interface 200.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] pim dm
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] pim dm
[SwitchC-Vlan-interface202] quit
```

3) Verify the configuration

Carry out the **display igmp interface** command to view the IGMP configuration and operation status on each switch interface. For example:

View IGMP information on VLAN-interface 200 of Switch B.

```
[SwitchB] display igmp interface vlan-interface 200
 Vlan-interface200(10.110.2.1):
   IGMP is enabled
   Current IGMP version is 2
   Value of query interval for IGMP(in seconds): 60
   Value of other querier present interval for IGMP(in seconds): 125
   Value of maximum query response time for IGMP(in seconds): 10
   Querier for IGMP: 10.110.2.1 (this router)
  Total 1 IGMP Group reported
```

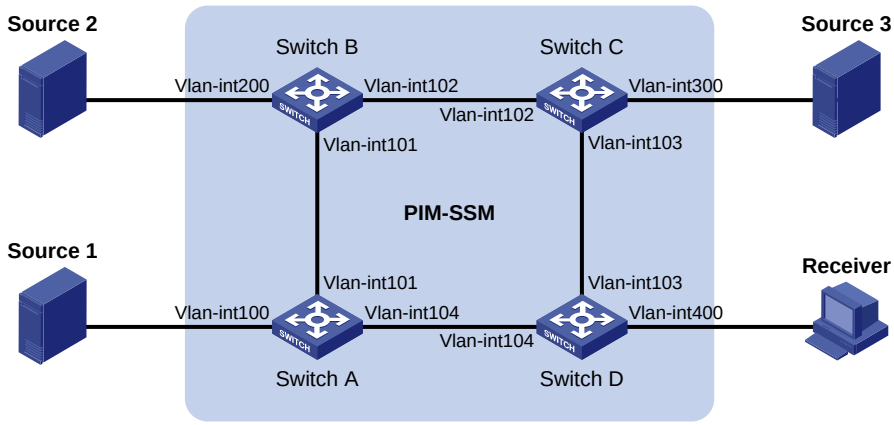
SSM Mapping Configuration Example

Network requirements

- On the PIM-SSM network shown in Figure 1-6, the receiver host receives VOD information through multicast. The receiver host runs IGMPv2, so it cannot specify the expected multicast sources in its membership reports.
- It is required to configure the IGMP SSM mapping feature on Switch D so that the receiver host will receive multicast data from Source 1 and Source 3 only.

Network diagram

Figure 1-6 Network diagram for IGMP SSM mapping configuration



Device	Interface	IP address	Device	Interface	IP address
Source 1	—	133.133.1.1/24	Source 3	—	133.133.3.1/24
Source 2	—	133.133.2.1/24	Receiver	—	133.133.4.1/24
Switch A	Vlan-int100	133.133.1.2/24	Switch C	Vlan-int300	133.133.3.2/24
	Vlan-int101	192.168.1.1/24		Vlan-int103	192.168.3.1/24
	Vlan-int104	192.168.4.2/24		Vlan-int102	192.168.2.2/24
Switch B	Vlan-int200	133.133.2.2/24	Switch D	Vlan-int400	133.133.4.2/24
	Vlan-int101	192.168.1.2/24		Vlan-int103	192.168.3.2/24
	Vlan-int102	192.168.2.1/24		Vlan-int104	192.168.4.1/24

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask of each interface as per Figure 1-6. The detailed configuration steps are omitted here.

Configure OSPF for interoperability among the switches. Ensure the network-layer interoperation on the PIM-SSM network and dynamic update of routing information among the switches through a unicast routing protocol. The detailed configuration steps are omitted here.

2) Enable IP multicast routing, enable PIM-SM on each interface, and enable IGMP and IGMP SSM mapping on the host-side interface.

Enable IP multicast routing on Switch D, enable PIM-SM on each interface, and enable IGMPv3 and IGMP SSM mapping on VLAN-interface 400.

```
<SwitchD> system-view
[SwitchD] multicast routing-enable
[SwitchD] interface vlan-interface 400
```

```
[SwitchD-Vlan-interface400] igmp enable
[SwitchD-Vlan-interface400] igmp version 3
[SwitchD-Vlan-interface400] igmp ssm-mapping enable
[SwitchD-Vlan-interface400] pim sm
[SwitchD-Vlan-interface400] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim sm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 104
[SwitchD-Vlan-interface104] pim sm
[SwitchD-Vlan-interface104] quit
```

Enable IP multicast routing on Switch A, and enable PIM-SM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 104
[SwitchA-Vlan-interface104] pim sm
[SwitchA-Vlan-interface104] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A.

3) Configure the SSM group range

Configure the SSM group range 232.1.1.0/24 on Switch D.

```
[SwitchD] acl number 2000
[SwitchD-acl-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[SwitchD-acl-basic-2000] quit
[SwitchD] pim
[SwitchD-pim] ssm-policy 2000
[SwitchD-pim] quit
```

The configuration on Switch A, Switch B and Switch C is similar to that on Switch D.

4) Configure IGMP SSM mappings

Configure IGMP SSM mappings on Switch D.

```
[SwitchD] igmp
[SwitchD-igmp] ssm-mapping 232.1.1.1 32 133.133.1.1
[SwitchD-igmp] ssm-mapping 232.1.1.1 32 133.133.3.1
[SwitchD-igmp] quit
```

5) Verify the configuration

Use the **display igmp ssm-mapping** command to view the IGMP SSM mappings on the switch.

View the IGMP SSM mapping information for multicast group 232.1.1.1 on Switch D.

```
[SwitchD] display igmp ssm-mapping 232.1.1.1
Group address: 232.1.1.1
Source list:
```

```
        133.133.1.1
        133.133.3.1
```

Use the **display igmp ssm-mapping group** command to view the multicast group information created based on the configured IGMP SSM mappings.

View the IGMP multicast group information created based on the IGMP SSM mappings on Switch D.

```
[SwitchD] display igmp ssm-mapping group
Total 1 IGMP SSM-mapping Group(s).
Interface group report information of VPN-Instance: public net
 Vlan-interface400(133.133.4.2):
  Total 1 IGMP SSM-mapping Group reported
   Group Address       Last Reporter   Uptime      Expires
   232.1.1.1           133.133.4.1     00:02:04    off
```

Use the **display pim routing-table** command to view the PIM routing table information on each switch.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
Total 0 (*, G) entry; 2 (S, G) entry

 (133.133.1.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface104
         Upstream neighbor: 192.168.4.2
         RPF prime neighbor: 192.168.4.2
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: Vlan-interface400
                 Protocol: igmp, UpTime: 00:13:25, Expires: never

 (133.133.3.1, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface103
         Upstream neighbor: 192.168.3.1
         RPF prime neighbor: 192.168.3.1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: Vlan-interface400
                 Protocol: igmp, UpTime: 00:13:25, Expires: never
```

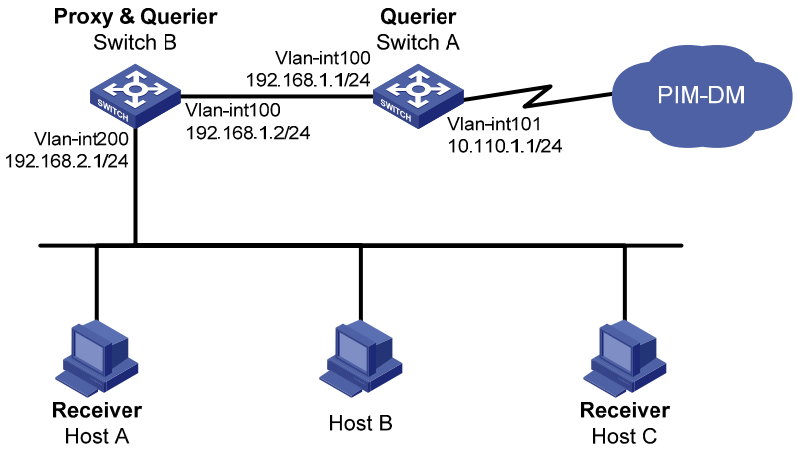
IGMP Proxying Configuration Example

Network requirements

- PIM-DM is required to run on the core network. Host A and Host C in the stub network receive VOD information destined to multicast group 224.1.1.1.
- It is required to configure the IGMP Proxying feature on Switch B so that Switch B can maintain group memberships and forward multicast traffic without running PIM-DM.

Network diagram

Figure 1-7 Network diagram for IGMP Proxying configuration



Configuration procedure

1) Configure IP addresses

Configure the IP address and subnet mask of each interface as per Figure 1-7. The detailed configuration steps are omitted here.

2) Enable IP multicast routing, PIM-DM, IGMP, and IGMP Proxying.

Enable IP multicast routing on Switch A, PIM-DM on VLAN-interface 101, and IGMP on VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim dm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim dm
[SwitchA-Vlan-interface100] quit
```

Enable IP multicast routing on Switch B, IGMP Proxying on VLAN-interface 100, and IGMP on VLAN-interface 200.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp proxying enable
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] igmp enable
[SwitchB-Vlan-interface200] quit
```

3) Verify the configuration

Use the **display igmp interface** command to view the IGMP configuration and operation information on an interface. For example

View the IGMP configuration and operation information on VLAN-interface 100 of Switch B.

```
[SwitchB] display igmp interface vlan-interface 100 verbose
 Vlan-interface100(192.168.1.2):
   IGMP proxy is enabled
   Current IGMP version is 2
   Multicast routing on this interface: enabled
   Require-router-alert: disabled
   Version1-querier-present-timer-expiry: 00:00:20
```

Use the **display igmp group** command to view the IGMP multicast group information. For example,

View the IGMP multicast group information on Switch A.

```
[SwitchA] display igmp group
Total 1 IGMP Group(s).
Interface group report information
 Vlan-interface100(192.168.1.1):
  Total 1 IGMP Groups reported
   Group Address   Last Reporter   Uptime      Expires
   224.1.1.1       192.168.1.2     00:02:04    00:01:15
```

As shown above, IGMP reports from the hosts are forwarded to Switch A through the proxy interface, VLAN-interface 100 on Switch B.

Troubleshooting IGMP

No Membership Information on the Receiver-Side Router

Symptom

When a host sends a report for joining multicast group G, there is no membership information of the multicast group G on the router closest to that host.

Analysis

- The correctness of networking and interface connections and whether the protocol layer of the interface is up directly affect the generation of group membership information.
- Multicast routing must be enabled on the router, and IGMP must be enabled on the interface connecting to the host.
- If the IGMP version on the router interface is lower than that on the host, the router will not be able to recognize the IGMP report from the host.
- If the **igmp group-policy** command has been configured on the interface, the interface cannot receive report messages that fail to pass filtering.

Solution

1) Check that the networking, interface connection, and IP address configuration are correct. Check the interface information with the display igmp interface command. If there is no information output, the interface is in an abnormal state. This is usually because you have configured the shutdown command on the interface, the interface is not properly connected, or the IP address configuration is not correctly done.
2) Check that multicast routing is enabled. Carry out the display current-configuration command to check whether the multicast routing-enable command has been executed. If not, carry out the multicast routing-enable command in system view to enable IP multicast routing. In addition, check that IGMP is enabled on the corresponding interfaces.

3) Check the IGMP version on the interface. You can use the display igmp interface command to check whether the IGMP version on the interface is lower than that on the host.
4) Check that no ACL rule has been configured to restrict the host from joining the multicast group G. Carry out the display current-configuration interface command to check whether the igmp group-policy command has been executed. If the host is restricted from joining the multicast group G, the ACL rule must be modified to allow receiving the reports for the multicast group G.

Inconsistent Memberships on Routers on the Same Subnet

Symptom

Different memberships are maintained on different IGMP routers on the same subnet.

Analysis

- A router running IGMP maintains multiple parameters for each interface, and these parameters influence one another, forming very complicated relationships. Inconsistent IGMP interface parameter configurations for routers on the same subnet will surely result in inconsistency of memberships.
- In addition, although an IGMP router is compatible with a host that is running a different IGMP version, all routers on the same subnet must run the same version of IGMP. Inconsistent IGMP versions running on routers on the same subnet will also lead to inconsistency of IGMP memberships.

Solution

1) Check the IGMP configuration. Carry out the display current-configuration command to view the IGMP configuration information on the interfaces.
2) Carry out the display igmp interface command on all routers on the same subnet to check the IGMP-related timer settings. Make sure that the settings are consistent on all the routers.
3) Use the display igmp interface command to check whether all the routers on the same subnet are running the same version of IGMP.

Table of Contents

1 PIM Configuration

When configuring PIM, go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running the PIM protocol.

PIM Overview

Protocol Independent Multicast (PIM) provides IP multicast forwarding by leveraging static routes or unicast routing tables generated by any unicast routing protocol, such as routing information protocol (RIP), open shortest path first (OSPF), intermediate system to intermediate system (IS-IS), or border gateway protocol (BGP). Independent of the unicast routing protocols running on the device, multicast routing can be implemented as long as the corresponding multicast routing entries are created through unicast routes. PIM uses the reverse path forwarding (RPF) mechanism to implement multicast forwarding. When a multicast packet arrives on an interface of the device, it is subject to an RPF check. If the RPF check succeeds, the device creates the corresponding routing entry and forwards the packet; if the RPF check fails, the device discards the packet. For more information about RPF, refer to *Multicast Routing and Forwarding Configuration* in the *IP Multicast Volume*.

Based on the implementation mechanism, PIM falls into two modes:

- Protocol Independent Multicast–Dense Mode (PIM-DM), and
- Protocol Independent Multicast–Sparse Mode (PIM-SM).

Note

To facilitate description, a network comprising PIM-capable routers is referred to as a “PIM domain” in this document.

Introduction to PIM-DM

PIM-DM is a type of dense mode multicast protocol. It uses the “push mode” for multicast forwarding, and is suitable for small-sized networks with densely distributed multicast members.

The basic implementation of PIM-DM is as follows:

- PIM-DM assumes that at least one multicast group member exists on each subnet of a network, and therefore multicast data is flooded to all nodes on the network. Then, branches without multicast forwarding are pruned from the forwarding tree, leaving only those branches that contain receivers. This “flood and prune” process takes place periodically, that is, pruned branches resume multicast forwarding when the pruned state times out and then data is re-flooded down these branches, and then are pruned again.
- When a new receiver on a previously pruned branch joins a multicast group, to reduce the join latency, PIM-DM uses a graft mechanism to resume data forwarding to that branch.

Generally speaking, the multicast forwarding path is a source tree, namely a forwarding tree with the multicast source as its “root” and multicast group members as its “leaves”. Because the source tree is the shortest path from the multicast source to the receivers, it is also called shortest path tree (SPT).

How PIM-DM Works

The working mechanism of PIM-DM is summarized as follows:

- Neighbor discovery
- SPT building
- Graft
- Assert

Neighbor discovery

In a PIM domain, a PIM router discovers PIM neighbors, maintains PIM neighboring relationships with other routers, and builds and maintains SPTs by periodically multicasting hello messages to all other PIM routers (224.0.0.13).

 **Note**

Every PIM-enabled interface on a router sends hello messages periodically, and thus learns the PIM neighboring information pertinent to the interface.

SPT establishment

The process of building an SPT is the process of “flood and prune”.

1) In a PIM-DM domain, when a multicast source S sends multicast data to multicast group G, the multicast packet is first flooded throughout the domain: The router first performs RPF check on the multicast packet. If the packet passes the RPF check, the router creates an (S, G) entry and forwards the data to all downstream nodes in the network. In the flooding process, an (S, G) entry is created on all the routers in the PIM-DM domain.
2) Then, nodes without receivers downstream are pruned: A router having no receivers downstream sends a prune message to the upstream node to “tell” the upstream node to delete the

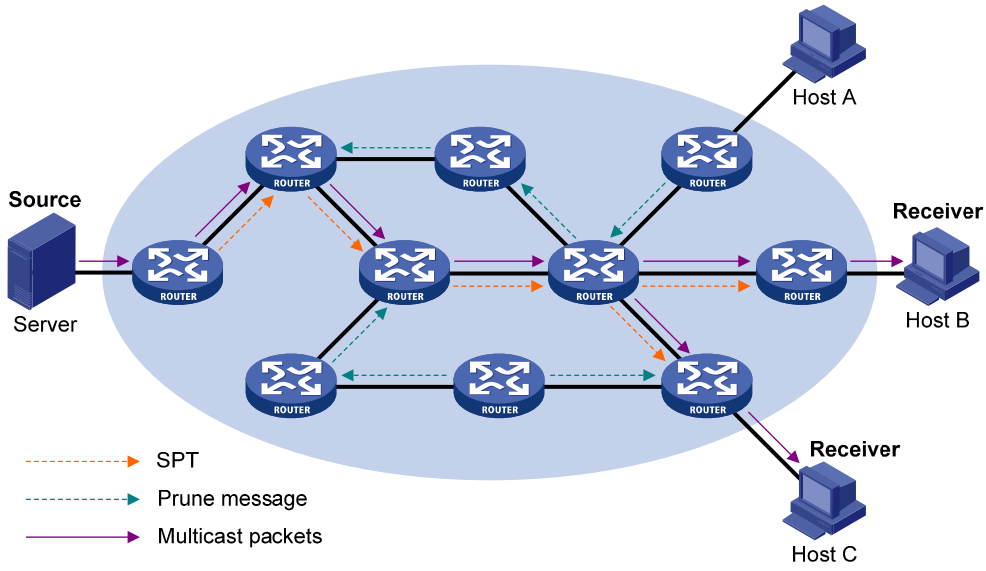
corresponding interface from the outgoing interface list in the (S, G) entry and stop forwarding subsequent packets addressed to that multicast group down to this node.

Note

- An (S, G) entry contains the multicast source address S, multicast group address G, outgoing interface list, and incoming interface.
- For a given multicast stream, the interface that receives the multicast stream is referred to as “upstream”, and the interfaces that forward the multicast stream are referred to as “downstream”.

A prune process is first initiated by a leaf router. As shown in Figure 1-1, a router without any receiver attached to it (the router connected with Host A, for example) sends a prune message, and this prune process goes on until only necessary branches are left in the PIM-DM domain. These branches constitute the SPT.

Figure 1-1 SPT establishment



The “flood and prune” process takes place periodically. A pruned state timeout mechanism is provided. A pruned branch restarts multicast forwarding when the pruned state times out and then is pruned again when it no longer has any multicast receiver.

Note

Pruning has a similar implementation in PIM-SM.

Graft

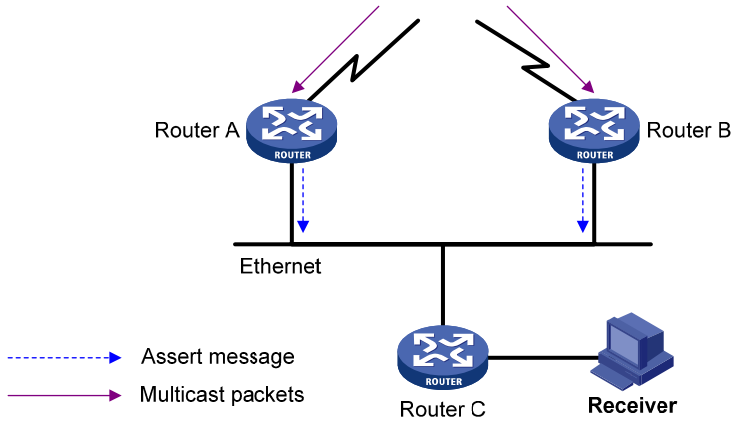
When a host attached to a pruned node joins a multicast group, to reduce the join latency, PIM-DM uses a graft mechanism to resume data forwarding to that branch. The process is as follows:

1) The node that needs to receive multicast data sends a graft message toward its upstream node, as a request to join the SPT again.
2) Upon receiving this graft message, the upstream node puts the interface on which the graft was received into the forwarding state and responds with a graft-ack message to the graft sender.
3) If the node that sent a graft message does not receive a graft-ack message from its upstream node, it will keep sending graft messages at a configurable interval until it receives an acknowledgment from its upstream node.

Assert

The assert mechanism is used to shutoff duplicate multicast flows onto the same multi-access network, where more than one multicast routers exists, by electing a unique multicast forwarder on the multi-access network.

Figure 1-2 Assert mechanism



As shown in Figure 1-2, after Router A and Router B receive an (S, G) packet from the upstream node, they both forward the packet to the local subnet. As a result, the downstream node Router C receives two identical multicast packets, and both Router A and Router B, on their own local interface, receive a duplicate packet forwarded by the other. Upon detecting this condition, both routers send an assert message to all PIM routers (224.0.0.13) through the interface on which the packet was received. The assert message contains the following information: the multicast source address (S), the multicast group address (G), and the preference and metric of the unicast route to the source. By comparing these parameters, either Router A or Router B becomes the unique forwarder of the subsequent (S, G) packets on the multi-access subnet. The comparison process is as follows:

1) The router with a higher unicast route preference to the source wins;
2) If both routers have the same unicast route preference to the source, the router with a smaller metric to the source wins;
3) If there is a tie in route metric to the source, the router with a higher IP address of the local interface wins.

Introduction to PIM-SM

PIM-DM uses the “flood and prune” principle to build SPTs for multicast data distribution. Although an SPT has the shortest path, it is built with a low efficiency. Therefore the PIM-DM mode is not suitable for large- and medium-sized networks.

PIM-SM is a type of sparse mode multicast protocol. It uses the “pull mode” for multicast forwarding, and is suitable for large- and medium-sized networks with sparsely and widely distributed multicast group members.

The basic implementation of PIM-SM is as follows:

- PIM-SM assumes that no hosts need to receive multicast data. In the PIM-SM mode, routers must specifically request a particular multicast stream before the data is forwarded to them. The core task for PIM-SM to implement multicast forwarding is to build and maintain rendezvous point trees (RPTs). An RPT is rooted at a router in the PIM domain as the common node, or rendezvous point (RP), through which the multicast data travels along the RPT and reaches the receivers.
- When a receiver is interested in the multicast data addressed to a specific multicast group, the router connected to this receiver sends a join message to the RP corresponding to that multicast group. The path along which the message goes hop by hop to the RP forms a branch of the RPT.
- When a multicast source sends multicast streams to a multicast group, the source-side designated router (DR) first registers the multicast source with the RP by sending register messages to the RP by unicast until it receives a register-stop message from the RP. The arrival of a register message at the RP triggers the establishment of an SPT. Then, the multicast source sends subsequent multicast packets along the SPT to the RP. Upon reaching the RP, the multicast packet is duplicated and delivered to the receivers along the RPT.

Note

Multicast traffic is duplicated only where the distribution tree branches, and this process automatically repeats until the multicast traffic reaches the receivers.

How PIM-SM Works

The working mechanism of PIM-SM is summarized as follows:

- Neighbor discovery
- DR election
- RP discovery
- RPT building
- Multicast source registration
- Switchover to SPT
- Assert

Neighbor discovery

PIM-SM uses a similar neighbor discovery mechanism as PIM-DM does. For details, refer to Neighbor discovery.

DR election

PIM-SM also uses hello messages to elect a DR for a multi-access network (such as Ethernet). The elected DR will be the only multicast forwarder on this multi-access network.

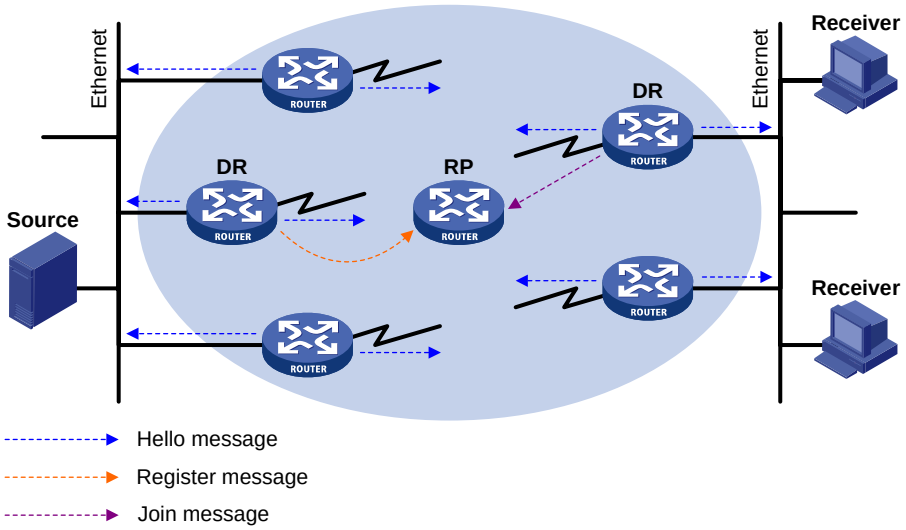
A DR must be elected in a multi-access network, no matter this network connects to multicast sources or to receivers. The DR at the receiver side sends join messages to the RP; the DR at the multicast source side sends register messages to the RP.

Note

- A DR is elected on a multi-access subnet by means of comparison of the priorities and IP addresses carried in hello messages. An elected DR is substantially meaningful to PIM-SM. PIM-DM itself does not require a DR. However, if IGMPv1 runs on any multi-access network in a PIM-DM domain, a DR must be elected to act as the IGMPv1 querier on that multi-access network.
- IGMP must be enabled on a device that acts as a receiver-side DR before receivers attached to this device can join multicast groups through this DR.

For details about IGMP, refer to *IGMP Configuration* in the *IP Multicast Volume*.

Figure 1-3 DR election



As shown in Figure 1-3, the DR election process is as follows:

1) Routers on the multi-access network send hello messages to one another. The hello messages contain the router priority for DR election. The router with the highest DR priority will become the DR.
2) In the case of a tie in the router priority, or if any router in the network does not support carrying the DR-election priority in hello messages, the router with the highest IP address will win the DR election.

When the DR fails, a timeout in receiving hello message triggers a new DR election process among the other routers.

RP discovery

The RP is the core of a PIM-SM domain. For a small-sized, simple network, one RP is enough for forwarding information throughout the network, and the position of the RP can be statically specified on each router in the PIM-SM domain. In most cases, however, a PIM-SM network covers a wide area and a huge amount of multicast traffic needs to be forwarded through the RP. To lessen the RP burden and

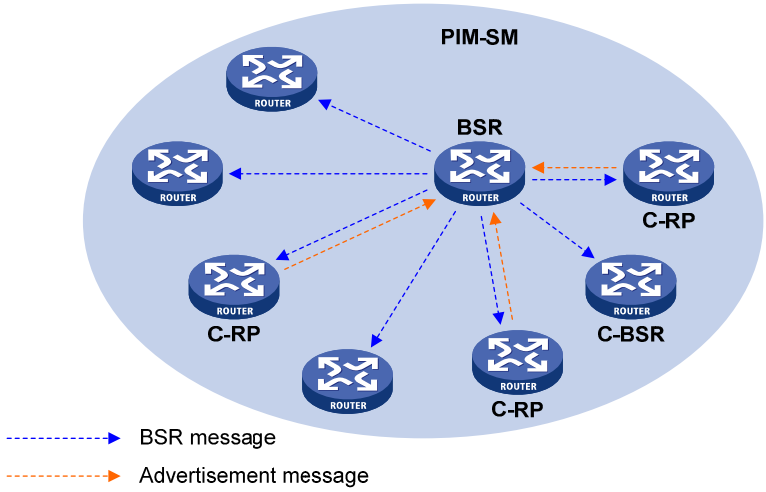
optimize the topological structure of the RPT, multiple candidate RPs (C-RPs) can be configured in a PIM-SM domain, among which an RP is dynamically elected through the bootstrap mechanism. Each elected RP serves a different multicast group range. For this purpose, a bootstrap router (BSR) must be configured. The BSR serves as the administrative core of the PIM-SM domain. A PIM-SM domain can have only one BSR, but can have multiple candidate-BSRs (C-BSRs). Once the BSR fails, a new BSR is automatically elected from the C-BSRs to avoid service interruption.

Note

- An RP can serve multiple multicast groups or all multicast groups. Only one RP can serve a given multicast group at a time.
- A device can serve as a C-RP and a C-BSR at the same time.

As shown in Figure 1-4, each C-RP periodically unicasts its advertisement messages (C-RP-Adv messages) to the BSR. A C-RP-Adv message contains the address of the advertising C-RP and the multicast group range it serves. The BSR collects these advertisement messages and chooses the appropriate C-RP information for each multicast group to form an RP-set, which is a database of mappings between multicast groups and RPs. The BSR then encapsulates the RP-set in the bootstrap messages it periodically originates and floods the bootstrap messages to the entire PIM-SM domain.

Figure 1-4 BSR and C-RPs



Based on the information in the RP-sets, all routers in the network can calculate the location of the corresponding RPs based on the following rules:

1) The C-RP with the highest priority wins.
2) If all the C-RPs have the same priority, their hash values are calculated through the hashing algorithm. The C-RP with the largest hash value wins.
3) If all the C-RPs have the same priority and hash value, the C-RP has the highest IP address wins.

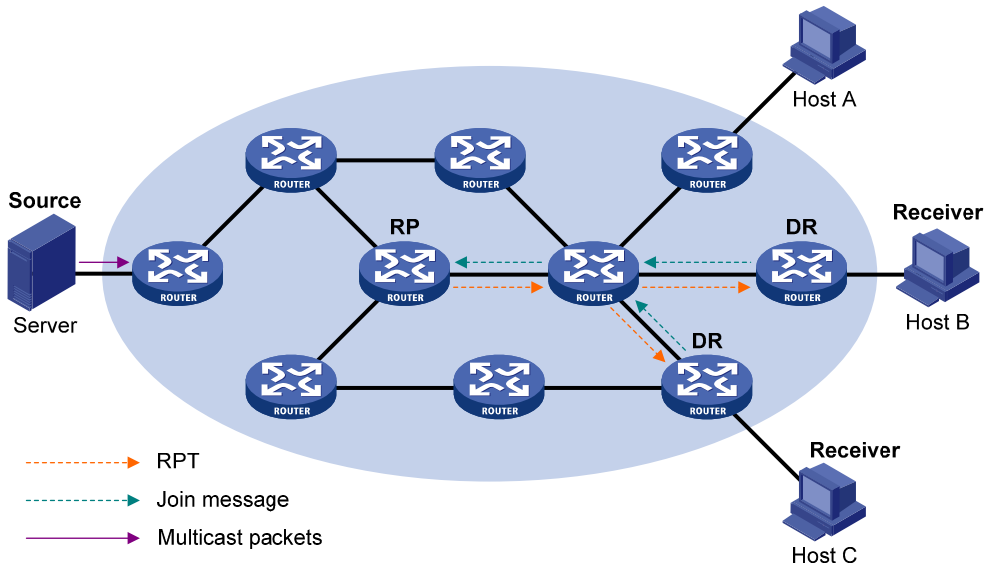
The hashing algorithm used for RP calculation is: Value (G, M, C_i) = (1103515245 * ((1103515245 * (G & M) + 12345) XOR C_i) + 12345) mod 2^{31}. The table below gives the meanings of the values in this algorithm.

Table 1-1 Values in the hashing algorithm

Value	Description
Value	Hash value
G	IP address of the multicast group
M	Hash mask length
C_i	IP address of the C-RP
&	Logical operator of “and”
XOR	Logical operator of “exclusive-or”
mod	Modulo operator, which gives the remainder of an integer division

RPT establishment

Figure 1-5 RPT establishment in a PIM-SM domain



As shown in Figure 1-5, the process of building an RPT is as follows:

1) When a receiver joins multicast group G, it uses an IGMP message to inform the directly connected DR.
2) Upon getting the receiver information, the DR sends a join message, which is hop by hop forwarded to the RP corresponding to the multicast group.
3) The routers along the path from the DR to the RP form an RPT branch. Each router on this branch generates a (*, G) entry in its forwarding table. The * means any multicast source. The RP is the root, while the DRs are the leaves, of the RPT.

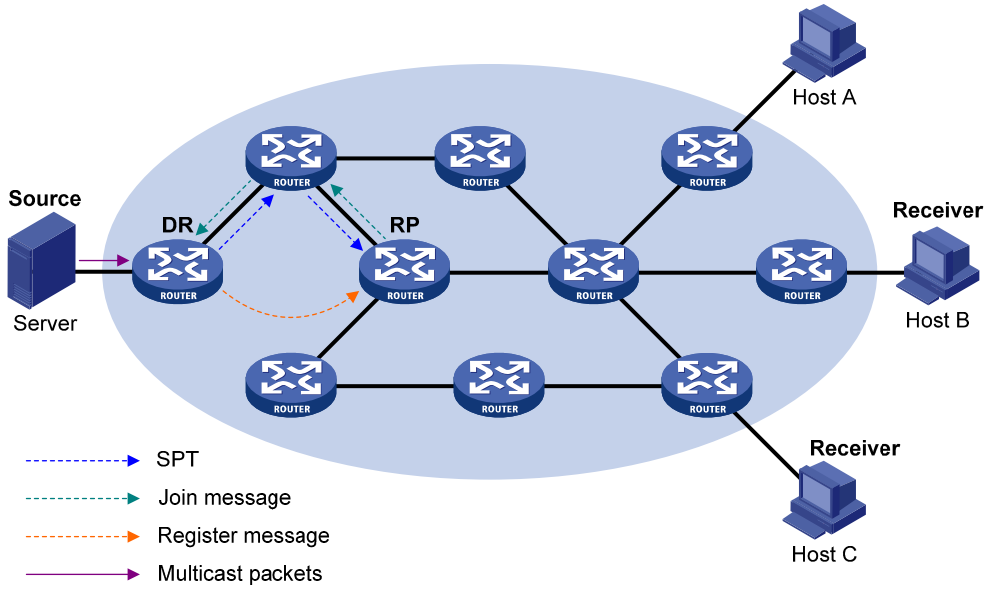
The multicast data addressed to the multicast group G flows through the RP, reaches the corresponding DR along the established RPT, and finally is delivered to the receiver.

When a receiver is no longer interested in the multicast data addressed to multicast group G, the directly connected DR sends a prune message, which goes hop by hop along the RPT to the RP. Upon receiving the prune message, the upstream node deletes the interface connected with this downstream node from the outgoing interface list and checks whether it itself has receivers for that multicast group. If not, the router continues to forward the prune message to its upstream router.

Multicast source registration

The purpose of multicast source registration is to inform the RP about the existence of the multicast source.

Figure 1-6 Multicast source registration



As shown in Figure 1-6, the multicast source registers with the RP as follows:

1) When the multicast source S sends the first multicast packet to multicast group G, the DR directly connected with the multicast source, upon receiving the multicast packet, encapsulates the packet in a PIM register message, and sends the message to the corresponding RP by unicast.
2) When the RP receives the register message, it extracts the multicast packet from the register message and forwards the multicast packet down the RPT, and sends an (S, G) join message hop by hop toward the multicast source. Thus, the routers along the path from the RP to the multicast source constitute an SPT branch. Each router on this branch generates an (S, G) entry in its forwarding table. The DR at the multicast source side is the root, while the RP is the leaf, of the SPT.
3) The subsequent multicast data from the multicast source travels along the established SPT to the RP, and then the RP forwards the data along the RPT to the receivers. When the multicast traffic arrives at the RP along the SPT, the RP sends a register-stop message to the source-side DR by unicast to stop the source registration process.

Note

The RP is configured to initiate an SPT switchover as described in this section. Otherwise, the DR at the multicast source side keeps encapsulating multicast data in register messages and the registration process will not stop unless no outgoing interfaces exist in the (S, G) entry on the RP.

Switchover to SPT

In a PIM-SM domain, a multicast group corresponds to one RP and RPT. Before the SPT switchover takes place, the DR at the multicast source side encapsulates all multicast data destined to the multicast group in register messages and sends these messages to the RP. Upon receiving these register messages, the RP abstracts the multicast data and sends the multicast data down the RPT to the DRs at the receiver side. The RP acts as a transfer station for all multicast packets. The whole process involves three issues as follows:

- The DR at the source side and the RP need to implement complicated encapsulation and decapsulation of multicast packets.
- Multicast packets are delivered along a path that is not necessarily the shortest one.
- When the multicast traffic increases, a great burden is added to the RP, increasing the risk of failure.

To solve the issues, PIM-SM allows an RP or the DR at the receiver side to initiate an SPT switchover process:

1) The RP initiates an SPT switchover process

Upon receiving the first multicast packet, the RP sends an (S, G) join message hop by hop toward the multicast source to establish an SPT between the DR at the source side and the RP. The subsequent multicast data from the multicast source travel along the established SPT to the RP.

Note

For details about the SPT switchover initiated by the RP, refer to Multicast source registration.

---

2) The receiver-side DR initiates an SPT switchover process

Upon receiving the first multicast packet, the receiver-side DR initiates an SPT switchover process, as follows:

- First, the receiver-side DR sends an (S, G) join message hop by hop toward the multicast source. When the join message reaches the source-side DR, all the routers on the path have installed the (S, G) entry in their forwarding table, and thus an SPT branch is established.
- When subsequent multicast packets arrive at the router at the junction of the RPT and SPT, the router drops those transmitted along the RPT and sends an RP-bit prune message hop by hop to the RP. Upon receiving this prune message, the RP sends a prune message toward the multicast source (suppose only one receiver exists), thus to implement the SPT switchover.
- Finally, multicast data is directly sent from the source to the receivers along the SPT.

PIM-SM builds SPTs through SPT switchover more economically than PIM-DM does through the “flood and prune” mechanism.

Assert

PIM-SM uses a similar assert mechanism as PIM-DM does. For details, refer to Assert.

Introduction to Administrative Scoping in PIM-SM

Division of PIM-SM domains

Typically, a PIM-SM domain contains only one BSR, which is responsible for advertising RP-set information within the entire PIM-SM domain. The information for all multicast groups is forwarded within the network scope administered by the BSR. We call this non-scoped BSR mechanism.

To implement refined management, a PIM-SM domain can be divided into one global scope zone and multiple administratively scoped zones (admin-scope zones). We call this administrative scoping mechanism.

The administrative scoping mechanism effectively releases stress on the management in a single-BSR domain and enables provision of zone-specific services using private group addresses.

Admin-scope zones are divided specific to multicast groups. The boundary of the admin-scope zone is formed by zone border routers (ZBRs). Each admin-scope zone maintains one BSR, which serves multicast groups within a specific range. Multicast protocol packets, such as assert messages and bootstrap messages, for a specific group range cannot cross the admin-scope zone boundary. Multicast group ranges served by different admin-scope zones can be overlapped. A multicast group is valid only within its local admin-scope zone, functioning as a private group address.

The global scope zone maintains a BSR, which serves the multicast groups that do not belong to any admin-scope zone.

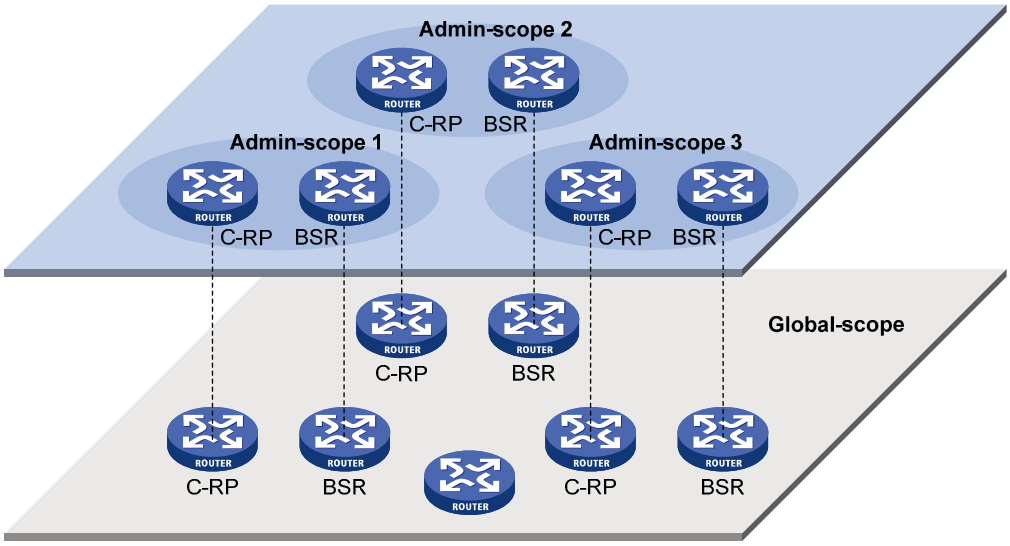
Relationship between admin-scope zones and the global scope zone

The global scope zone and each admin-scope zone have their own C-RPs and BSRs. These devices are effective only in their respective admin-scope zones. Namely, BSR election and RP election are implemented independently within each admin-scope zone. Each admin-scope zone has its own boundary. The multicast information cannot cross this border in either direction. A better understanding of the global scope zone and admin-scope zones should be based on two aspects: geographical space and group address range.

Geographical space

Admin-scope zones are logical zones specific to particular multicast groups, and each admin-scope zone must be geographically independent of every other one, as shown in Figure 1-7.

Figure 1-7 Relationship between admin-scope zones and the global scope zone in geographic space

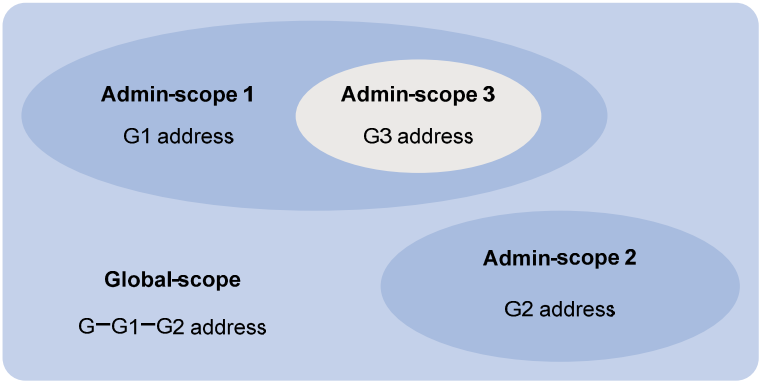


Admin-scope zones are geographically separated from one another. Namely, a router must not serve different admin-scope zones. In other words, different admin-scope zones contain different routers, whereas the global scope zone covers all routers in the PIM-SM domain.

In terms of multicast group address ranges

Each admin-scope zone serves specific multicast groups. Usually, these addresses have no intersections; however, they may overlap one another.

Figure 1-8 Relationship between admin-scope zones and the global scope zone in group address ranges



In Figure 1-8, the group address ranges of admin-scope 1 and 2 have no intersection, whereas the group address range of admin-scope 3 is a subset of the address range of admin-scope 1. The group address range of the global scope zone covers all the group addresses other than those of all the admin-scope zones. That is, the group address range of the global scope zone is G-G1-G2. In other words, there is a supplementary relationship between the global scope zone and all the admin-scope zones in terms of group address ranges.

SSM Model Implementation in PIM

The source-specific multicast (SSM) model and the any-source multicast (ASM) model are two opposite models. Presently, the ASM model includes the PIM-DM and PIM-SM modes. The SSM model can be implemented by leveraging part of the PIM-SM technique.

The SSM model provides a solution for source-specific multicast. It maintains the relationships between hosts and routers through IGMPv3.

In actual application, part of the PIM-SM technique is adopted to implement the SSM model. In the SSM model, receivers know exactly where a multicast source is located by means of advertisements, consultancy, and so on. Therefore, no RP is needed, no RPT is required, there is no source registration process, and there is no need of using the multicast source discovery protocol (MSDP) for discovering sources in other PIM domains.

Compared with the ASM model, the SSM model only needs the support of IGMPv3 and some subsets of PIM-SM. The operation mechanism of PIM-SSM can be summarized as follows:

- Neighbor discovery
- DR election
- SPT building

Neighbor discovery

PIM-SSM uses the same neighbor discovery mechanism as in PIM-DM and PIM-SM. Refer to Neighbor discovery.

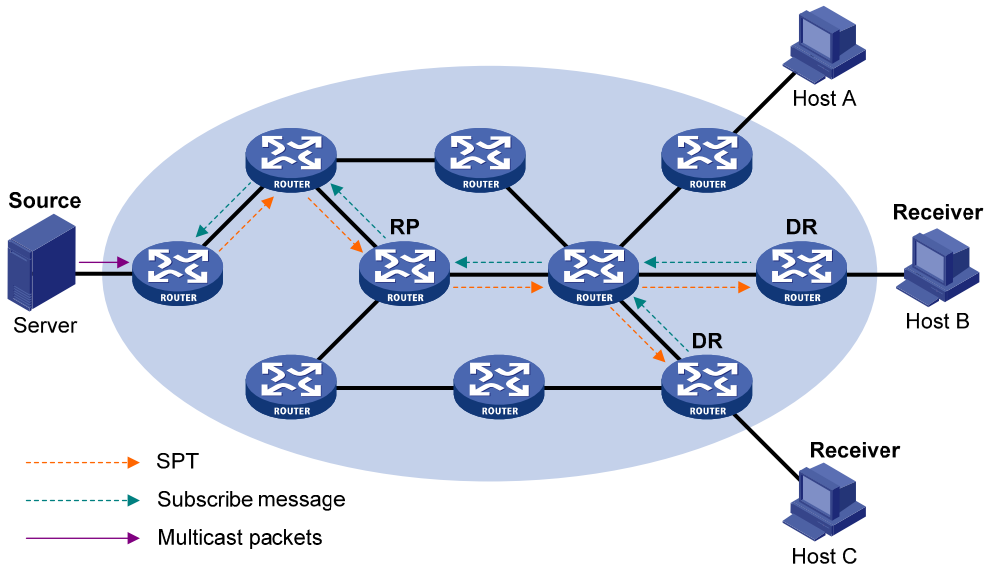
DR election

PIM-SSM uses the same DR election mechanism as in PIM-SM. Refer to DR election.

Construction of SPT

Whether to build an RPT for PIM-SM or an SPT for PIM-SSM depends on whether the multicast group the receiver is to join falls in the SSM group range (SSM group range reserved by IANA is 232.0.0.0/8).

Figure 1-9 SPT establishment in PIM-SSM



As shown in Figure 1-9, Host B and Host C are multicast information receivers. They send IGMPv3 report messages to the respective DRs to express their interest in the information of the specific multicast source S.

Upon receiving a report message, the DR first checks whether the group address in this message falls in the SSM group range:

- If so, the DR sends a subscribe message for channel subscription hop by hop toward the multicast source S. An (S, G) entry is created on all routers on the path from the DR to the source. Thus, an SPT is built in the network, with the source S as its root and receivers as its leaves. This SPT is the transmission channel in PIM-SSM.
- If not, the PIM-SM process is followed: the DR needs to send a (*, G) join message to the RP, and a multicast source registration process is needed.

Note

In PIM-SSM, the “channel” concept is used to refer to a multicast group, and the “channel subscription” concept is used to refer to a join message.

Protocols and Standards

PIM-related specifications are as follows:

- RFC 4601: Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)
- RFC 3973: Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification(Revised)
- RFC 4607: Source-Specific Multicast for IP
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- Draft-ietf-ssm-overview-05: An Overview of Source-Specific Multicast (SSM)

Configuring PIM-DM

PIM-DM Configuration Task List

Complete these tasks to configure PIM-DM:

Task	Remarks
Enabling PIM-DM	Required
Enabling State-Refresh Capability	Optional
Configuring State-Refresh Parameters	Optional
Configuring PIM-DM Graft Retry Period	Optional
Configuring PIM Common Features	Optional

Configuration Prerequisites

Before configuring PIM-DM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-DM, prepare the following data:

- The interval between state-refresh messages
- Minimum time to wait before receiving a new refresh message
- TTL value of state-refresh messages
- Graft retry period

Enabling PIM-DM

With PIM-DM enabled, a router sends hello messages periodically to discover PIM neighbors and processes messages from the PIM neighbors. When deploying a PIM-DM domain, you are recommended to enable PIM-DM on all non-border interfaces of the routers.

Enabling PIM-DM globally

Follow these steps to enable PIM-DM globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable PIM-DM	**pim dm**	Required Disabled by default

Caution

- All the interfaces on the same device must work in the same PIM mode.
- PIM-DM does not work with multicast groups in the SSM group grange.

Note

For details about the **multicast routing-enable** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Enabling State-Refresh Capability

A multi-access subnet can have the state-refresh capability only if the state-refresh capability is enabled on all PIM routers on the subnet.

Follow these steps to enable the state-refresh capability:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable state-refresh	**pim state-refresh-capable**	Optional Enabled by default

Configuring State-Refresh Parameters

To avoid the resource-consuming reflooding of unwanted traffic caused by timeout of pruned interfaces, the router directly connected with the multicast source periodically sends an (S, G) state-refresh message, which is forwarded hop by hop along the initial multicast flooding path of the PIM-DM domain, to refresh the prune timer state of all the routers on the path.

A router may receive multiple state-refresh messages within a short time, of which some may be duplicated messages. To keep a router from receiving such duplicated messages, you can configure the time the router must wait before receiving the next state-refresh message. If a new state-refresh message is received within the waiting time, the router will discard it; if this timer times out, the router will accept a new state-refresh message, refresh its own PIM-DM state, and reset the waiting timer.

The TTL value of a state-refresh message decrements by 1 whenever it passes a router before it is forwarded to the downstream node until the TTL value comes down to 0. In a small network, a state-refresh message may cycle in the network. To effectively control the propagation scope of state-refresh messages, you need to configure an appropriate TTL value based on the network size.

It is recommended to perform the following configurations on all routers in the PIM domain.

Follow these steps to configure state-refresh parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the interval between state-refresh messages	**state-refresh-interval** *interval*	Optional 60 seconds by default
Configure the time to wait before receiving a new state-refresh message	**state-refresh-rate-limit** *interval*	Optional 30 seconds by default
Configure the TTL value of state-refresh messages	**state-refresh-ttl** *ttl-value*	Optional 255 by default

Configuring PIM-DM Graft Retry Period

In PIM-DM, graft is the only type of message that uses the acknowledgment mechanism. In a PIM-DM domain, if a router does not receive a graft-ack message from the upstream router within the specified time after it sends a graft message, the router keeps sending new graft messages at a configurable interval, namely graft retry period, until it receives a graft-ack from the upstream router.

Follow these steps to configure graft retry period:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure graft retry period	**pim timer graft-retry** *interval*	Optional 3 seconds by default

Note

For the configuration of other timers in PIM-DM, refer to Configuring PIM Common Timers.

Configuring PIM-SM

PIM-SM Configuration Task List

Complete these tasks to configure PIM-SM:

Task		Remarks
Configuring PIM-SM		Required
Configuring an RP	Configuring a static RP	Optional
	Configuring a C-RP	Optional
	Enabling auto-RP	Optional
	Configuring C-RP timers globally	Optional
Configuring a BSR	Configuring a C-BSR	Optional
	Configuring a PIM domain border	Optional
	Configuring global C-BSR parameters	Optional
	Configuring C-BSR timers	Optional
Configuring Administrative Scoping	Enabling administrative scoping	Optional
	Configuring an admin-scope zone boundary	Optional
	Configuring a C-BSR for each zone	Optional
Configuring Multicast Source Registration		Optional
Disabling SPT Switchover		Optional
Configuring PIM Common Features		Optional

Configuration Prerequisites

Before configuring PIM-SM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-SM, prepare the following data:

- The IP address of a static RP and an ACL rule defining the range of multicast groups to be served by the static RP
- C-RP priority and an ACL rule defining the range of multicast groups to be served by each C-RP
- A legal C-RP address range and an ACL rule defining the range of multicast groups to be served
- C-RP-Adv interval
- C-RP timeout
- C-BSR priority
- Hash mask length
- An ACL rule defining a legal BSR address range
- BS period
- BS timeout
- An ACL rule for register message filtering
- Register suppression time
- Register probe time
- The ACL rule and sequencing rule for SPT switchover

Enabling PIM-SM

With PIM-SM enabled, a router sends hello messages periodically to discover PIM neighbors and processes messages from the PIM neighbors. When deploying a PIM-SM domain, you are recommended to enable PIM-SM on all non-border interfaces of the routers.

Enabling PIM-SM globally

Follow these steps to enable PIM-SM:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable PIM-SM	**pim sm**	Required Disabled by default

Caution

All the interfaces on the same router must work in the same PIM mode.

Note

For details about the **multicast routing-enable** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring an RP

An RP can be manually configured or dynamically elected through the BSR mechanism. For a large PIM network, static RP configuration is a tedious job. Generally, static RP configuration is just a backup means for the dynamic RP election mechanism to enhance the robustness and operation manageability of a multicast network.

Configuring a static RP

If there is only one dynamic RP in a network, manually configuring a static RP can avoid communication interruption due to single-point failures and avoid frequent message exchange between C-RPs and the BSR.

Perform this configuration on all the routers in the PIM-SM domain.

Follow these steps to configure a static RP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure a static RP	**static-rp** *rp-address* [*acl-number*] [**preferred**]	Required No static RP by default

Caution

To enable a static RP to work normally, you must perform this configuration on all the routers in the PIM-SM domain and specify the same RP address.

Configuring a C-RP

In a PIM-SM domain, you can configure routers that intend to become the RP as C-RPs. The BSR collects the C-RP information by receiving the C-RP-Adv messages from C-RPs or auto-RP announcements from other routers and organizes the information into an RP-set, which is flooded throughout the entire network. Then, the other routers in the network calculate the mappings between specific group ranges and the corresponding RPs based on the RP-set. We recommend that you configure C-RPs on backbone routers.

To guard against C-RP spoofing, you need to configure a legal C-RP address range and the range of multicast groups to be served on the BSR. In addition, because every C-BSR has a chance to become the BSR, you need to configure the same filtering policy on all C-BSRs in the PIM-SM domain.

Follow these steps to configure a C-RP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure an interface to be a C-RP	**c-rp** *interface-type interface-number* [**group-policy** *acl-number* | **priority** *priority* | **holdtime** *hold-interval* | **advertisement-interval** *adv-interval*] *	Required No C-RPs are configured by default
Configure a legal C-RP address range and the range of multicast groups to be served	**crp-policy** *acl-number*	Optional No restrictions by default

Note

- When configuring a C-RP, ensure a relatively large bandwidth between this C-RP and the other devices in the PIM-SM domain.
- An RP can serve multiple multicast groups or all multicast groups. Only one RP can forward multicast traffic for a multicast group at a moment.

Enabling auto-RP

Auto-RP announcement and discovery messages are addressed to the multicast group addresses 224.0.1.39 and 224.0.1.40 respectively. With auto-RP enabled on a device, the device can receive these two types of messages and record the RP information carried in such messages.

Follow these steps to enable auto-RP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Enable auto-RP	**auto-rp enable**	Required Disabled by default

Configuring C-RP timers globally

To enable the BSR to distribute the RP-set information within the PIM-SM domain, C-RPs must periodically send C-RP-Adv messages to the BSR. The BSR learns the RP-set information from the received messages, and encapsulates its own IP address together with the RP-set information in its bootstrap messages. The BSR then floods the bootstrap messages to all PIM routers (224.0.0.13) in the network.

Each C-RP encapsulates a timeout value in its C-RP-Adv messages. Upon receiving a C_RP-Adv message, the BSR obtains this timeout value and starts a C-RP timeout timer. If the BSR fails to hear a subsequent C-RP-Adv message from the C-RP when the timer times out, the BSR assumes the C-RP to have expired or become unreachable.

The C-RP timers need to be configured on C-RP routers.

Follow these steps to configure C-RP timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the C-RP-Adv interval	**c-rp advertisement-interval** *interval*	Optional 60 seconds by default
Configure C-RP timeout time	**c-rp holdtime** *interval*	Optional 150 seconds by default

Note

For the configuration of other timers in PIM-SM, refer to Configuring PIM Common Timers.

Configuring a BSR

A PIM-SM domain can have only one BSR, but must have at least one C-BSR. Any router can be configured as a C-BSR. Elected from C-BSRs, the BSR is responsible for collecting and advertising RP information in the PIM-SM domain.

Configuring a C-BSR

C-BSRs should be configured on routers in the backbone network. When configuring a router as a C-BSR, be sure to specify a PIM-SM-enabled interface on the router. The BSR election process is summarized as follows:

- Initially, every C-BSR assumes itself to be the BSR of this PIM-SM domain, and uses its interface IP address as the BSR address to send bootstrap messages.
- When a C-BSR receives the bootstrap message of another C-BSR, it first compares its own priority with the other C-BSR's priority carried in message. The C-BSR with a higher priority wins. If there is a tie in the priority, the C-BSR with a higher IP address wins. The loser uses the winner's BSR address to replace its own BSR address and no longer assumes itself to be the BSR, while the winner retains its own BSR address and continues assuming itself to be the BSR.

Configuring a legal range of BSR addresses enables filtering of bootstrap messages based on the address range, thus to prevent a maliciously configured host from masquerading as a BSR. The same configuration needs to be made on all routers in the PIM-SM domain. The following are typical BSR spoofing cases and the corresponding preventive measures:

1) Some maliciously configured hosts can forge bootstrap messages to fool routers and change RP mappings. Such attacks often occur on border routers. Because a BSR is inside the network whereas hosts are outside the network, you can protect a BSR against attacks from external hosts by enabling the border routers to perform neighbor checks and RPF checks on bootstrap messages and discard unwanted messages.
2) When a router in the network is controlled by an attacker or when an illegal router is present in the network, the attacker can configure this router as a C-BSR and make it win BSR election to control the right of advertising RP information in the network. After being configured as a C-BSR, a router automatically floods the network with bootstrap messages. As a bootstrap message has a TTL

value of 1, the whole network will not be affected as long as the neighbor router discards these bootstrap messages. Therefore, with a legal BSR address range configured on all routers in the entire network, all these routers will discard bootstrap messages from out of the legal address range.

The above-mentioned preventive measures can partially protect the security of BSRs in a network. However, if a legal BSR is controlled by an attacker, the above-mentioned problem will still occur.

Follow these steps to configure a C-BSR:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure an interface as a C-BSR	**c-bsr** *interface-type interface-number* [*hash-length* [*priority*]]	Required No C-BSRs are configured by default.
Configure a legal BSR address range	**bsr-policy** *acl-number*	Optional No restrictions on BSR address range by default

Note

- Since a large amount of information needs to be exchanged between a BSR and the other devices in the PIM-SM domain, a relatively large bandwidth should be provided between the C-BSRs and the other devices in the PIM-SM domain.
- For C-BSRs interconnected via a Generic Routing Encapsulation (GRE) tunnel, multicast static routes need to be configured to ensure that the next hop to a C-BSR is a GRE interface. For more information about multicast static routes, refer to Multicast Routing and Forwarding Configuration in the IP Multicast Volume.

Configuring a PIM domain border

As the administrative core of a PIM-SM domain, the BSR sends the collected RP-Set information in the form of bootstrap messages to all routers in the PIM-SM domain.

A PIM domain border is a bootstrap message boundary. Each BSR has its specific service scope. A number of PIM domain border interfaces partition a network into different PIM-SM domains. Bootstrap messages cannot cross a domain border in either direction

Perform the following configuration on routers that can become a PIM domain border.

Follow these steps to configure a PIM domain border:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Configure a PIM domain border	**pim bsr-boundary**	Required By default, no PIM domain border is configured.

Configuring global C-BSR parameters

In each PIM-SM domain, a unique BSR is elected from C-BSRs. The C-RPs in the PIM-SM domain send advertisement messages to the BSR. The BSR summarizes the advertisement messages to form an RP-set and advertises it to all routers in the PIM-SM domain. All the routers use the same Hash algorithm to get the RP address corresponding to specific multicast groups.

Perform the following configuration on C-BSR routers.

Follow these steps to configure C-BSR parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the Hash mask length	**c-bsr hash-length** *hash-length*	Optional 30 by default
Configure the C-BSR priority	**c-bsr priority** *priority*	Optional By default, the C-BSR priority is 0.

Note

About the Hash mask length and C-BSR priority:

- You can configure these parameters at three levels: global configuration level, global scope zone level, and admin-scope zone level.
- The value of these parameters configured at the global scope zone level or admin-scope zone level have preference over the global values.
- If you do not configure these parameters at the global scope zone level or admin-scope zone level, the corresponding global values will be used.

For configuration of C-BSR parameters for an admin-scope zone and global scope zone, see Configuring a C-BSR for each zone.

Configuring C-BSR timers

The BSR election winner multicasts its own IP address and RP-Set information through bootstrap messages within the entire zone it serves. The BSR floods bootstrap messages throughout the network at the interval of BS (BSR state) period. Any C-BSR that receives a bootstrap message retains the RP-set for the length of BS timeout, during which no BSR election takes place. If the BSR state times out and no bootstrap message is received from the BSR, a new BSR election process is triggered among the C-BSRs.

Perform the following configuration on C-BSR routers.

Follow these steps to configure C-BSR timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the BS period	**c-bsr interval** *interval*	Optional For the default value, see the note below.
Configure the BS timeout	**c-bsr holdtime** *interval*	Optional For the default value, see the note below.

Note

About the BS period:

- By default, the BS period is determined by this formula: BS period = (BS timeout – 10) / 2. The default BS timeout is 130 seconds, so the default BS period = (130 – 10) / 2 = 60 (seconds).
- If this parameter is manually configured, the system will use the configured value.

About the BS timeout:

- By default, the BS timeout value is determined by this formula: BS timeout = BS period × 2 + 10. The default BS period is 60 seconds, so the default BS timeout = 60 × 2 + 10 = 130 (seconds).
- If this parameter is manually configured, the system will use the configured value.

Caution

In configuration, make sure that the BS period value is smaller than the BS timeout value.

Configuring Administrative Scoping

With administrative scoping disabled, a PIM-SM domain has only one BSR. The BSR manages the whole network. To manage your network more effectively and specifically, you can partition the PIM-SM domain into multiple admin-scope zones. Each admin-scope zone maintains a BSR, which serves a specific multicast group range; while the global scope zone also maintains a BSR, which serves all the rest multicast groups.

Enabling administrative scoping

Before configuring an admin-scope zone, you must enable administrative scoping first.

Perform the following configuration on all routers in the PIM-SM domain.

Follow these steps to enable administrative scoping:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—

To do…	Use the command…	Remarks
Enable administrative scoping	**c-bsr admin-scope**	Required Disabled by default

Configuring an admin-scope zone boundary

The boundary of each admin-scope zone is formed by ZBRs. Each admin-scope zone maintains a BSR, which serves a specific multicast group range. Multicast protocol packets (such as assert messages and bootstrap messages) that belong to this range cannot cross the admin-scope zone boundary.

Perform the following configuration on routers that can become a ZBR.

Follow these steps to configure an admin-scope zone boundary:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a multicast forwarding boundary	**multicast boundary** *group-address* { *mask* | *mask-length* }	Required By default, no multicast forwarding boundary is configured.

Note

For details about the **multicast boundary** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring a C-BSR for each zone

In a network with administrative scoping enabled, group-range-specific BSRs are elected from C-BSRs. C-RPs in the network send advertisement messages to the specific BSR. The BSR summarizes the advertisement messages to form an RP-set and advertises it to all routers in the specific admin-scope zone. All the routers use the same Hash algorithm to get the RP address corresponding to the specific multicast group.

Perform the following configuration on the routers that may become C-BSRs in each admin-scope zone or the global scope zone.

Follow these steps to configure a C-BSR for each zone:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure a C-BSR for an admin-scope zone	**c-bsr group** *group-address* { *mask* | *mask-length* } [**hash-length** *hash-length* | **priority** *priority*] *	Required No C-BSRs are configured for a admin-scope zone by default

To do…	Use the command…	Remarks
Configure a C-BSR for the global-scope zone	**c-bsr global** [**hash-length** *hash-length* | **priority** *priority*] *	Required No C-BSRs are configured for the global-scope zone by default

Note

About the Hash mask length and C-BSR priority:

- You can configure these parameters at three levels: global configuration level, global scope zone level, and admin-scope zone level.
- The value of these parameters configured at the global scope zone level or admin-scope zone level have preference over the global values.
- If you do not configure these parameters at the global scope zone level or admin-scope zone level, the corresponding global values will be used.

For configuration of global C-BSR parameters, see Configuring global C-BSR parameters.

Configuring Multicast Source Registration

Within a PIM-SM domain, the source-side DR sends register messages to the RP, and these register messages have different multicast source or group addresses. You can configure a filtering rule to filter register messages so that the RP can serve specific multicast groups. If an (S, G) entry is denied by the filtering rule, or the action for this entry is not defined in the filtering rule, the RP will send a register-stop message to the DR to stop the registration process for the multicast data.

In view of information integrity of register messages in the transmission process, you can configure the device to calculate the checksum based on the entire register messages. However, to reduce the workload of encapsulating data in register messages and for the sake of interoperability, this method of checksum calculation is not recommended.

When receivers stop receiving multicast data addressed to a certain multicast group through the RP (that is, the RP stops serving the receivers of that multicast group), or when the RP formally starts receiving multicast data from the multicast source, the RP sends a register-stop message to the source-side DR. Upon receiving this message, the DR stops sending register messages encapsulated with multicast data and starts a register-stop timer. When the register-stop timer expires, the DR sends a null register message (a register message without encapsulated multicast data) to the RP. If the DR receives a register-stop message during the register probe time, it will reset its register-stop timer; otherwise, the DR starts sending register messages with encapsulated data again when the register-stop timer expires.

The register-stop timer is set to a random value chosen uniformly from the interval (0.5 times register_suppression_time, 1.5 times register_suppression_time) minus register_probe_time.

Configure a filtering rule for register messages on all C-RP routers and configure them to calculate the checksum based on the entire register messages. Configure the register suppression time and the register probe time on all routers that may become source-side DRs.

Follow these steps to configure register-related parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure a filtering rule for register messages	**register-policy** *acl-number*	Optional No register filtering rule by default
Configure the device to calculate the checksum based on the entire register messages	**register-whole-checksum**	Optional By default, the checksum is calculated based on the header of register messages
Configure the register suppression time	**register-suppression-timeout** *interval*	Optional 60 seconds by default
Configure the register probe time	**probe-interval** *interval*	Optional 5 seconds by default

Disabling SPT Switchover

If an 3Com Switch 4800G acts as an RP or the receiver-side DR, it initiates an SPT switchover process (by default) upon receiving the first multicast packet along the RPT. You can disable the switchover from RPT to SPT.

Perform the following operations to disable the SPT switchover:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Apply the ACL for initiating an SPT switchover	**spt-switch-threshold infinity** [**group-policy** *acl-number* [**order** *order-value*]]	Optional By default, the device switches to the SPT immediately after it receives the first multicast packet.

Configuring PIM-SSM

Note

The PIM-SSM model needs the support of IGMPv3. Therefore, be sure to enable IGMPv3 on PIM routers with multicast receivers.

PIM-SSM Configuration Task List

Complete these tasks to configure PIM-SSM:

Task	Remarks
Enabling PIM-SM	Required
Configuring the SSM Group Range	Optional
Configuring PIM Common Features	Optional

Configuration Prerequisites

Before configuring PIM-SSM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-SSM, prepare the following data:

- The SSM group range

Enabling PIM-SM

The SSM model is implemented based on some subsets of PIM-SM. Therefore, a router is PIM-SSM capable after you enable PIM-SM on it.

When deploying a PIM-SM domain, you are recommended to enable PIM-SM on non-border interfaces of the routers.

Follow these steps to enable PIM-SM globally

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable PIM-SM	**pim sm**	Required Disabled by default

Caution

All the interfaces on the same device must work in the same PIM mode.

Note

For details about the **multicast routing-enable** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring the SSM Group Range

As for whether the information from a multicast source is delivered to the receivers based on the PIM-SSM model or the PIM-SM model, this depends on whether the group address in the (S, G) channel subscribed by the receivers falls in the SSM group range. All PIM-SM-enabled interfaces assume that multicast groups within this address range are using the PIM-SSM model.

Perform the following configuration on all routers in the PIM-SM domain.

Follow these steps to configure an SSM multicast group range:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the SSM group range	**ssm-policy** *acl-number*	Optional 232.0.0.0/8 by default

Caution

- Make sure that the same SSM group range is configured on all routers in the entire domain. Otherwise, multicast information cannot be delivered through the SSM model.
- When a member of a multicast group in the SSM group range sends an IGMPv1 or IGMPv2 report message, the device does not trigger a (*, G) join.

Configuring PIM Common Features

Note

For the functions or parameters that can be configured in both PIM view and interface view described in this section:

- Configurations performed in PIM view are effective to all interfaces, while configurations performed in interface view are effective to the current interface only.
- If the same function or parameter is configured in both PIM view and interface view, the configuration made in interface view has preference over the configuration made in PIM view, regardless of the configuration sequence.

PIM Common Feature Configuration Task List

Complete these tasks to configure PIM common features:

Task	Remarks
Configuring a Multicast Data Filter	Optional
Configuring a Hello Message Filter	Optional

Task	Remarks
Configuring PIM Hello Options	Optional
Configuring PIM Common Timers	Optional
Configuring Join/Prune Message Sizes	Optional

Configuration Prerequisites

Before configuring PIM common features, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure PIM-DM, or PIM-SM, or PIM-SSM.

Before configuring PIM common features, prepare the following data:

- An ACL rule for filtering multicast data
- An ACL rule defining a legal source address range for hello messages
- Priority for DR election (global value/interface level value)
- PIM neighbor timeout time (global value/interface value)
- Prune delay (global value/interface level value)
- Prune override interval (global value/interface level value)
- Hello interval (global value/interface level value)
- Maximum delay between hello message (interface level value)
- Assert timeout time (global value/interface value)
- Join/prune interval (global value/interface level value)
- Join/prune timeout (global value/interface value)
- Multicast source lifetime
- Maximum size of join/prune messages
- Maximum number of (S, G) entries in a join/prune message

Configuring a Multicast Data Filter

No matter in a PIM-DM domain or a PIM-SM domain, routers can check passing-by multicast data based on the configured filtering rules and determine whether to continue forwarding the multicast data. In other words, PIM routers can act as multicast data filters. These filters can help implement traffic control on one hand, and control the information available to receivers downstream to enhance data security on the other hand.

Follow these steps to configure a multicast data filter:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure a multicast group filter	**source-policy** *acl-number*	Required No multicast data filter by default

Note

- Generally, a smaller distance from the filter to the multicast source results in a more remarkable filtering effect.
- This filter works not only on independent multicast data but also on multicast data encapsulated in register messages.

Configuring a Hello Message Filter

Along with the wide applications of PIM, the security requirement for the protocol is becoming more and more demanding. The establishment of correct PIM neighboring relationships is the prerequisite for secure application of PIM. You can configure a legal source address range for hello messages on interfaces of routers to ensure the correct PIM neighboring relationships, and thus to guard against PIM message attacks.

Follow these steps to configure a hello message filter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a hello message filter	**pim neighbor-policy** *acl-number*	Required No hello message filter by default.

Note

With the hello message filter configured, if hello messages of an existing PIM neighbor fail to pass the filter, the PIM neighbor will be removed automatically when it times out.

Configuring PIM Hello Options

No matter in a PIM-DM domain or a PIM-SM domain, the hello messages sent among routers contain many configurable options, including:

- DR_Priority (for PIM-SM only): priority for DR election. The device with the highest priority wins the DR election. You can configure this parameter on all the routers in a multi-access network directly connected to multicast sources or receivers.
- Holdtime: the timeout time of PIM neighbor reachability state. When this timer times out, if the router has received no hello message from a neighbor, it assumes that this neighbor has expired or become unreachable.
- LAN_Prune_Delay: the delay of prune messages on a multi-access network. This option consists of LAN-delay (namely, prune delay), override-interval, and neighbor tracking flag. If the LAN-delay or override-interval values of different PIM routers on a multi-access subnet are different, the

largest value will take effect. If you want to enable neighbor tracking, the neighbor tracking feature should be enabled on all PIM routers on a multi-access subnet.

The LAN-delay setting will cause the upstream routers to delay processing received prune messages. If the LAN-delay setting is too small, it may cause the upstream router to stop forwarding multicast packets before a downstream router sends a prune override message. Therefore, be cautious when configuring this parameter.

The override-interval sets the length of time a downstream router is allowed to wait before sending a prune override message. When a router receives a prune message from a downstream router, it does not perform the prune action immediately; instead, it maintains the current forwarding state for a period of LAN-delay plus override-interval. If the downstream router needs to continue receiving multicast data, it must send a prune override message within the prune override interval; otherwise, the upstream route will perform the prune action when the period of LAN-delay plus override-interval time out.

A hello message sent from a PIM router contains a generation ID option. The generation ID is a random value for the interface on which the hello message is sent. Normally, the generation ID of a PIM router does not change unless the status of the router changes (for example, when PIM is just enabled on the interface or the device is restarted). When the router starts or restarts sending hello messages, it generates a new generation ID. If a PIM router finds that the generation ID in a hello message from the upstream router has changed, it assumes that the status of the upstream neighbor is lost or the upstream neighbor has changed. In this case, it triggers a join message for state update.

If you disable join suppression (namely, enable neighbor tracking), the join suppression feature should be disabled on all PIM routers on a multi-access subnet; otherwise, the upstream router will fail to explicitly track which downstream routers are joined to it.

Configuring hello options globally

Follow these steps to configure hello options globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the priority for DR election	**hello-option dr-priority** *priority*	Optional 1 by default
Configure PIM neighbor timeout time	**hello-option holdtime** *interval*	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	**hello-option lan-delay** *interval*	Optional 500 milliseconds by default
Configure the prune override interval	**hello-option override-interval** *interval*	Optional 2,500 milliseconds by default
Disable join suppression	**hello-option neighbor-tracking**	Required Enabled by default

Configuring hello options on an interface

Follow these steps to configure hello options on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the priority for DR election	**pim hello-option dr-priority** *priority*	Optional 1 by default
Configure PIM neighbor timeout time	**pim hello-option holdtime** *interval*	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	**pim hello-option lan-delay** *interval*	Optional 500 milliseconds by default
Configure the prune override interval	**pim hello-option override-interval** *interval*	Optional 2,500 milliseconds by default
Disable join suppression	**pim hello-option neighbor-tracking**	Required Enabled by default
Configure the interface to reject hello messages without a generation ID	**pim require-genid**	Required By default, hello messages without Generation_ID are accepted

Configuring PIM Common Timers

PIM routers discover PIM neighbors and maintain PIM neighboring relationships with other routers by periodically sending out hello messages.

Upon receiving a hello message, a PIM router waits a random period, which is smaller than the maximum delay between hello messages, before sending out a hello message. This avoids collisions that occur when multiple PIM routers send hello messages simultaneously.

A PIM router periodically sends join/prune messages to its upstream for state update. A join/prune message contains the join/prune timeout time. The upstream router sets a join/prune timeout timer for each pruned downstream interface.

Any router that has lost assert election will prune its downstream interface and maintain the assert state for a period of time. When the assert state times out, the assert losers will resume multicast forwarding.

When a router fails to receive subsequent multicast data from multicast source S, the router does not immediately delete the corresponding (S, G) entry; instead, it maintains the (S, G) entry for a period of time, namely the multicast source lifetime, before deleting the (S, G) entry.

Configuring PIM common timers globally

Follow these steps to configure PIM common timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the hello interval	**timer hello** *interval*	Optional 30 seconds by default

To do...	Use the command...	Remarks
Configure the join/prune interval	**timer join-prune** *interval*	Optional 60 seconds by default
Configure the join/prune timeout time	**holdtime join-prune** *interval*	Optional 210 seconds by default
Configure assert timeout time	**holdtime assert** *interval*	Optional 180 seconds by default
Configure the multicast source lifetime	**source-lifetime** *interval*	Optional 210 seconds by default

Configuring PIM common timers on an interface

Follow these steps to configure PIM common timers on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the hello interval	**pim timer hello** *interval*	Optional 30 seconds by default
Configure the maximum delay between hello messages	**pim triggered-hello-delay** *interval*	Optional 5 seconds by default
Configure the join/prune interval	**pim timer join-prune** *interval*	Optional 60 seconds by default
Configure the join/prune timeout time	**pim holdtime join-prune** *interval*	Optional 210 seconds by default
Configure assert timeout time	**pim holdtime assert** *interval*	Optional 180 seconds by default

Note

If there are no special networking requirements, we recommend that you use the default settings.

Configuring Join/Prune Message Sizes

A larger join/prune message size will result in loss of a larger amount of information when a message is lost; with a reduced join/message size, the loss of a single message will bring relatively minor impact.

By controlling the maximum number of (S, G) entries in a join/prune message, you can effectively reduce the number of (S, G) entries sent per unit of time.

Follow these steps to configure join/prune message sizes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter PIM view	**pim**	—
Configure the maximum size of a join/prune message	**jp-pkt-size** *packet-size*	Optional 8,100 bytes by default
Configure the maximum number of (S, G) entries in a join/prune message	**jp-queue-size** *queue-size*	Optional 1,020 by default

Displaying and Maintaining PIM

To do...	Use the command...	Remarks
View the BSR information in the PIM-SM domain and locally configured C-RP information in effect	**display pim bsr-info**	Available in any view
View the information of unicast routes used by PIM	**display pim claimed-route** [*source-address*]	Available in any view
View the number of PIM control messages	**display pim control-message counters** [**message-type** { **probe** | **register** | **register-stop** } | [**interface** *interface-type interface-number* | **message-type** { **assert** | **bsr** | **crp** | **graft** | **graft-ack** | **hello** | **join-prune** | **state-refresh** }] *]	Available in any view
View the information about unacknowledged graft messages	**display pim grafts**	Available in any view
View the PIM information on an interface or all interfaces	**display pim interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View the information of join/prune messages to send	**display pim join-prune mode** { **sm** [**flags** *flag-value*] | **ssm** } [**interface** *interface-type interface-number* | **neighbor** *neighbor-address*] * [**verbose**]	Available in any view
View PIM neighboring information	**display pim neighbor** [**interface** *interface-type interface-number* | *neighbor-address* | **verbose**] *	Available in any view
View the content of the PIM routing table	**display pim routing-table** [*group-address* [**mask** { *mask-length* | *mask* }] | *source-address* [**mask** { *mask-length* | *mask* }] | **incoming-interface** [*interface-type interface-number* | **register**] | **outgoing-interface** { **include** | **exclude** | **match** } { *interface-type interface-number* | **register** } | **mode** *mode-type* | **flags** *flag-value* | **fsm**] *	Available in any view
View the RP information	**display pim rp-info** [*group-address*]	Available in any view
Reset PIM control message counters	**reset pim control-message counters** [**interface** *interface-type interface-number*]	Available in user view

PIM Configuration Examples

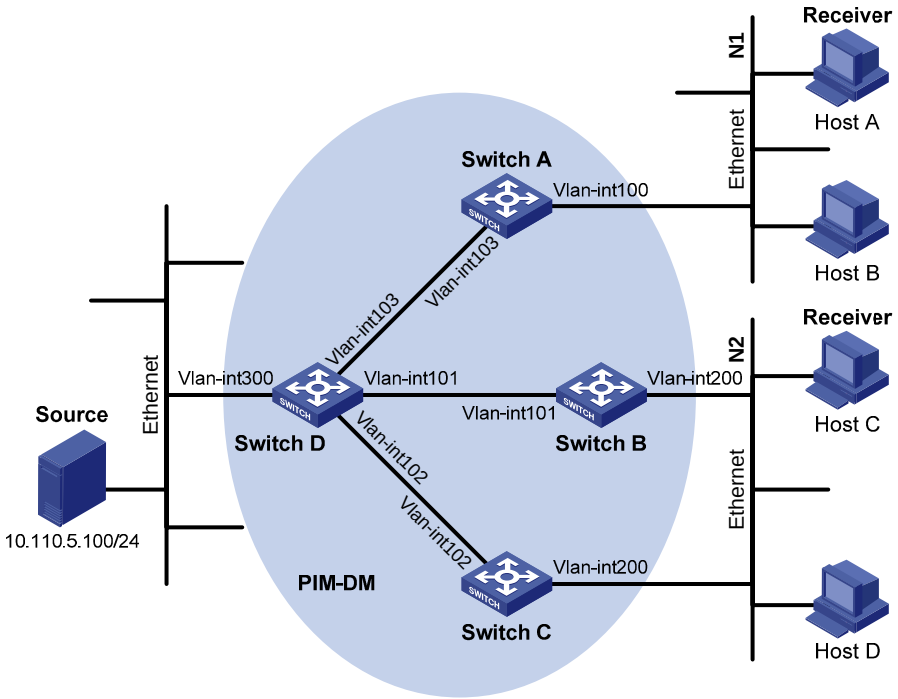
PIM-DM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the dense mode.
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D through VLAN-interface 103.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch D through VLAN-interface 101 and VLAN-interface 102 respectively.
- IGMPv2 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 1-10 Network diagram for PIM-DM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int103	192.168.1.1/24		Vlan-int103	192.168.1.2/24
Switch B	Vlan-int200	10.110.2.1/24		Vlan-int101	192.168.2.2/24
	Vlan-int101	192.168.2.1/24		Vlan-int102	192.168.3.2/24
Switch C	Vlan-int200	10.110.2.2/24			
	Vlan-int102	192.168.3.1/24			

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-10. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation among the switches in the PIM-DM domain. Ensure the network-layer interoperation in the PIM-DM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-DM and IGMP

Enable IP multicast routing on Switch A, enable PIM-DM on each interface, and enable IGMP on VLAN-interface 100, which connects Switch A to the stub network.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim dm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
[SwitchA-Vlan-interface103] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A.

Enable IP multicast routing on Switch D, and enable PIM-DM on each interface.

```
<SwitchD> system-view
[SwitchD] multicast routing-enable
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] pim dm
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim dm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 101
[SwitchD-Vlan-interface101] pim dm
[SwitchD-Vlan-interface101] quit
[SwitchD] interface vlan-interface 102
[SwitchD-Vlan-interface102] pim dm
[SwitchD-Vlan-interface102] quit
```

3) Verify the configuration

Use the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch D.

```
[SwitchD] display pim interface
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
 Vlan300             0      30         1          10.110.5.1     (local)
 Vlan103             1      30         1          192.168.1.2    (local)
 Vlan101             1      30         1          192.168.2.2    (local)
 Vlan102             1      30         1          192.168.3.2    (local)
```

Carry out the **display pim neighbor** command to view the PIM neighboring relationships among the switches. For example:

View the PIM neighboring relationships on Switch D.

```
[SwitchD] display pim neighbor
 Total Number of Neighbors = 3

 Neighbor        Interface           Uptime   Expires  Dr-Priority
 192.168.1.1     Vlan103             00:02:22 00:01:27 1
 192.168.2.1     Vlan101             00:00:22 00:01:29 3
 192.168.3.1     Vlan102             00:00:23 00:01:31 5
```

Assume that Host A needs to receive the information addressed to multicast group G (225.1.1.1). After multicast source S (10.110.5.100/24) sends multicast packets to the multicast group G, an SPT is established through traffic flooding. Switches on the SPT path (Switch A and Switch D) have their (S, G) entries. Host A sends an IGMP report to Switch A to join the multicast group G, and a (*, G) entry is generated on Switch A. You can use the **display pim routing-table** command to view the PIM routing table information on each switch. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:04:25
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: igmp, UpTime: 00:04:25, Expires: never

 (10.110.5.100, 225.1.1.1)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:06:14
     Upstream interface: Vlan-interface103,
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: pim-dm, UpTime: 00:04:25, Expires: never
```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.1)
     Protocol: pim-dm, Flag: LOC ACT
```

```
UpTime: 00:03:27
Upstream interface: Vlan-interface300
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
Downstream interface(s) information:
Total number of downstreams: 3
    1: Vlan-interface103
        Protocol: pim-dm, UpTime: 00:03:27, Expires: never
    2: Vlan-interface101
        Protocol: pim-dm, UpTime: 00:03:27, Expires: never
    3: Vlan-interface102
        Protocol: pim-dm, UpTime: 00:03:27, Expires: never
```

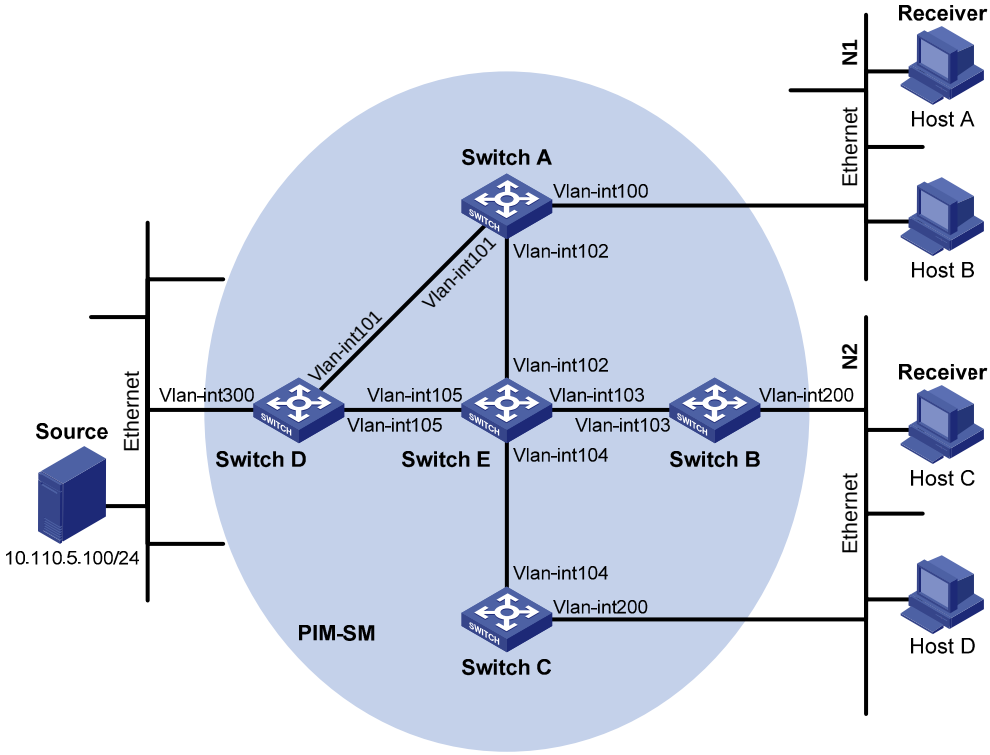
PIM-SM Non-Scoped Zone Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM-SM domain contains only one BSR.
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Vlan-interface 105 on Switch D and Vlan-interface 102 on Switch E act as C-BSRs and C-RPs; the C-BSR on Switch E has a higher priority; the multicast group range served by the C-RP is 225.1.1.0/24; modify the hash mask length to map a certain number of consecutive group addresses within the range to the two C-RPs.
- IGMPv2 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 1-11 Network diagram for PIM-SM non- scoped zone configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int102	192.168.9.1/24		Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
	Vlan-int103	192.168.2.1/24		Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24		Vlan-int102	192.168.9.2/24
	Vlan-int104	192.168.3.1/24		Vlan-int105	192.168.4.1/24

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-11. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation in the PIM-SM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-SM and IGMP

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMP on VLAN-interface 100, which connects Switch A to the stub network.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
```

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable IGMP on the corresponding interfaces on these two switches.

3) Configure a C-BSR and a C-RP

On Switch D, configure the service scope of RP, specify a C-BSR and a C-RP, and set the hash mask length to 32 and the priority of the C-BSR to 10.

```
<SwitchD> system-view
[SwitchD] acl number 2005
[SwitchD-acl-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[SwitchD-acl-basic-2005] quit
[SwitchD] pim
[SwitchD-pim] c-bsr vlan-interface 105 32 10
[SwitchD-pim] c-rp vlan-interface 105 group-policy 2005
[SwitchD-pim] quit
```

On Switch E, configure the service scope of RP advertisements, specify a C-BSR and a C-RP; and set the hash mask length to 32 and the priority of the C-BSR to 20.

```
<SwitchE> system-view
[SwitchE] acl number 2005
[SwitchE-acl-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[SwitchE-acl-basic-2005] quit
[SwitchE] pim
[SwitchE-pim] c-bsr vlan-interface 102 32 20
[SwitchE-pim] c-rp vlan-interface 102 group-policy 2005
[SwitchE-pim] quit
```

4) Verify the configuration

Carry out the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch A.

```
[SwitchA] display pim interface
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
 Vlan100             0      30         1          10.110.1.1     (local)
 Vlan101             1      30         1          192.168.1.2
 Vlan102             1      30         1          192.168.9.2
```

To view the BSR election information and the locally configured C-RP information in effect on a switch, use the **display pim bsr-info** command. For example:

View the BSR information and the locally configured C-RP information in effect on Switch A.

```
[SwitchA] display pim bsr-info
 Elected BSR Address: 192.168.9.2
     Priority: 20
```

```
     Hash mask length: 32
     State: Accept Preferred
     Scope: Not scoped
     Uptime: 00:40:40
     Expires: 00:01:42
```

View the BSR information and the locally configured C-RP information in effect on Switch D.

```
[SwitchD] display pim bsr-info
 Elected BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Accept Preferred
     Scope: Not scoped
     Uptime: 00:05:26
     Expires: 00:01:45
 Candidate BSR Address: 192.168.4.2
     Priority: 10
     Hash mask length: 32
     State: Candidate
     Scope: Not scoped

 Candidate RP: 192.168.4.2(Vlan-interface105)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:34
```

View the BSR information and the locally configured C-RP information in effect on Switch E.

```
[SwitchE] display pim bsr-info
 Elected BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Elected
     Scope: Not scoped
     Uptime: 00:01:18
     Next BSR message scheduled at: 00:01:52
 Candidate BSR Address: 192.168.9.2
     Priority: 20
     Hash mask length: 32
     State: Elected
     Scope: Not scoped

Candidate RP: 192.168.9.2(Vlan-interface102)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:48
```

To view the RP information discovered on a switch, use the **display pim rp-info** command. For example:

View the RP information on Switch A.

```
[SwitchA] display pim rp-info
 PIM-SM BSR RP information:
 Group/MaskLen: 225.1.1.0/24
     RP: 192.168.4.2
     Priority: 0
     HoldTime: 150
     Uptime: 00:51:45
     Expires: 00:02:22

     RP: 192.168.9.2
     Priority: 0
     HoldTime: 150
     Uptime: 00:51:45
     Expires: 00:02:22
```

Assume that Host A needs to receive information addressed to the multicast group G (225.1.1.0). The RP corresponding to the multicast group G is Switch E as a result of hash calculation, so an RPT will be built between Switch A and Switch E. When the multicast source S (10.110.5.100/24) registers with the RP, an SPT will be built between Switch D and Switch E. Upon receiving multicast data, Switch A immediately switches from the RPT to the SPT. Switches on the RPT path (Switch A and Switch E) have a (*, G) entry, while switches on the SPT path (Switch A and Switch D) have an (S, G) entry. You can use the **display pim routing-table** command to view the PIM routing table information on the switches. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.0)
     RP: 192.168.9.2
     Protocol: pim-sm, Flag: WC
     UpTime: 00:13:46
     Upstream interface: Vlan-interface102
         Upstream neighbor: 192.168.9.2
         RPF prime neighbor: 192.168.9.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: igmp, UpTime: 00:13:46, Expires: 00:03:06

 (10.110.5.100, 225.1.1.0)
     RP: 192.168.9.2
     Protocol: pim-sm, Flag: SPT ACT
     UpTime: 00:00:42
     Upstream interface: Vlan-interface101
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
```

```
    Total number of downstreams: 1
        1: Vlan-interface100
            Protocol: pim-sm, UpTime: 00:00:42, Expires: 00:03:06
```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.0)
     RP: 192.168.9.2
     Protocol: pim-sm, Flag: SPT LOC ACT
     UpTime: 00:00:42
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface105
             Protocol: pim-sm, UpTime: 00:00:42, Expires: 00:02:26
```

View the PIM routing table information on Switch E.

```
[SwitchE] display pim routing-table
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, 225.1.1.0)
     RP: 192.168.9.2 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:13:16
     Upstream interface: Register
         Upstream neighbor: 192.168.4.2
         RPF prime neighbor: 192.168.4.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface102
             Protocol: pim-sm, UpTime: 00:13:16, Expires: 00:03:22
```

PIM-SM Admin-Scope Zone Configuration Example

Network requirements

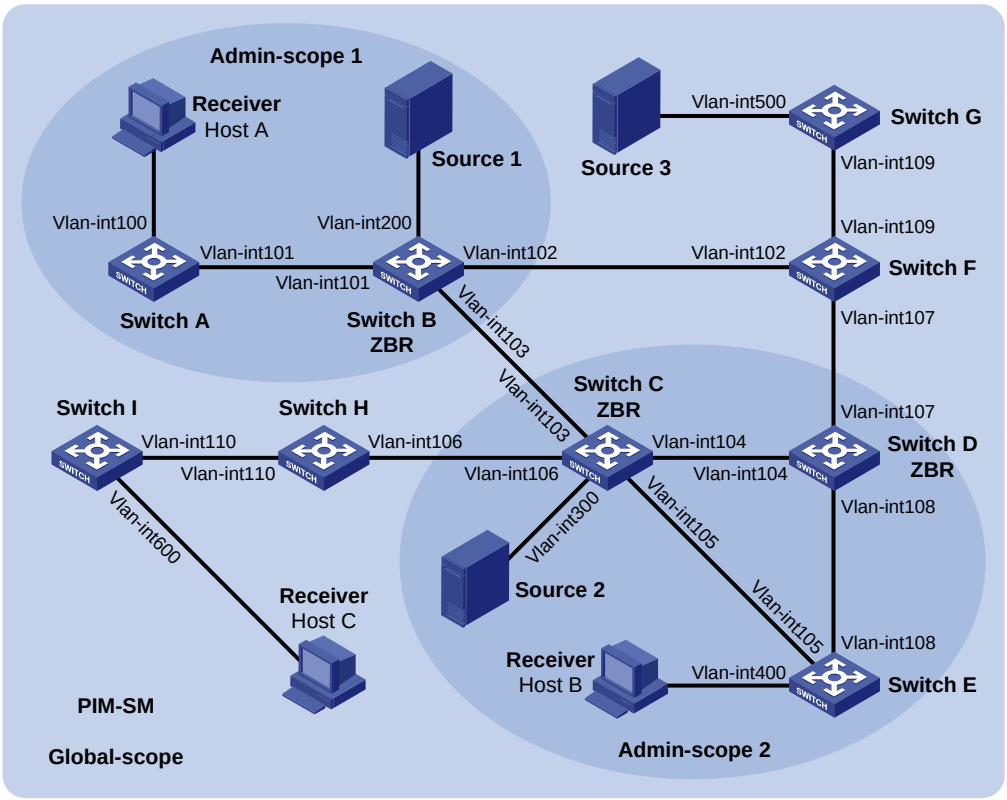
- Receivers receive VOD information through multicast. The entire PIM-SM domain is divided into admin-scope zone 1, admin-scope zone 2, and the global zone. Switch B, Switch C, and Switch D are ZBRs of these three domains respectively.
- Source 1 and Source 2 send different multicast information to multicast group 239.1.1.1. Host A receives the multicast information from only Source 1, while Host B receives the multicast information from only Source 2. Source 3 sends multicast information to multicast group 224.1.1.1. Host C is a multicast receiver for this multicast group.
- VLAN-interface 101 of Switch B acts as a C-BSR and C-RP of admin-scope zone 1, which serve the multicast group range 239.0.0.0/8. VLAN-interface 104 of Switch D acts as a C-BSR and C-RP

of admin-scope zone 2, which also serve the multicast group range 239.0.0.0/8. Both VLAN-interface 109 of Switch F and VLAN-interface 110 of Switch H act as C-BSRs and C-RPs of the global scope zone, which serve all the multicast groups other than those in the 239.0.0.0/8 range. Switch F has a higher priority.

- IGMPv2 is required between Switch A, Switch E, Switch I and their respective receivers.

Network diagram

Figure 1-12 Network diagram for PIM-SM admin-scope zone configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	192.168.1.1/24	Switch D	Vlan-int104	10.110.4.2/24
	Vlan-int101	10.110.1.1/24		Vlan-int108	10.110.7.1/24
Switch B	Vlan-int200	192.168.2.1/24		Vlan-int107	10.110.8.1/24
	Vlan-int101	10.110.1.2/24	Switch E	Vlan-int400	192.168.4.1/24
	Vlan-int103	10.110.2.1/24		Vlan-int105	10.110.5.2/24
	Vlan-int102	10.110.3.1/24		Vlan-int108	10.110.7.2/24
Switch C	Vlan-int300	192.168.3.1/24	Switch F	Vlan-int109	10.110.9.1/24
	Vlan-int104	10.110.4.1/24		Vlan-int107	10.110.8.2/24
	Vlan-int105	10.110.5.1/24		Vlan-int102	10.110.3.2/24
	Vlan-int103	10.110.2.2/24	Switch G	Vlan-int500	192.168.5.1/24
	Vlan-int106	10.110.6.1/24		Vlan-int109	10.110.9.2/24
Switch H	Vlan-int110	10.110.10.1/24	Source 1	—	192.168.2.10/24
	Vlan-int106	10.110.6.2/24	Source 2	—	192.168.3.10/24
Switch I	Vlan-int600	192.168.6.1/24	Source 3	—	192.168.5.10/24
	Vlan-int110	10.110.10.2/24			

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-12. The detailed configuration steps are omitted here.

Configure OSPF for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation among the switches in the PIM-SM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing and administrative scoping, and enable PIM-SM and IGMP

Enable IP multicast routing and administrative scoping on Switch A, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] pim
[SwitchA-pim] c-bsr admin-scope
[SwitchA-pim] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
```

The configuration on Switch E and Switch I is similar to the configuration on Switch A.

On Switch B, enable IP multicast routing and administrative scoping, and enable PIM-SM on each interface.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] pim
[SwitchB-pim] c-bsr admin-scope
[SwitchB-pim] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] pim sm
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim sm
[SwitchB-Vlan-interface102] quit
[SwitchB] interface vlan-interface 103
[SwitchB-Vlan-interface103] pim sm
[SwitchB-Vlan-interface103] quit
```

The configuration on Switch C, Switch D, Switch F, Switch G, and Switch H is similar to the configuration on Switch B. The specific configuration steps are omitted here.

3) Configure an admin-scope zone boundary

On Switch B, configure VLAN-interface 102 and VLAN-interface 103 to be the boundary of admin-scope zone 1.

```
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] multicast boundary 239.0.0.0 8
[SwitchB-Vlan-interface102] quit
[SwitchB] interface vlan-interface 103
[SwitchB-Vlan-interface103] multicast boundary 239.0.0.0 8
[SwitchB-Vlan-interface103] quit
```

On Switch C, configure VLAN-interface 103 and VLAN-interface 106 to be the boundary of admin-scope zone 2.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 103
[SwitchC-Vlan-interface103] multicast boundary 239.0.0.0 8
[SwitchC-Vlan-interface103] quit
[SwitchC] interface vlan-interface 106
[SwitchC-Vlan-interface106] multicast boundary 239.0.0.0 8
[SwitchC-Vlan-interface106] quit
```

On Switch D, configure VLAN-interface 107 to be the boundary of admin-scope zone 2.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 107
[SwitchD-Vlan-interface107] multicast boundary 239.0.0.0 8
[SwitchD-Vlan-interface107] quit
```

4) Configure C-BSRs and C-RPs

On Switch B, configure the service scope of RP advertisements and configure VLAN-interface 101 as a C-BSR and C-RP of admin-scope zone 1.

```
[SwitchB] acl number 2001
[SwitchB-acl-basic-2001] rule permit source 239.0.0.0 0.255.255.255
[SwitchB-acl-basic-2001] quit
[SwitchB] pim
[SwitchB-pim] c-bsr group 239.0.0.0 8
[SwitchB-pim] c-bsr vlan-interface 101
[SwitchB-pim] c-rp vlan-interface 101 group-policy 2001
[SwitchB-pim] quit
```

On Switch D, configure the service scope of RP advertisements and configure VLAN-interface 104 as a C-BSR and C-RP of admin-scope zone 2.

```
[SwitchD] acl number 2001
[SwitchD-acl-basic-2001] rule permit source 239.0.0.0 0.255.255.255
[SwitchD-acl-basic-2001] quit
[SwitchD] pim
[SwitchD-pim] c-bsr group 239.0.0.0 8
[SwitchD-pim] c-bsr vlan-interface 104
[SwitchD-pim] c-rp vlan-interface 104 group-policy 2001
[SwitchD-pim] quit
```

On Switch F, configure VLAN-interface 109 as a C-BSR and C-RP in the global scope zone.

```
<SwitchF> system-view
```

```
[SwitchF] pim
[SwitchF-pim] c-bsr global
[SwitchF-pim] c-bsr vlan-interface 109
[SwitchF-pim] c-rp vlan-interface 109
[SwitchF-pim] quit
```

5) Verify the configuration

To view the BSR election information and the C-RP information on a switch, use the **display pim bsr-info** command. For example:

View the BSR information and the locally configured C-RP information on Switch B.

```
[SwitchB] display pim bsr-info
 Elected BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Accept Preferred
     Scope: Global
     Uptime: 00:01:45
     Expires: 00:01:25
 Elected BSR Address: 10.110.1.2
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: 239.0.0.0/8
     Uptime: 00:04:54
     Next BSR message scheduled at: 00:00:06
 Candidate BSR Address: 10.110.1.2
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: 239.0.0.0/8

 Candidate RP: 10.110.1.2(Vlan-interface101)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:15
```

View the BSR information and the locally configured C-RP information on Switch D.

```
[SwitchD] display pim bsr-info
 Elected BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Accept Preferred
     Scope: Global
     Uptime: 00:01:45
     Expires: 00:01:25
 Elected BSR Address: 10.110.4.2
     Priority: 0
     Hash mask length: 30
```

```
     State: Elected
     Scope: 239.0.0.0/8
     Uptime: 00:03:48
     Next BSR message scheduled at: 00:01:12
 Candidate BSR Address: 10.110.4.2
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: 239.0.0.0/8

 Candidate RP: 10.110.4.2(Vlan-interface104)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:10
```

View the BSR information and the locally configured C-RP information on Switch F.

```
[SwitchF] display pim bsr-info
 Elected BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: Global
     Uptime: 00:11:11
     Next BSR message scheduled at: 00:00:49
 Candidate BSR Address: 10.110.9.1
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: Global

 Candidate RP: 10.110.9.1(Vlan-interface109)
     Priority: 0
     HoldTime: 150
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:55
```

To view the RP information learned on a switch, use the **display pim rp-info** command. For example:

View the RP information on Switch B.

```
[SwitchB] display pim rp-info
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
     RP: 10.110.9.1
     Priority: 0
     HoldTime: 150
     Uptime: 00:03:39
     Expires: 00:01:51

 Group/MaskLen: 239.0.0.0/8
```

```
    RP: 10.110.1.2 (local)
    Priority: 0
    HoldTime: 150
    Uptime: 00:07:44
    Expires: 00:01:51
```

View the RP information on Switch D.

```
[SwitchD] display pim rp-info
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
     RP: 10.110.9.1
     Priority: 0
     HoldTime: 150
     Uptime: 00:03:42
     Expires: 00:01:48

 Group/MaskLen: 239.0.0.0/8
     RP: 10.110.4.2 (local)
     Priority: 0
     HoldTime: 150
     Uptime: 00:06:54
     Expires: 00:02:41
```

View the RP information on Switch F.

```
[SwitchF] display pim rp-info
 PIM-SM BSR RP information:
 Group/MaskLen: 224.0.0.0/4
     RP: 10.110.9.1 (local)
     Priority: 0
     HoldTime: 150
     Uptime: 00:00:32
     Expires: 00:01:58
```

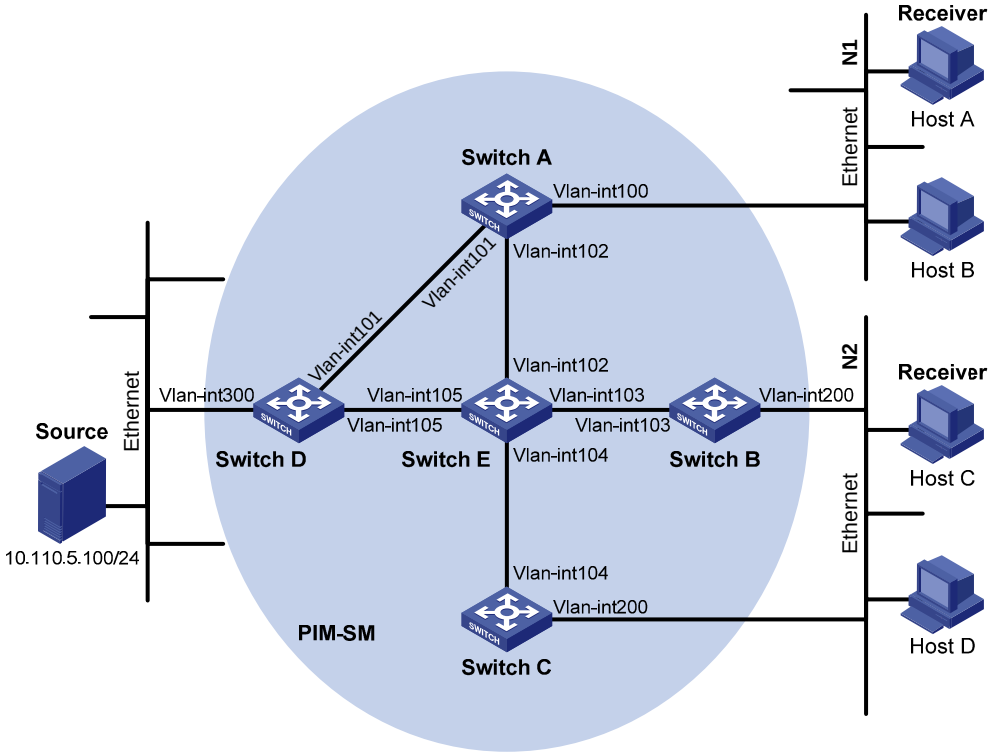
PIM-SSM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the SSM mode.
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Switch E connects to Switch A, Switch B, Switch C and Switch D.
- The SSM group range is 232.1.1.0/24.
- IGMPv3 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 1-13 Network diagram for PIM-SSM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int102	192.168.9.1/24		Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
	Vlan-int103	192.168.2.1/24		Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24		Vlan-int102	192.168.9.2/24
	Vlan-int104	192.168.3.1/24		Vlan-int105	192.168.4.1/24

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-13. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation in the PIM-SM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-SM and IGMP

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and run IGMPv3 on VLAN-interface 100, which connects Switch A to the stub network.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] igmp version 3
[SwitchA-Vlan-interface100] pim sm
```

```
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable IGMP on the corresponding interfaces on these two switches.

3) Configure the SSM group range

Configure the SSM group range to be 232.1.1.0/24 one Switch A.

```
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[SwitchA-acl-basic-2000] quit
[SwitchA] pim
[SwitchA-pim] ssm-policy 2000
[SwitchA-pim] quit
```

The configuration on Switch B, Switch C, Switch D and Switch E is similar to that on Switch A.

4) Verify the configuration

Carry out the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch A.

```
[SwitchA] display pim interface
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
 Vlan100             0      30         1          10.110.1.1     (local)
 Vlan101             1      30         1          192.168.1.2
 Vlan102             1      30         1          192.168.9.2
```

Assume that Host A needs to receive the information a specific multicast source S (10.110.5.100/24) sends to multicast group G (232.1.1.1). Switch A builds an SPT toward the multicast source. Switches on the SPT path (Switch A and Switch D) have generated an (S, G) entry, while Switch E, which is not on the SPT path, does not have multicast routing entries. You can use the **display pim routing-table** command to view the PIM routing table information on each switch. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface101
         Upstream neighbor: 192.168.1.2
         RPF prime neighbor: 192.168.1.2
     Downstream interface(s) information:
     Total number of downstreams: 1
```

```
        1: Vlan-interface100
            Protocol: igmp, UpTime: 00:13:25, Expires: 00:03:25
```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol: pim-ssm, Flag: LOC
     UpTime: 00:12:05
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface105
             Protocol: pim-ssm, UpTime: 00:12:05, Expires: 00:03:25
```

Troubleshooting PIM Configuration

Failure of Building a Multicast Distribution Tree Correctly

Symptom

None of the routers in the network (including routers directly connected with multicast sources and receivers) has multicast forwarding entries. That is, a multicast distribution tree cannot be built correctly and clients cannot receive multicast data.

Analysis

- When PIM-DM runs on the entire network, multicast data is flooded from the first hop router connected with the multicast source to the last hop router connected with the clients. When the multicast data is flooded to a router, no matter which router is, it creates (S, G) entries only if it has a route to the multicast source. If the router does not have a route to the multicast source, or if PIM-DM is not enabled on the router's RPF interface to the multicast source, the router cannot create (S, G) entries.
- When PIM-SM runs on the entire network, and when a router is to join the SPT, the router creates (S, G) entries only if it has a route to the multicast source. If the router does not have a router to the multicast source, or if PIM-DM is not enabled on the router's RPF interface to the multicast source, the router cannot create (S, G) entries.
- When a multicast router receives a multicast packet, it searches the existing unicast routing table for the optimal route to the RPF check object. The outgoing interface of this route will act as the RPF interface and the next hop will be taken as the RPF neighbor. The RPF interface completely relies on the existing unicast route, and is independent of PIM. The RPF interface must be PIM-enabled, and the RPF neighbor must also be a PIM neighbor. If PIM is not enabled on the router where the RPF interface or the RPF neighbor resides, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.
- Because a hello message does not carry the PIM mode information, a router running PIM is unable to know what PIM mode its PIM neighbor is running. If different PIM modes are enabled on the RPF

interface and on the corresponding interface of the RPF neighbor router, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.

- The same PIM mode must run on the entire network. Otherwise, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.

Solution

1) Check unicast routes. Use the **display ip routing-table** command to check whether a unicast route exists from the receiver host to the multicast source.
2) Check that PIM is enabled on the interfaces, especially on the RPF interface. Use the **display pim interface** command to view the PIM information on each interface. If PIM is not enabled on the interface, use the **pim dm** or **pim sm** command to enable PIM-DM or PIM-SM.
3) Check that the RPF neighbor is a PIM neighbor. Use the **display pim neighbor** command to view the PIM neighbor information.
4) Check that PIM and IGMP are enabled on the interfaces directly connecting to the multicast source and to the receivers.
5) Check that the same PIM mode is enabled on related interfaces. Use the **display pim interface verbose** command to check whether the same PIM mode is enabled on the RPF interface and the corresponding interface of the RPF neighbor router.
6) Check that the same PIM mode is enabled on all the routers in the entire network. Make sure that the same PIM mode is enabled on all the routers: PIM-SM on all routers, or PIM-DM on all routers. In the case of PIM-SM, also check that the BSR and RP configurations are correct.

Multicast Data Abnormally Terminated on an Intermediate Router

Symptom

An intermediate router can receive multicast data successfully, but the data cannot reach the last hop router. An interface on the intermediate router receives data but no corresponding (S, G) entry is created in the PIM routing table.

Analysis

- If a multicast forwarding boundary has been configured through the **multicast boundary** command, any multicast packet will be kept from crossing the boundary, and therefore no routing entry can be created in the PIM routing table.
- In addition, the **source-policy** command is used to filter received multicast packets. If the multicast data fails to pass the ACL rule defined in this command, PIM cannot create the route entry, either.

Solution

1) Check the multicast forwarding boundary configuration. Use the **display current-configuration** command to check the multicast forwarding boundary settings. Use the **multicast boundary** command to change the multicast forwarding boundary settings.
2) Check the multicast filter configuration. Use the **display current-configuration** command to check the multicast filter configuration. Change the ACL rule defined in the **source-policy** command so that the source/group address of the multicast data can pass ACL filtering.

RPs Unable to Join SPT in PIM-SM

Symptom

An RPT cannot be established correctly, or the RPs cannot join the SPT to the multicast source.

Analysis

- As the core of a PIM-SM domain, the RPs serve specific multicast groups. Multiple RPs can coexist in a network. Make sure that the RP information on all routers is exactly the same, and a specific group is mapped to the same RP. Otherwise, multicast forwarding will fail.
- If the static RP mechanism is used, the same static RP command must be executed on all the routers in the entire network. Otherwise, multicast forwarding will fail.

Solution

1) Check that a route is available to the RP. Carry out the **display ip routing-table** command to check whether a route is available on each router to the RP.
2) Check the dynamic RP information. Use the **display pim rp-info** command to check whether the RP information is consistent on all routers.
3) Check the configuration of static RPs. Use the **display pim rp-info** command to check whether the same static RP address has been configured on all the routers in the entire network.

RPT Establishment Failure or Source Registration Failure in PIM-SM

Symptom

C-RPs cannot unicast advertise messages to the BSR. The BSR does not advertise bootstrap messages containing C-RP information and has no unicast route to any C-RP. An RPT cannot be established correctly, or the DR cannot perform source register with the RP.

Analysis

- The C-RPs periodically send C-RP-Adv messages to the BSR by unicast. If a C-RP has no unicast route to the BSR, the BSR cannot receive C-RP-Adv messages from that C-RP and the bootstrap message of the BSR will not contain the information of that C-RP.
- In addition, if the BSR does not have a unicast router to a C-RP, it will discard the C-RP-Adv messages from that C-RP, and therefore the bootstrap messages of the BSR will not contain the information of that C-RP.
- The RP is the core of a PIM-SM domain. Make sure that the RP information on all routers is exactly the same, a specific group G is mapped to the same RP, and unicast routes are available to the RP.

Solution

1) Check whether routes to C-RPs and the BSR are available. Carry out the **display ip routing-table** command to check whether routes are available on each router to the RP and the BSR, and whether a route is available between the RP and the BSR. Make sure that each C-RP has a unicast route to the BSR, the BSR has a unicast route to each C-RP, and all the routers in the entire network have a unicast route to the RP.
2) Check the RP and BSR information. PIM-SM needs the support of the RP and BSR. Use the **display pim bsr-info** command to check whether the BSR information is available on each router, and then use the **display pim rp-info** command to check whether the RP information is correct.
3) View PIM neighboring relationships. Use the **display pim neighbor** command to check whether the normal PIM neighboring relationships have been established among the routers.

Table of Contents

1 MSDP Configuration

When configuring MSDP, go to these sections for information you are interested in:

Note

- The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running the MSDP protocol.
- For details about the concepts of designated router (DR), bootstrap router (BSR), candidate-BSR (C-BSR), rendezvous point (RP), candidate RP (C-RP), shortest path tree (SPT) and rendezvous point tree (RPT) mentioned in this manual, refer to *PIM Configuration* in the *IP Multicast Volume*.

MSDP Overview

Introduction to MSDP

Multicast source discovery protocol (MSDP) is an inter-domain multicast solution developed to address the interconnection of protocol independent multicast sparse mode (PIM-SM) domains. It is used to discover multicast source information in other PIM-SM domains.

In the basic PIM-SM mode, a multicast source registers only with the RP in the local PIM-SM domain, and the multicast source information of a domain is isolated from that of another domain. As a result, the RP is aware of the source information only within the local domain and a multicast distribution tree is built only within the local domain to deliver multicast data from a local multicast source to local receivers. If there is a mechanism that allows RPs of different PIM-SM domains to share their multicast source information, the local RP will be able to join multicast sources in other domains and multicast data can be transmitted among different domains.

MSDP achieves this goal. With MSDP peer relationships established between appropriate routers in the network, the RPs of different PIM-SM domains are interconnected with one another. Source active (SA) messages are exchanged between these MSDP peers and thus the multicast source information is shared among these different domains.

Caution

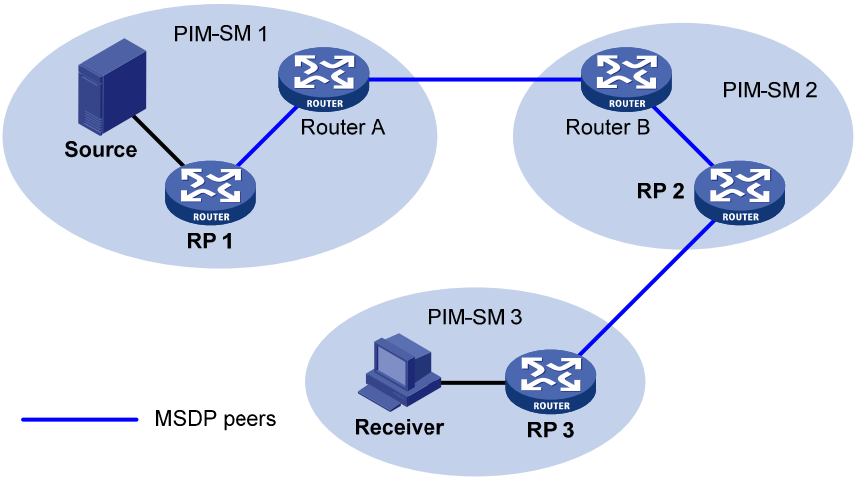
- MSDP is applicable only if the intra-domain multicast protocol is PIM-SM.
- MSDP is meaningful only for the any-source multicast (ASM) model.

How MSDP Works

MSDP peers

With one or more pairs of MSDP peers configured in the network, an MSDP interconnection map is formed, where the RPs of different PIM-SM domains are interconnected in series. Relayed by these MSDP peers, an SA message sent by an RP can be delivered to all other RPs.

Figure 1-1 Where MSDP peers are in the network



As shown in Figure 1-1, an MSDP peer can be created on any PIM-SM router. MSDP peers created on PIM-SM routers that assume different roles function differently.

1) MSDP peers on RPs

- Source-side MSDP peer: the MSDP peer nearest to the multicast source (Source), typically the source-side RP, like RP 1. The source-side RP creates SA messages and sends the messages to its remote MSDP peer to notify the MSDP peer of the locally registered multicast source information. A source-side MSDP peer must be created on the source-side RP; otherwise it will not be able to advertise the multicast source information out of the PIM-SM domain.
- Receiver-side MSDP peer: the MSDP peer nearest to the receivers, typically the receiver-side RP, like RP 3. Upon receiving an SA message, the receiver-side MSDP peer resolves the multicast source information carried in the message and joins the SPT rooted at the source across the PIM-SM domain. When multicast data from the multicast source arrives, the receiver-side MSDP peer forwards the data to the receivers along the RPT.
- Intermediate MSDP peer: an MSDP peer with multicast remote MSDP peers, like RP 2. An intermediate MSDP peer forwards SA messages received from one remote MSDP peer to other remote MSDP peers, functioning as a relay of multicast source information.

2) MSDP peers created on common PIM-SM routers (other than RPs)

Router A and Router B are MSDP peers on common multicast routers. Such MSDP peers just forward received SA messages.

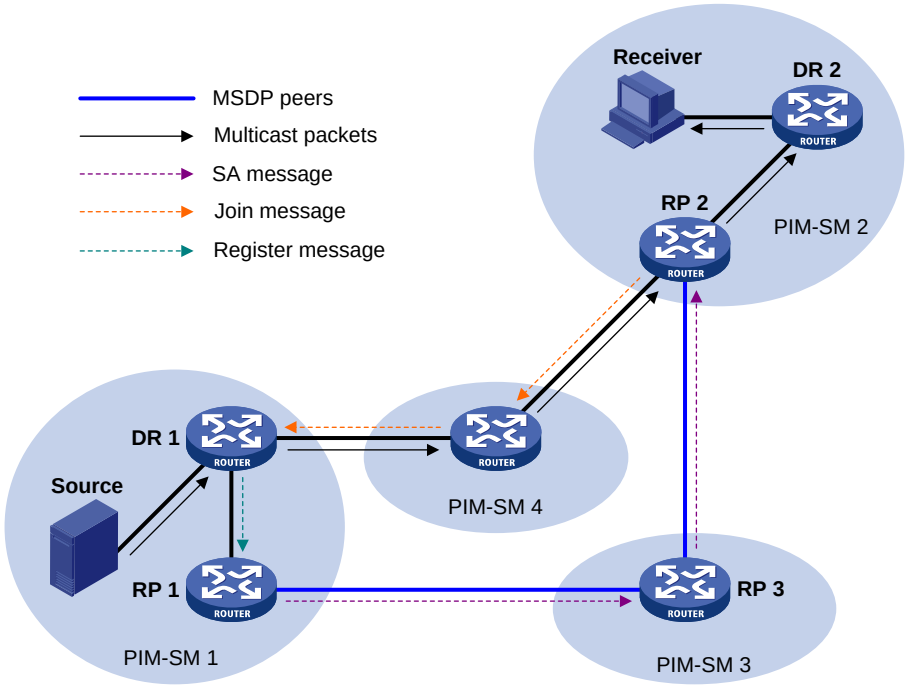
Note

In a PIM-SM network running the BSR mechanism, the RP is dynamically elected from C-RPs. To enhance network robustness, a PIM-SM network typically has more than one C-RP. As the RP election result is unpredictable, MSDP peering relationships should be built among all C-RPs so that the winner C-RP is always on the "MSDP interconnection map", while loser C-RPs will assume the role of common PIM-SM routers on the "MSDP interconnection map".

Implementing inter-domain multicast delivery by leveraging MSDP peers

As shown in Figure 1-2, an active source (Source) exists in the domain PIM-SM 1, and RP 1 has learned the existence of Source through multicast source registration. If RPs in PIM-SM 2 and PIM-SM 3 also wish to know the specific location of Source so that receiver hosts can receive multicast traffic originated from it, MSDP peering relationships should be established between RP 1 and RP 3 and between RP 3 and RP 2 respectively.

Figure 1-2 MSDP peering relationships



The process of implementing inter-domain multicast delivery by leveraging MSDP peers is as follows:

1) When the multicast source in PIM-SM 1 sends the first multicast packet to multicast group G, DR 1 encapsulates the multicast data within a register message and sends the register message to RP 1. Then, RP 1 gets aware of the information related to the multicast source.
2) As the source-side RP, RP 1 creates SA messages and periodically sends the SA messages to its MSDP peer. An SA message contains the source address (S), the multicast group address (G), and the address of the RP which has created this SA message (namely RP 1).

3) On MSDP peers, each SA message is subject to a reverse path forwarding (RPF) check and multicast policy–based filtering, so that only SA messages that have arrived along the correct path and passed the filtering are received and forwarded. This avoids delivery loops of SA messages. In addition, you can configure MSDP peers into an MSDP mesh group so as to avoid flooding of SA messages between MSDP peers.
4) SA messages are forwarded from one MSDP peer to another, and finally the information of the multicast source traverses all PIM-SM domains with MSDP peers (PIM-SM 2 and PIM-SM 3 in this example).
5) Upon receiving the SA message create by RP 1, RP 2 in PIM-SM 2 checks whether there are any receivers for the multicast group in the domain.

- If so, the RPT for the multicast group G is maintained between RP 2 and the receivers. RP 2 creates an (S, G) entry, and sends an (S, G) join message hop by hop towards DR 1 at the multicast source side, so that it can directly join the SPT rooted at the source over other PIM-SM domains. Then, the multicast data can flow along the SPT to RP 2 and is forwarded by RP 2 to the receivers along the RPT. Upon receiving the multicast traffic, the DR at the receiver side (DR 2) decides whether to initiate an RPT-to-SPT switchover process.
- If no receivers for the group exist in the domain, RP 2 does not create an (S, G) entry and does join the SPT rooted at the source.

Note

- An MSDP mesh group refers to a group of MSDP peers that have MSDP peering relationships among one another and share the same group name.
- When using MSDP for inter-domain multicasting, once an RP receives information form a multicast source, it no longer relies on RPs in other PIM-SM domains. The receivers can override the RPs in other domains and directly join the multicast source-based SPT.

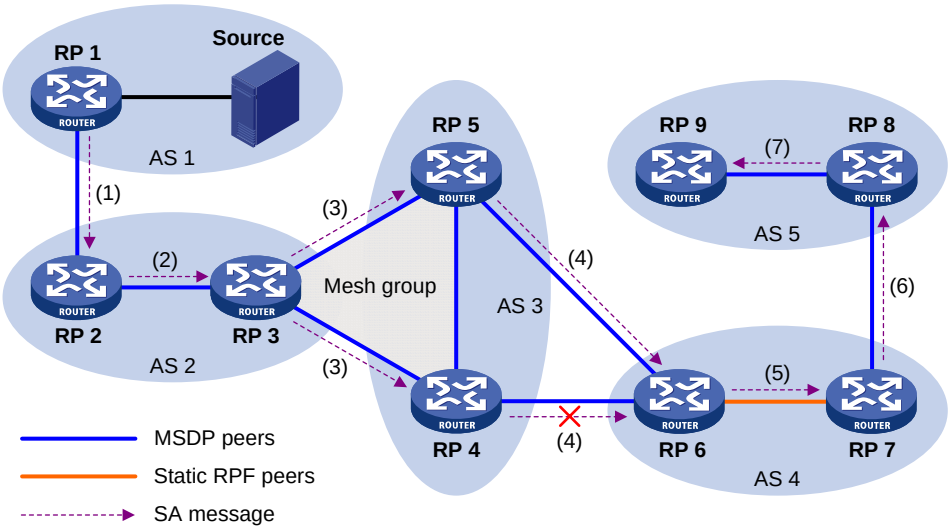
RPF check rules for SA messages

As shown in Figure 1-3, there are five autonomous systems in the network, AS 1 through AS 5, with IGP enabled on routers within each AS and BGP or MBGP as the interoperation protocol among different ASs. Each AS contains at least one PIM-SM domain and each PIM-SM domain contains one ore more RPs. MSDP peering relationships have been established among different RPs. RP 3, RP 4 and RP 5 are in an MSDP mesh group. On RP 7, RP 6 is configured as its static RPF peer.

Note

If only one MSDP peer exists in a PIM-SM domain, this PIM-SM domain is also called a stub domain. For example, AS 4 in Figure 1-3 is a stub domain. The MSDP peer in a stub domain can have multiple remote MSDP peers at the same time. You can configure one or more remote MSDP peers as static RPF peers. When an RP receives an SA message from a static RPF peer, the RP accepts the SA message and forwards it to other peers without performing an RPF check.

Figure 1-3 Diagram for RPF check for SA messages



As illustrated in Figure 1-3, these MSDP peers dispose of SA messages according to the following RPF check rules:

1) When RP 2 receives an SA message from RP 1

Because the source-side RP address carried in the SA message is the same as the MSDP peer address, which means that the MSDP peer where the SA is from is the RP that has created the SA message, RP 2 accepts the SA message and forwards it to its other MSDP peer (RP 3).

2) When RP 3 receives the SA message from RP 2

Because the SA message is from an MSDP peer (RP 2) in the same AS, and the MSDP peer is the next hop on the optimal path to the source-side RP, RP 3 accepts the message and forwards it to other peers (RP 4 and RP 5).

3) When RP 4 and RP 5 receive the SA message from RP 3

Because the SA message is from an MSDP peer (RP 3) in the same mesh group, RP 4 and RP 5 both accept the SA message, but they do not forward the message to other members in the mesh group; instead, they forward it to other MSDP peers (RP 6 in this example) out of the mesh group.

4) When RP 6 receives the SA messages from RP 4 and RP 5 (suppose RP 5 has a higher IP address)

Although RP 4 and RP 5 are in the same AS (AS 3) and both are MSDP peers of RP 6, because RP 5 has a higher IP address, RP 6 accepts only the SA message from RP 5.

5) When RP 7 receives the SA message from RP 6

Because the SA message is from a static RPF peer (RP 6), RP 7 accepts the SA message and forwards it to other peer (RP 8).

6) When RP 8 receives the SA message from RP 7

A BGP or MBGP route exists between two MSDP peers in different ASs. Because the SA message is from an MSDP peer (RP 7) in a different AS, and the MSDP peer is the next hop on the BGP or MBGP route to the source-side RP, RP 8 accepts the message and forwards it to its other peer (RP 9).

7) When RP 9 receives the SA message from RP 8

Because RP 9 has only one MSDP peer, RP 9 accepts the SA message.

SA messages from other paths than described above will not be accepted nor forwarded by MSDP peers.

Implementing intra-domain Anycast RP by leveraging MSDP peers

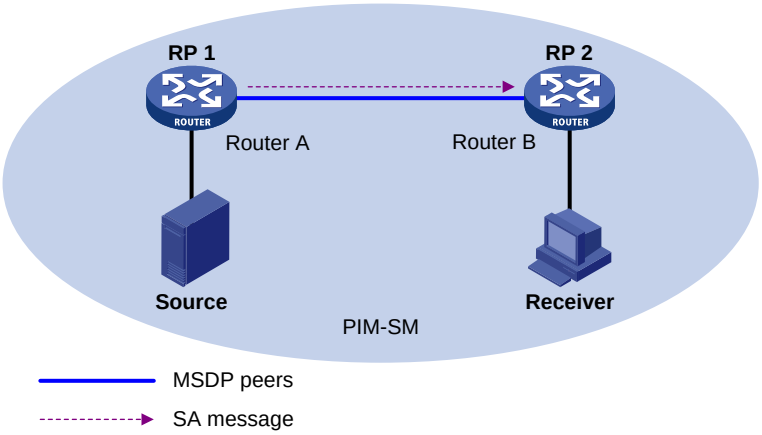
Anycast RP refers to such an application that enables load balancing and redundancy backup between two or more RPs within a PIM-SM domain by configuring the same IP address for, and establishing MSDP peering relationships between, these RPs.

As shown in Figure 1-4, within a PIM-SM domain, a multicast source sends multicast data to multicast group G, and Receiver is a member of the multicast group. To implement Anycast RP, configure the same IP address (known as anycast RP address, typically a private address) on Router A and Router B, configure these interfaces as C-RPs, and establish an MSDP peering relationship between Router A and Router B.

Note

Usually an Anycast RP address is configured on a logic interface, like a loopback interface.

Figure 1-4 Typical network diagram of Anycast RP



The work process of Anycast RP is as follows:

1) The multicast source registers with the nearest RP. In this example, Source registers with RP 1, with its multicast data encapsulated in the register message. When the register message arrives at RP 1, RP 1 decapsulates the message.
2) Receivers send join messages to the nearest RP to join in the RPT rooted as this RP. In this example, Receiver joins the RPT rooted at RP 2.
3) RPs share the registered multicast information by means of SA messages. In this example, RP 1 creates an SA message and sends it to RP 2, with the multicast data from Source encapsulated in the SA message. When the SA message reaches RP 2, RP 2 decapsulates the message.
4) Receivers receive the multicast data along the RPT and directly join the SPT rooted at the multicast source. In this example, RP 2 forwards the multicast data down the RPT. When Receiver receives the multicast data from Source, it directly joins the SPT rooted at Source.

The significance of Anycast RP is as follows:

- Optimal RP path: A multicast source registers with the nearest RP so that an SPT with the optimal path is built; a receiver joins the nearest RP so that an RPT with the optimal path is built.
- Load balancing between RPs: Each RP just needs to maintain part of the source/group information within the PIM-SM domain and forward part of the multicast data, thus achieving load balancing between different RPs.
- Redundancy backup between RPs: When an RP fails, the multicast source previously registered on it or the receivers previous joined it will register with or join another nearest RP, thus achieving redundancy backup between RPs.

Caution

- Be sure to configure a 32-bit subnet mask (255.255.255.255) for the Anycast RP address, namely configure the Anycast RP address into a host address.
- An MSDP peer address must be different from the Anycast RP address.

Protocols and Standards

MSDP is documented in the following specifications:

- RFC 3618: Multicast Source Discovery Protocol (MSDP)
- RFC 3446: Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)

MSDP Configuration Task List

Complete these tasks to configure MSDP:

Task		Remarks
Configuring Basic Functions of MSDP	Enabling MSDP	Required
	Creating an MSDP Peer Connection	Required
	Configuring a Static RPF Peer	Optional
Configuring an MSDP Peer Connection	Configuring MSDP Peer Description	Optional
	Configuring an MSDP Mesh Group	Optional
	Configuring MSDP Peer Connection Control	Optional
Configuring SA Messages Related Parameters	Configuring SA Message Content	Optional
	Configuring SA Request Messages	Optional
	Configuring SA Message Filtering Rules	Optional
	Configuring the SA Cache Mechanism	Optional

Configuring Basic Functions of MSDP

Note

All the configuration tasks should be carried out on RPs in PIM-SM domains, and each of these RPs acts as an MSDP peer.

Configuration Prerequisites

Before configuring the basic functions of MSDP, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring PIM-SM to enable intra-domain multicast forwarding.

Before configuring the basic functions of MSDP, prepare the following data:

- IP addresses of MSDP peers
- Address prefix list for an RP address filtering policy

Enabling MSDP

Follow these steps to enable MSDP globally

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IP multicast routing	**multicast routing-enable**	Required Disabled by default
Enable MSDP and enter MSDP view	**msdp**	Required Disabled by default

Note

For details about the **multicast routing-table** command, see *Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Creating an MSDP Peer Connection

An MSDP peering relationship is identified by an address pair, namely the address of the local MSDP peer and that of the remote MSDP peer. An MSDP peer connection must be created on both devices that are a pair of MSDP peers.

Follow these steps to create an MSDP peer connection:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Create an MSDP peer connection	**peer** *peer-address* **connect-interface** *interface-type interface-number*	Required No MSDP peer connection created by default

Note

If an interface of the router is shared by an MSDP peer and a BGP/MBGP peer at the same time, we recommend that you use the IP address of the BGP/MBGP peer as the IP address of the for the MSDP peer.

Configuring a Static RPF Peer

Configuring static RPF peers avoids RPF check of SA messages.

Follow these steps to configure a static RPF peer:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Configure a static RPF peer	**static-rpf-peer** *peer-address* [**rp-policy** *ip-prefix-name*]	Required No static RPF peer configured by default

Note

If only one MSDP peer is configured on a router, this MSDP will be registered as a static RPF peer.

Configuring an MSDP Peer Connection

Configuration Prerequisites

Before configuring MSDP peer connection, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring basic functions of MSDP

Before configuring an MSDP peer connection, prepare the following data:

- Description information of MSDP peers
- Name of an MSDP mesh group
- MSDP peer connection retry interval

Configuring MSDP Peer Description

With the MSDP peer description information, the administrator can easily distinguish different MSDP peers and thus better manage MSDP peers.

Follow these steps to configure description for an MSDP peer:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Configure description for an MSDP peer	**peer** *peer-address* **description** *text*	Required No description for MSDP peers by default

Configuring an MSDP Mesh Group

An AS may contain multiple MSDP peers. You can use the MSDP mesh group mechanism to avoid SA message flooding among these MSDP peers and optimize the multicast traffic.

On one hand, an MSDP peer in an MSDP mesh group forwards SA messages from outside the mesh group that have passed the RPF check to the other members in the mesh group; on the other hand, a mesh group member accepts SA messages from inside the group without performing an RPF check, and does not forward the message within the mesh group either. This mechanism not only avoids SA flooding but also simplifies the RPF check mechanism, because BGP or MBGP is not needed to run between these MSDP peers.

By configuring the same mesh group name for multiple MSDP peers, you can create a mesh group assign those MSDP peers to that mesh group.

Follow these steps to create an MSDP mesh group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Create an MSDP mesh group and assign an MSDP peer to that mesh group	**peer** *peer-address* **mesh-group** *name*	Required An MSDP peer does not belong to any mesh group by default

Note

- Before grouping multiple routers into an MSDP mesh group, make sure that these routers are interconnected with one another.
- If you configure more than one mesh group name on an MSDP peer, only the last configuration is effective.

Configuring MSDP Peer Connection Control

MSDP peers are interconnected over TCP (port number 639). You can flexibly control sessions between MSDP peers by manually deactivating and reactivating the MSDP peering connections. When the connection between two MSDP peers is deactivated, SA messages will no longer be delivered between them, and the TCP connection is closed without any connection setup retry, but the configuration information will remain unchanged.

When a new MSDP peer is created, or when a previously deactivated MSDP peer connection is reactivated, or when a previously failed MSDP peer attempts to resume operation, a TCP connection is required. You can flexibly adjust the interval between MSDP peering connection retries.

Follow these steps to configure MSDP peer connection control:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Deactivate an MSDP peer	**shutdown** *peer-address*	Optional Active by default
Configure the interval between MSDP peer connection retries	**timer retry** *interval*	Optional 30 seconds by default

Configuring SA Messages Related Parameters

Configuration Prerequisites

Before configuring SA message delivery, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring basic functions of MSDP

Before configuring SA message delivery, prepare the following data:

- ACL rules for filtering SA request messages
- ACL rules as SA message creation rules
- ACL rules for filtering SA messages to be received and forwarded
- TTL threshold for multicast packet encapsulation in SA messages
- Maximum number of (S, G) entries learned from the specified MSDP peer that the router can cache

Configuring SA Message Content

Some multicast sources send multicast data at an interval longer than the aging time of (S, G) entries. In this case, the source-side DR has to encapsulate multicast data packet by packet in register messages and send them to the source-side RP. The source-side RP transmits the (S, G) information to the remote RP through SA messages. Then the remote RP joins the source-side DR and builds an SPT. Since the (S, G) entries have timed out, remote receivers can never receive the multicast data from the multicast source.

If the source-side RP is enabled to encapsulate register messages in SA messages, when there is a multicast packet to deliver, the source-side RP encapsulates a register message containing the multicast packet in an SA message and sends it out. After receiving the SA message, the remote RP

decapsulates the SA message and delivers the multicast data contained in the register message to the receivers along the RPT.

The MSDP peers deliver SA messages to one another. Upon receiving an SA message, a router performs RPF check on the message. If the router finds that the remote RP address is the same as the local RP address, it will discard the SA message. In the Anycast RP application, however, you need to configure RPs with the same IP address on two or more routers in the same PIM-SM domain, and configure these routers as MSDP peers to one another. Therefore, a logic RP address (namely the RP address on the logic interface) that is different from the actual RP address must be designated for SA messages so that the messages can pass the RPF check.

Follow these steps to configure the SA message content:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Enable encapsulation of a register message	**encap-data-enable**	Optional Disabled by default
Configure the interface address as the RP address in SA messages	**originating-rp** *interface-type interface-number*	Optional PIM RP address by default

Configuring SA Request Messages

By default, upon receiving a new Join message, a router does not send an SA request message to any MSDP peer; instead, it waits for the next SA message from its MSDP peer. This will cause the receiver to delay obtaining multicast source information. To enable a new receiver to get the currently active multicast source information as early as possible, you can configure routers to send SA request messages to the designated MSDP peers upon receiving a Join message of a new receiver.

Follow these steps to configure SA message transmission and filtering:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Enable the device to send SA request messages	**peer** *peer-address* **request-sa-enable**	Optional Disabled by default
Configure a filtering rule for SA request messages	**peer** *peer-address* **sa-request-policy** [**acl** *acl-number*]	Optional SA request messages are not filtered by default

Caution

Before you can enable the device to send SA requests, be sure to disable the SA message cache mechanism.

Configuring SA Message Filtering Rules

By configuring an SA message creation rule, you can enable the router to filter the (S, G) entries to be advertised when creating an SA message, so that the propagation of messages of multicast sources is controlled.

By configuring a filtering rule for receiving or forwarding SA messages, you can enable the router to filter the (S, G) forwarding entries to be advertised when receiving or forwarding an SA message, so that the propagation of multicast source information is controlled at SA message reception or forwarding.

By configuring a TTL threshold for multicast data packet encapsulation in SA messages, you can control the multicast data packet encapsulation in SA messages and limit the propagation range of SA messages:

- Before creating an SA message with an encapsulated multicast data packet, the router checks the TTL value of the multicast data packet. If the TTL value is less than the threshold, the router does not create an SA message; if the TTL value is greater than or equal to the threshold, the router encapsulates the multicast data in an SA message and sends the SA message out.
- Upon receiving an SA message with an encapsulated multicast data packet, the router decrements the TTL value of the multicast packet by 1 and then checks the TTL value. If the TTL value is less than the threshold, the router does not forward the SA message to the designated MSDP peer; if the TTL value is greater than or equal to the threshold, the router re-encapsulates the multicast data in an SA message and sends the SA message out.

Follow these steps to configure a filtering rule for receiving or forwarding SA messages:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Configure an SA message creation rule	**import-source** [**acl** *acl-number*]	Required No restrictions on (S, G) entries by default
Configure a filtering rule for receiving or forwarding SA messages	**peer** *peer-address* **sa-policy** { **import** | **export** } [**acl** *acl-number*]	Required No filtering rule by default
Configure the TTL threshold for multicast data packet encapsulation in SA messages	**peer** *peer-address* **minimum-ttl** *ttl-value*	Optional 0 by default

Configuring the SA Cache Mechanism

To reduce the time spent in obtaining the multicast information, you can enable the SA cache mechanism to cache (S, G) entries contained in SA messages locally on the router. However, the more (S, G) entries are cached, the larger memory space of the router is used.

With the SA cache mechanism enabled, when receiving a new (*, G) join message, the router searches its SA cache first:

- If the corresponding (S, G) entry does not exist in the cache, the router waits for the SA message its MSDP peer will send in the next cycle;
- If the corresponding (S, G) entry exists in the cache, the router joins the corresponding SPT rooted at S.

To protect the router effectively against denial of service (DoS) attacks, you can set a limit on the number of (S, G) entries the router can cache.

Follow these steps to configure the SA message cache:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MSDP view	**msdp**	—
Enable the SA cache mechanism	**cache-sa-enable**	Optional Enabled by default
Configure the maximum number of (S, G) entries learned from the specified MSDP peer that the router can cache	**peer** *peer-address* **sa-cache-maximum** *sa-limit*	Optional 8192 by default

Displaying and Maintaining MSDP

To do...	Use the command...	Remarks
View the brief information of MSDP peers	**display msdp brief** [**state** { **connect** | **down** | **listen** | **shutdown** | **up** }]	Available in any view
View the detailed information about the status of MSDP peers	**display msdp peer-status** [*peer-address*]	Available in any view
View the (S, G) entry information in the SA cache	**display msdp sa-cache** [*group-address* | *source-address* | *as-number*] *	Available in any view
View the number of (S, G) entries in the SA cache	**display msdp sa-count** [*as-number*]	Available in any view
Reset the TCP connection with an MSDP peer	**reset msdp peer** [*peer-address*]	Available in user view
Clear (S, G) entries in the SA cache	**reset msdp sa-cache** [*group-address*]	Available in user view
Clear all statistics information of an MSDP peer	**reset msdp statistics** [*peer-address*]	Available in user view

MSDP Configuration Examples

Inter-AS Multicast Configuration Leveraging BGP Routes

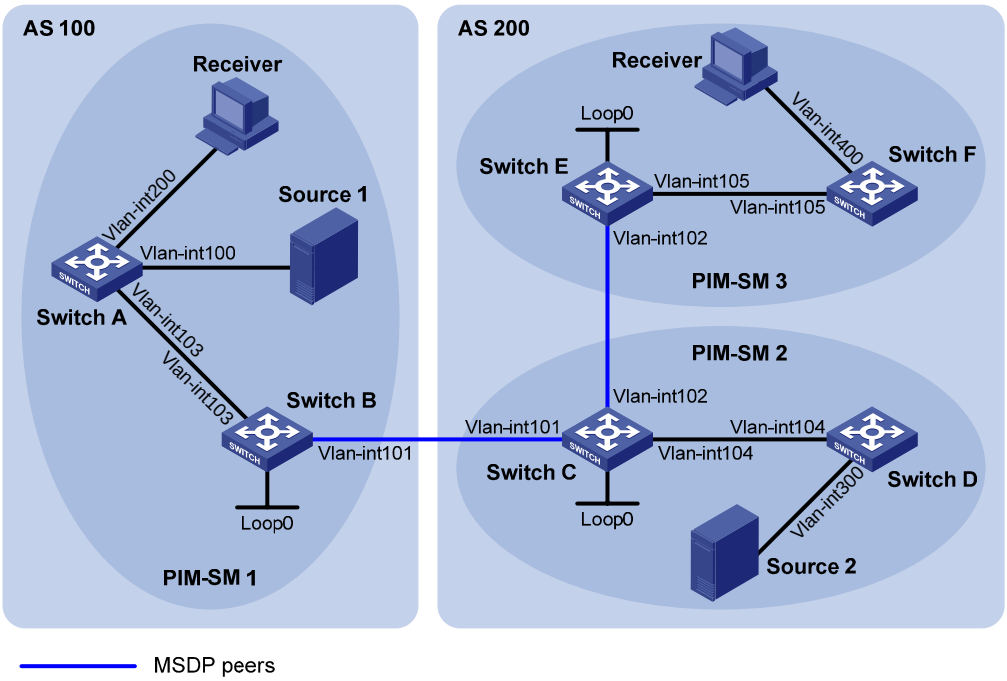
Network requirements

- There are two ASs in the network, AS 100 and AS 200 respectively. OSPF is running within each AS, and BGP is running between the two ASs.
- PIM-SM 1 belongs to AS 100, while PIM-SM 2 and PIM-SM 3 belong to AS 200.
- Each PIM-SM domain has zero or one multicast source and receiver. OSPF runs within each domain to provide unicast routes.
- It is required that the respective Loopback 0 of Switch B, Switch C and Switch E be configured as the C-BSR and C-RP of the respective PIM-SM domains.

- It is required that an MSDP peering relationship be set up between Switch B and Switch C through EBGP, and between Switch C and Switch E through IBGP.

Network diagram

Figure 1-5 Network diagram for inter-AS multicast configuration leveraging BGP routes (on switches)



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int103	10.110.1.2/24	Switch D	Vlan-int104	10.110.4.2/24
	Vlan-int100	10.110.2.1/24		Vlan-int300	10.110.5.1/24
	Vlan-int200	10.110.3.1/24	Switch E	Vlan-int105	10.110.6.1/24
Switch B	Vlan-int103	10.110.1.1/24		Vlan-int102	192.168.3.2/24
	Vlan-int101	192.168.1.1/24		Loop0	3.3.3.3/32
	Loop0	1.1.1.1/32	Switch F	Vlan-int105	10.110.6.2/24
Switch C	Vlan-int104	10.110.4.1/24		Vlan-int400	10.110.7.1/24
	Vlan-int102	192.168.3.1/24	Source 1	—	10.110.2.100/24
	Vlan-int101	192.168.1.2/24	Source 2	—	10.110.5.100/24
	Loop0	2.2.2.2/32			

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-5. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between switches in each AS. Ensure the network-layer interoperation among each AS, and ensure the dynamic update of routing information between the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, enable PIM-SM on each interface, and configure a PIM-SM domain border

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
```

```
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim sm
[SwitchA-Vlan-interface103] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] igmp enable
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit
```

The configuration on Switch B, Switch C, Switch D, Switch E, and Switch F is similar to the configuration on Switch A.

Configure a PIM domain border on Switch B.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

3) Configure C-BSRs and C-RPs

Configure Loopback 0 as a C-BSR and a C-RP on Switch B.

```
[SwitchB] pim
[SwitchB-pim] c-bsr loopback 0
[SwitchB-pim] c-rp loopback 0
[SwitchB-pim] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

4) Configure BGP for mutual route redistribution between BGP and OSPF

Configure EBGP on Switch B, and redistribute OSPF routes.

```
[SwitchB] bgp 100
[SwitchB-bgp] router-id 1.1.1.1
[SwitchB-bgp] peer 192.168.1.2 as-number 200
[SwitchB-bgp] import-route ospf 1
[SwitchB-bgp] quit
```

Configure IBGP and EBGP on Switch C, and redistribute OSPF routes.

```
[SwitchC] bgp 200
[SwitchC-bgp] router-id 2.2.2.2
[SwitchC-bgp] peer 192.168.1.1 as-number 100
[SwitchC-bgp] peer 192.168.3.2 as-number 200
[SwitchC-bgp] import-route ospf 1
[SwitchC-bgp] quit
```

Configure IBGP on Switch E, and redistribute OSPF routes.

```
[SwitchE] bgp 200
[SwitchE-bgp] router-id 3.3.3.3
[SwitchE-bgp] peer 192.168.3.1 as-number 200
[SwitchE-bgp] import-route ospf 1
[SwitchE-bgp] quit
```

Redistribute BGP routes into OSPF on Switch B.

```
[SwitchB] ospf 1
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

5) Configure MSDP peers

Configure an MSDP peer on Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchB-msdp] quit
```

Configure an MSDP peer on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchC-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchC-msdp] quit
```

Configure MSDP peers on Switch E.

```
[SwitchE] msdp
[SwitchE-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchE-msdp] quit
```

6) Verify the configuration

Carry out the **display bgp peer** command to view the BGP peering relationships between the switches. For example:

View the information about BGP peering relationships on Switch B.

```
[SwitchB] display bgp peer

 BGP local router ID : 1.1.1.1
 Local AS number : 100
 Total number of peers : 1                 Peers in established state : 1

  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.2  4 200       24       21     0       6 00:13:09 Established
```

View the information about BGP peering relationships on Switch C.

```
[SwitchC] display bgp peer

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 2                 Peers in established state : 2

  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.1  4 100       18       16     0       1 00:12:04 Established
  192.168.3.2  4 200       21       20     0       6 00:12:05 Established
```

View the information about BGP peering relationships on Switch E.

```
[SwitchE] display bgp peer

BGP local router ID : 3.3.3.3
 Local AS number : 200
 Total number of peers : 1                 Peers in established state : 1

  Peer         V  AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.3.1  4 200       16       14     0       1 00:10:58 Established
```

To view the BGP routing table information on the switches, use the **display bgp routing-table** command. For example:

View the BGP routing table information on Switch C.

```
[SwitchC] display bgp routing-table

 Total Number of Routes: 13

 BGP Local router ID is 2.2.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>   1.1.1.1/32        192.168.1.1   0                  0       100?
 *>i  2.2.2.2/32        192.168.3.2   0          100     0       ?
 *>   3.3.3.3/32        0.0.0.0       0                  0       ?
 *>   192.168.1.0       0.0.0.0       0                  0       ?
 *                      192.168.1.1   0                  0       100?
 *>   192.168.1.1/32    0.0.0.0       0                  0       ?
 *>   192.168.1.2/32    0.0.0.0       0                  0       ?
 *                      192.168.1.1   0                  0       100?
 *>   192.168.3.0       0.0.0.0       0                  0       ?
 * i                    192.168.3.2   0          100     0       ?
 *>   192.168.3.1/32    0.0.0.0       0                  0       ?
 *>   192.168.3.2/32    0.0.0.0       0                  0       ?
 * i                    192.168.3.2   0          100     0       ?
```

When the multicast source in PIM-SM 1 (Source 1) and the multicast source in PIM-SM 2 (Source 2) send multicast information, receivers in PIM-SM 1 and PIM-SM 3 can receive the multicast data. You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches. For example:

View the brief information about MSDP peering relationships on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address     State     Up/Down time    AS     SA Count   Reset Count
```

```
  192.168.1.2        Up        00:12:27       200    13         0
```

View the brief information about MSDP peering relationships on Switch C.

```
[SwitchC] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.2       Up       00:15:32        200    8          0
  192.168.1.1       Up       00:06:39        100    13         0
```

View the brief information about MSDP peering relationships on Switch E.

```
[SwitchE] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.1       Up       01:07:08        200    8          0
```

View the detailed MSDP peer information on Switch B.

```
[SwitchB] display msdp peer-status
  MSDP Peer 192.168.1.2, AS 200
  Description:
  Information about connection status:
    State: Up
    Up/down time: 00:15:47
    Resets: 0
    Connection interface: Vlan-interface101 (192.168.1.1)
    Number of sent/received messages: 16/16
    Number of discarded output messages: 0
    Elapsed time since last connection or counters clear: 00:17:51
  Information about (Source, Group)-based SA filtering policy:
    Import policy: none
    Export policy: none
  Information about SA-Requests:
    Policy to accept SA-Request messages: none
    Sending SA-Requests status: disable
  Minimum TTL to forward SA with encapsulated data: 0
  SAs learned from this peer: 0, SA-cache maximum for the peer: none
  Input queue size: 0, Output queue size: 0
  Counters for MSDP message:
    Count of RPF check failure: 0
    Incoming/outgoing SA messages: 0/0
    Incoming/outgoing SA requests: 0/0
    Incoming/outgoing SA responses: 0/0
    Incoming/outgoing data packets: 0/0
```

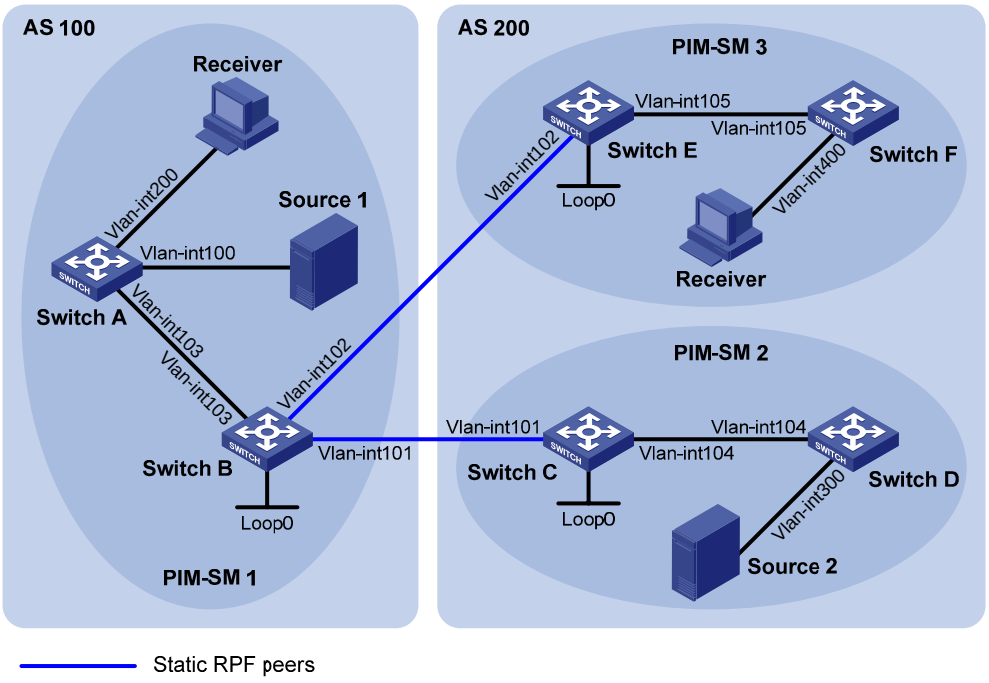
Inter-AS Multicast Configuration Leveraging Static RPF Peers

Network requirements

- There are two ASs in the network, AS 100 and AS 200 respectively. OSPF is running within each AS, and BGP is running between the two ASs.
- PIM-SM 1 belongs to AS 100, while PIM-SM 2 and PIM-SM 3 belong to AS 200.
- Each PIM-SM domain has zero or one multicast source and receiver. OSPF runs within each domain to provide unicast routes.
- PIM-SM 2 and PIM-SM 3 are both stub domains, and BGP or MBGP is not required between these two domains and PIM-SM 1. Instead, static RPF peers are configured to avoid RPF check on SA messages.
- It is required that the respective loopback 0 of Switch B, Switch C and Switch E be configured as the C-BSR and C-RP of the respective PIM-SM domains.
- It is required that Switch C and Switch E be configured as static RPF peers of Switch B, and Switch B be configured as the only static RPF peer of Switch C and Switch E, so that any switch can receive SA messages only from its static RPF peer(s) and permitted by the corresponding filtering policy.

Network diagram

Figure 1-6 Network diagram for inter-AS multicast configuration leveraging static RPF peers



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int103	10.110.1.2/24	Switch D	Vlan-int104	10.110.4.2/24
	Vlan-int100	10.110.2.1/24		Vlan-int300	10.110.5.1/24
	Vlan-int200	10.110.3.1/24	Switch E	Vlan-int105	10.110.6.1/24
Switch B	Vlan-int103	10.110.1.1/24		Vlan-int102	192.168.3.2/24
	Vlan-int101	192.168.1.1/24		Loop0	3.3.3.3/32
	Vlan-int102	192.168.3.1/24	Switch F	Vlan-int105	10.110.6.2/24
	Loop0	1.1.1.1/32		Vlan-int400	10.110.7.1/24
Switch C	Vlan-int101	192.168.1.2/24	Source 1	—	10.110.2.100/24
	Vlan-int104	10.110.4.1/24	Source 2	—	10.110.5.100/24
	Loop0	2.2.2.2/32			

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-6. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between the switches. Ensure the network-layer interoperation in each AS, and ensure the dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, enable PIM-SM and IGMP, and configure a PIM-SM domain border

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim sm
[SwitchA-Vlan-interface103] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] igmp enable
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit
```

The configuration on Switch B, Switch C, Switch D, Switch E, and Switch F is similar to the configuration on Switch A.

Configure PIM domain borders on Switch B.

```
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim bsr-boundary
[SwitchB-Vlan-interface102] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

3) Configure C-BSRs and C-RPs

Configure Loopback 0 as a C-BSR and a C-RP on Switch B.

```
[SwitchB] pim
[SwitchB-pim] c-bsr loopback 0
[SwitchB-pim] c-rp loopback 0
[SwitchB-pim] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

4) Configure a static RPF peer

Configure Switch C and Switch E as a static RPF peers of Switch B.

```
[SwitchB] ip ip-prefix list-df permit 192.168.0.0 16 greater-equal 16 less-equal 32
[SwitchB] msdp
```

```
[SwitchB-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchB-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchB-msdp] static-rpf-peer 192.168.3.2 rp-policy list-df
[SwitchB-msdp] static-rpf-peer 192.168.1.2 rp-policy list-df
[SwitchB-msdp] quit
```

Configure Switch B as a static RPF peer of Switch C.

```
[SwitchC] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16 less-equal 32
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchC-msdp] static-rpf-peer 192.168.1.1 rp-policy list-c
[SwitchC-msdp] quit
```

Configure Switch B as a static RPF peer of Switch E.

```
[SwitchE] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16 less-equal 32
[SwitchE] msdp
[SwitchE-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchE-msdp] static-rpf-peer 192.168.3.1 rp-policy list-c
[SwitchE-msdp] quit
```

5) Verify the configuration

Carry out the **display bgp peer** command to view the BGP peering relationships between the switches. If the command gives no output information, a BGP peering relationship has not been established between the switches.

When the multicast source in PIM-SM 1 (Source 1) and the multicast source in PIM-SM 2 (Source 2) send multicast information, receivers in PIM-SM 1 and PIM-SM 3 can receive the multicast data. You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches. For example:

View the brief MSDP peer information on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.3.2       Up       01:07:08        ?      8          0
  192.168.1.2       Up       00:16:39        ?      13         0
```

View the brief MSDP peer information on Switch C.

```
[SwitchC] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.1.1       Up       01:07:09        ?      8          0
```

View the brief MSDP peer information on Switch E.

```
[SwitchE] display msdp brief
MSDP Peer Brief Information
```

```
Configured    Up          Listen      Connect     Shutdown    Down
1             1           0           0           0           0

Peer's Address    State    Up/Down time    AS     SA Count    Reset Count
192.168.3.1       Up       00:16:40        ?      13          0
```

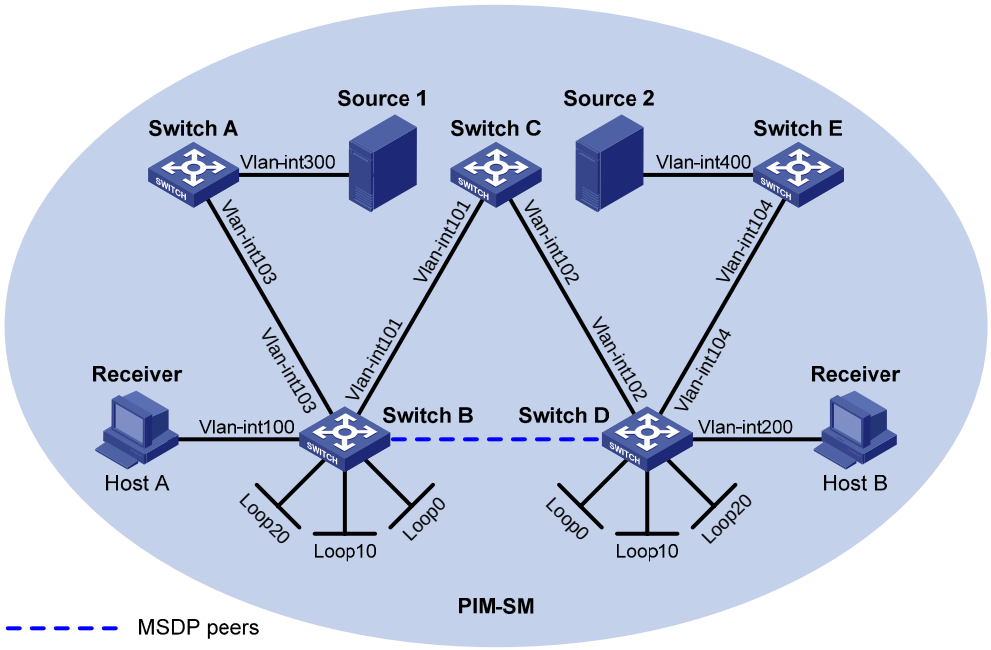
Anycast RP Configuration

Network requirements

- The PIM-SM domain has multiple multicast sources and receivers. OSPF runs within the domain to provide unicast routes.
- It is required to configure the anycast RP application so that the receiver-side DRs and the source-side DRs can initiate a Join message to their respective RPs that are the topologically nearest to them.
- On Switch B and Switch D, configure the interface Loopback 10 as a C-BSR, and Loopback 20 as a C-RP.
- The router ID of Switch B is 1.1.1.1, while the router ID of Switch D is 2.2.2.2. Set up an MSDP peering relationship between Switch B and Switch D.

Network diagram

Figure 1-7 Network diagram for anycast RP configuration



Device	Interface	IP address	Device	Interface	IP address
Source 1	—	10.110.5.100/24	Switch C	Vlan-int101	192.168.1.2/24
Source 2	—	10.110.6.100/24		Vlan-int102	192.168.2.2/24
Switch A	Vlan-int300	10.110.5.1/24	Switch D	Vlan-int200	10.110.3.1/24
	Vlan-int103	10.110.2.2/24		Vlan-int104	10.110.4.1/24
Switch B	Vlan-int100	10.110.1.1/24		Vlan-int102	192.168.2.1/24
	Vlan-int103	10.110.2.1/24		Loop0	2.2.2.2/32
	Vlan-int101	192.168.1.1/24		Loop10	4.4.4.4/32
	Loop0	1.1.1.1/32		Loop20	10.1.1.1/32
	Loop10	3.3.3.3/32	Switch E	Vlan-int400	10.110.6.1/24
	Loop20	10.1.1.1/32		Vlan-int104	10.110.4.2/24

Configuration procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-7. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between the switches. Ensure the network-layer interoperation among the switches, and ensure the dynamic update of routing information between the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IP multicast routing, and enable PIM-SM and IGMP

Enable IP multicast routing on Switch B, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 100.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] pim sm
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 103
[SwitchB-Vlan-interface103] pim sm
[SwitchB-Vlan-interface103] quit
[SwitchB] interface Vlan-interface 101
[SwitchB-Vlan-interface101] pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
[SwitchB] interface loopback 10
[SwitchB-LoopBack10] pim sm
[SwitchB-LoopBack10] quit
[SwitchB] interface loopback 20
[SwitchB-LoopBack20] pim sm
[SwitchB-LoopBack20] quit
```

The configuration on Switch A, Switch C, Switch D, and Switch E is similar to the configuration on Switch B.

3) Configure C-BSRs and C-RPs

Configure Loopback 10 as a C-BSR and Loopback 20 as a C-RP on Switch B.

```
[SwitchB] pim
[SwitchB-pim] c-bsr loopback 10
[SwitchB-pim] c-rp loopback 20
[SwitchB-pim] quit
```

The configuration on Switch D is similar to the configuration on Switch B.

4) Configure MSDP peers

Configure an MSDP peer on Loopback 0 of Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] originating-rp loopback 0
```

```
[SwitchB-msdp] peer 2.2.2.2 connect-interface loopback 0
[SwitchB-msdp] quit
```

Configure an MSDP peer on Loopback 0 of Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] originating-rp loopback 0
[SwitchD-msdp] peer 1.1.1.1 connect-interface loopback 0
[SwitchD-msdp] quit
```

5) Verify the configuration

You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches.

View the brief MSDP peer information on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  2.2.2.2           Up       00:10:17        ?      0          0
```

View the brief MSDP peer information on Switch D.

```
[SwitchD] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0

  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  1.1.1.1           Up       00:10:18        ?      0          0
```

To view the PIM routing information on the switches, use the **display pim routing-table** command. When Source 1 (10.110.5.100/24) sends multicast data to multicast group G (225.1.1.1), Host A joins multicast group G. By comparing the PIM routing information displayed on Switch B with that displayed on Switch D, you can see that Switch B acts now as the RP for Source 1 and Host A.

View the PIM routing information on Switch B.

```
[SwitchB] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:15:04
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
```

```
            Protocol: igmp, UpTime: 00:15:04, Expires: never

 (10.110.5.100, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT 2MSDP ACT
     UpTime: 00:46:28
     Upstream interface: Vlan-interface103
         Upstream neighbor: 10.110.2.2
         RPF prime neighbor: 10.110.2.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: pim-sm, UpTime:  - , Expires: never
```

View the PIM routing information on Switch D.

```
[SwitchD] display pim routing-table
```

No information is output on Switch D.

Host A has left multicast group G. Source 1 has stopped sending multicast data to multicast group G. When Source 2 (10.110.6.100/24) sends multicast data to G, Host B joins G. By comparing the PIM routing information displayed on Switch B with that displayed on Switch D, you can see that Switch D acts now as the RP for Source 2 and Host B.

View the PIM routing information on Switch B.

```
[SwitchB] display pim routing-table
```

No information is output on Switch B.

View the PIM routing information on Switch D.

```
[SwitchD] display pim routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:12:07
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface200
             Protocol: igmp, UpTime: 00:12:07, Expires: never

 (10.110.6.100, 225.1.1.1)
     RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT 2MSDP ACT
     UpTime: 00:40:22
     Upstream interface: Vlan-interface104
         Upstream neighbor: 10.110.4.2
         RPF prime neighbor: 10.110.4.2
```

```
Downstream interface(s) information:
Total number of downstreams: 1
    1: Vlan-interface200
        Protocol: pim-sm, UpTime:  never , Expires:  never
```

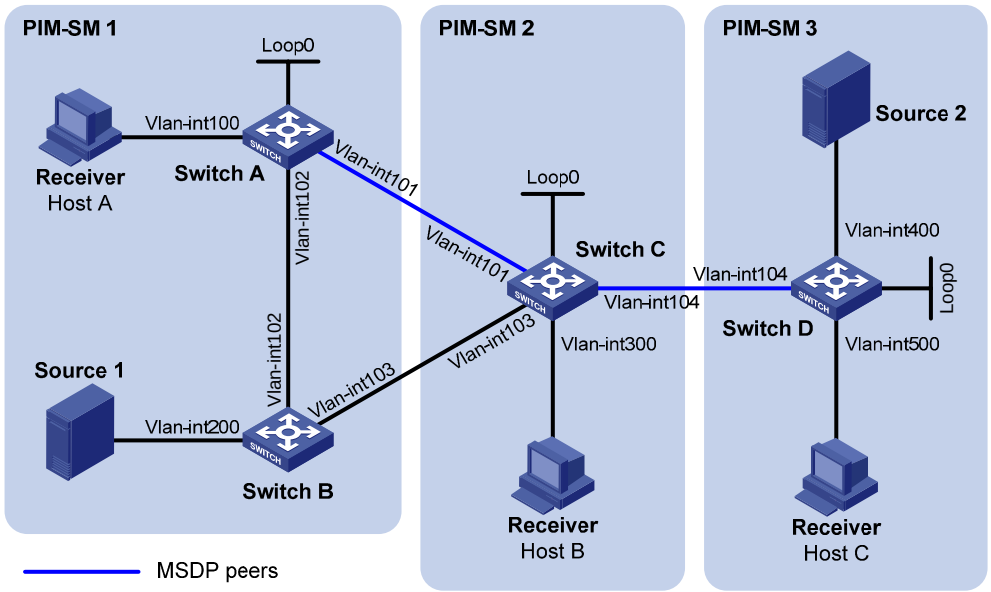
SA Message Filtering Configuration

Network requirements

- Three PIM-SM domains exist in the network, and OSPF runs within and among the domains to provide unicast routing.
- Configure respective Loopback 0 of Switch A, Switch C and Switch D as a C-BSR and C-RP in the respective PIM-SM domain.
- Set up an MSDP peering relationship between Switch A and Switch C and between Switch C and Switch D.
- Source 1 sends multicast data to multicast groups 225.1.1.0/30 and 226.1.1.0/30, and Source 2 sends multicast data to multicast group 227.1.1.0/30.
- Configure SA message filtering rules so that receivers Host A and Host B can receive only the multicast data addressed to multicast groups 225.1.1.0/30 and 226.1.1.0/30, while Host can receive only the multicast data addressed to multicast groups 226.1.1.0/30 and 227.1.1.0/30.

Network diagram

Figure 1-8 Network diagram for SA message filtering configuration (on switches)



Device	Interface	IP address	Device	Interface	IP address
Source 1	—	10.110.3.100/24	Switch C	Vlan-int300	10.110.4.1/24
Source 2	—	10.110.6.100/24		Vlan-int104	10.110.5.1/24
Switch A	Vlan-int100	10.110.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int102	10.110.2.1/24		Vlan-int103	192.168.2.2/24
	Vlan-int101	192.168.1.1/24		Loop0	2.2.2.2/32
	Loop0	1.1.1.1/32	Switch D	Vlan-int400	10.110.6.1/24
Switch B	Vlan-int200	10.110.3.1/24		Vlan-int500	10.110.7.1/24
	Vlan-int102	10.110.2.2/24		Vlan-int104	10.110.5.2/24
	Vlan-int103	192.168.2.1/24		Loop0	3.3.3.3/32

Configuration Procedure

1) Configure IP addresses and unicast routing

Configure the IP address and subnet mask for each interface as per Figure 1-8. The detailed configuration steps are omitted here.

Configure OSPF for interoperation among the switches. Ensure the network-layer interoperation within and between the PIM-SM domains and ensure dynamic update of routing information among the switches by leveraging unicast routing. The detailed configuration steps are omitted here.

2) Enable IP multicast routing, PIM-SM and IGMP, and configure a PIM domain border

On Switch A, enable IP multicast routing, enable PIM-SM on each interface, and enable IGMP on the host-side interface, VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] pim sm
[SwitchA-LoopBack0] quit
```

The configuration on Switch B, Switch C and Switch D is similar to the configuration on Switch A. The specific configuration steps are omitted here.

Configure a PIM domain border on Switch C.

```
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim bsr-boundary
[SwitchC-Vlan-interface101] quit
[SwitchC] interface vlan-interface 103
[SwitchC-Vlan-interface103] pim bsr-boundary
[SwitchC-Vlan-interface103] quit
[SwitchC] interface vlan-interface 104
[SwitchC-Vlan-interface104] pim bsr-boundary
[SwitchC-Vlan-interface104] quit
```

The configuration on Switch A, Switch B and Switch D is similar to the configuration on Switch C. The specific configuration steps are omitted here.

3) Configure C-BSRs and C-RPs

Configure Loopback 0 on Switch A as a C-BSR and a C-RP.

```
[SwitchA] pim
[SwitchA-pim] c-bsr loopback 0
[SwitchA-pim] c-rp loopback 0
```

```
[SwitchA-pim] quit
```

The configuration on Switch C and Switch D is similar to the configuration on Switch A. The specific configuration steps are omitted here.

4) Configure MSDP peers

Configure an MSDP peer on Switch A.

```
[SwitchA] msdp
[SwitchA-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchA-msdp] quit
```

Configure MSDP peers on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchC-msdp] peer 10.110.5.2 connect-interface vlan-interface 104
[SwitchC-msdp] quit
```

Configure an MSDP peer on Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] peer 10.110.5.1 connect-interface vlan-interface 104
[SwitchD-msdp] quit
```

5) Configure SA message filtering rules

Configure an SA message rule on Switch C so that Switch C will not forward SA messages for (Source 1, 225.1.1.0/30) to Switch D.

```
[SwitchC] acl number 3001
[SwitchC-acl-adv-3001] rule deny ip source 10.110.3.100 0 destination 225.1.1.0 0.0.0.3
[SwitchC-acl-adv-3001] rule permit ip source any destination any
[SwitchC-acl-adv-3001] quit
[SwitchC] msdp
[SwitchC-msdp] peer 10.110.5.2 sa-policy export acl 3001
[SwitchC-msdp] quit
```

Configure an SA message rule on Switch D so that Switch D will not create SA messages for Source 2.

```
[SwitchD] acl number 2001
[SwitchD-acl-basic-2001] rule deny source 10.110.6.100 0
[SwitchD-acl-basic-2001] quit
[SwitchD] msdp
[SwitchD-msdp] import-source acl 2001
[SwitchD-msdp] quit
```

6) Verify the configuration

View the (S, G) entries cached in the SA cache on the switches using the **display msdp sa-cache** command. For example:

View the (S, G) entries cached in the SA cache on Switch C.

```
[SwitchC] display msdp sa-cache
MSDP Source-Active Cache Information of VPN-Instance: public net
 MSDP Total Source-Active Cache - 8 entries
 MSDP matched 8 entries
```

```
(Source, Group)              Origin RP       Pro  AS     Uptime   Expires
(10.110.3.100, 225.1.1.0)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.1)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.2)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 225.1.1.3)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.0)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.1)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.2)    1.1.1.1         ?    ?      02:03:30 00:05:31
(10.110.3.100, 226.1.1.3)    1.1.1.1         ?    ?      02:03:30 00:05:31
```

View the (S, G) entries cached in the SA cache on Switch D.

```
[SwitchD] display msdp sa-cache
MSDP Source-Active Cache Information of VPN-Instance: public net
 MSDP Total Source-Active Cache - 4 entries
 MSDP matched 4 entries

(Source, Group)              Origin RP       Pro  AS     Uptime   Expires
(10.110.3.100, 226.1.1.0)    1.1.1.1         ?    ?      00:32:53 00:05:07
(10.110.3.100, 226.1.1.1)    1.1.1.1         ?    ?      00:32:53 00:05:07
(10.110.3.100, 226.1.1.2)    1.1.1.1         ?    ?      00:32:53 00:05:07
(10.110.3.100, 226.1.1.3)    1.1.1.1         ?    ?      00:32:53 00:05:07
```

Troubleshooting MSDP

MSDP Peers Stay in Down State

Symptom

The configured MSDP peers stay in the down state.

Analysis

- A TCP connection–based MSDP peering relationship is established between the local interface address and the MSDP peer after the configuration.
- The TCP connection setup will fail if there is a consistency between the local interface address and the MSDP peer address configured on the router.
- If no route is available between the MSDP peers, the TCP connection setup will also fail.

Solution

1) Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
2) Check that a unicast route is available between the two routers that will become MSDP peers to each other.
3) Verify the interface address consistency between the MSDP peers. Use the **display current-configuration** command to verify that the local interface address and the MSDP peer address of the remote router are the same.

No SA Entries in the Router's SA Cache

Symptom

MSDP fails to send (S, G) entries through SA messages.

Analysis

- The **import-source** command is used to control sending (S, G) entries through SA messages to MSDP peers. If this command is executed without the *acl-number* argument, all the (S, G) entries will be filtered off, namely no (S, G) entries of the local domain will be advertised.
- If the **import-source** command is not executed, the system will advertise all the (S, G) entries of the local domain. If MSDP fails to send (S, G) entries through SA messages, check whether the **import-source** command has been correctly configured.

Solution

1) Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
2) Check that a unicast route is available between the two routers that will become MSDP peers to each other.
3) Check configuration of the **import-source** command and its *acl-number* argument and make sure that ACL rule can filter appropriate (S, G) entries.

Inter-RP Communication Faults in Anycast RP Application

Symptom

RPs fail to exchange their locally registered (S, G) entries with one another in the Anycast RP application.

Analysis

- In the Anycast RP application, RPs in the same PIM-SM domain are configured to be MSDP peers to achieve load balancing among the RPs.
- An MSDP peer address must be different from the anycast RP address, and the C-BSR and C-RP must be configured on different devices or interfaces.
- If the **originating-rp** command is executed, MSDP will replace the RP address in the SA messages with the address of the interface specified in the command.
- When an MSDP peer receives an SA message, it performs RPF check on the message. If the MSDP peer finds that the remote RP address is the same as the local RP address, it will discard the SA message.

Solution

1) Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
2) Check that a unicast route is available between the two routers that will become MSDP peer to each other.
3) Check the configuration of the **originating-rp** command. In the Anycast RP application environment, be sure to use the **originating-rp** command to configure the RP address in the SA messages, which must be the local interface address.
4) Verify that the C-BSR address is different from the anycast RP address.

Table of Contents

1 MBGP Configuration

Note

The term “router” refers to a router or a Layer 3 switch in this document.

When configuring MBGP, go to these sections for information you are interested in:

MBGP Overview

BGP-4 is capable of carrying routing information for IPv4 only. IETF defined multiprotocol BGP extensions to carry routing information for multiple network layer protocols.

For a network, the multicast topology may be different from the unicast topology. To meet the requirement, the multiprotocol BGP extensions enable BGP to carry the unicast Network Layer Reachability Information (NLRI) and multicast NLRI separately, and the multicast NLRI is used to perform reverse path forwarding (RPF) exclusively. In this way, route selection for a destination through the unicast routing table and through the multicast routing table will have different results, ensuring normal unicast and multicast routing.

Multi-protocol BGP is defined in RFC 2858 (Multiprotocol Extensions for BGP-4).

Multi-protocol BGP for IP multicast is referred to as Multicast BGP (MBGP) for short.

Note

- This document covers configuration tasks related to multiprotocol BGP for IP multicast only. For information about BGP, refer to *BGP Configuration* in the *IP Routing Volume*.
- For information about RPF, refer to *Multicast Routing and Forwarding* in the *IP Multicast Volume*.

Protocols and Standards

- RFC2858: Multiprotocol Extensions for BGP-4
- RFC3392: Capabilities Advertisement with BGP-4
- draft-ietf-idmr-bgp-mcast-attr-00: BGP Attributes for Multicast Tree Construction

MBGP Configuration Task List

Complete the following tasks to configure MBGP:

Task		Remarks
Configuring MBGP Basic Functions		Required
Controlling Route Advertisement and Reception	Configuring MBGP Route Redistribution	Optional
	Configure Default Route Redistribution into MBGP	Optional
	Configuring MBGP Route Summarization	Optional
	Advertising a Default Route to an IPv4 MBGP Peer or Peer Group	Optional
	Configuring Outbound MBGP Route Filtering	Optional
	Configuring Inbound MBGP Route Filtering	Optional
	Configuring MBGP Route Dampening	
Configuring MBGP Route Attributes	Configuring MBGP Route Preferences	Optional
	Configuring the Default Local Preference	
	Configuring the MED Attribute	
	Configuring the Next Hop Attribute	
	Configuring the AS-PATH Attribute	
Tuning and Optimizing MBGP Networks	Configuring MBGP Soft Reset	Optional
	Configuring the Maximum Number of MBGP Routes	Optional
Configuring a Large Scale MBGP Network	Configuring IPv4 MBGP Peer Groups	Optional
	Configuring MBGP Community	Optional
	Configuring an MBGP Route Reflector	Optional

Configuring MBGP Basic Functions

Prerequisites

Before configuring MBGP, make sure neighboring nodes can access each other at the network layer.

Configuration Procedure

Follow these steps to configure MBGP basic functions:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Specify a peer or peer group and its AS number	**peer** { *group-name* | *ip-address* } **as-number** *as-number*	Required Not specified by default.
Enter IPv4 MBGP address family view	**ipv4-family multicast**	Required
Enable a peer or peer group created in IPv4 unicast view	**peer** { *group-name* | *ip-address* } **enable**	Required Not enabled by default
Specify a preferred value for routes from an IPv4 MBGP peer or peer group	**peer** { *group-name* | *ip-address* } **preferred-value** *value*	Optional The default preferred value is 0.

Controlling Route Advertisement and Reception

Prerequisites

You need to configure MBGP basic functions before configuring this task.

Configuring MBGP Route Redistribution

MBGP can advertise routing information in the local AS to neighboring ASs. It redistributes such routing information from IGP into its routing table rather than learns the information by itself.

Follow these steps to configure MBGP route redistribution:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Enable route redistribution from another routing protocol	**import-route** *protocol* [*process-id* [**med** *med-value* | **route-policy** *route-policy-name*] *]	At least one of these approaches is required. No route redistribution is configured by default.
Inject a network into the MBGP routing table	**network** *ip-address* [*mask* | *mask-length*] [**short-cut** | **route-policy** *route-policy-name*]	

Note

- The Origin attribute of routes redistributed into the MBGP routing table with the **import-route** command is Incomplete.
- The Origin attribute of routes injected into the MBGP routing table with the **network** command is IGP.
- The networks to be injected must exist in the local IP routing table, and using a route policy makes route control more flexible.

Configure Default Route Redistribution into MBGP

Using the **import-route** command cannot redistribute any default route into MBGP. This task allows you to do so.

Follow these steps to configure MBGP to redistribute a default route from another protocol:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter MBGP address family view	**ipv4-family multicast**	—
Enable route redistribution from another routing protocol	**import-route** *protocol* [*process-id* [**med** *med-value* | **route-policy** *route-policy-name*] *]	Required No route redistribution is configured by default.
Enable default route redistribution into the MBGP routing table	**default-route imported**	Required Not enabled by default

Configuring MBGP Route Summarization

To reduce the routing table size on medium and large MBGP networks, you need to configure route summarization on peers. MBGP supports two summarization modes: automatic and manual.

- Automatic summarization: Summarizes subnets redistributed from IGP. With the feature configured, MBGP advertises only summary natural networks rather than subnets. The default routes and routes injected with the **network** command are not summarized.
- Manual summarization: Summarizes MBGP local routes. A manual summary route has a higher priority than an automatic one.

Follow these steps to configure MBGP route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—

To do…		Use the command…	Remarks
Configure MBGP route summarization	Enable automatic route summarization	**summary automatic**	Required No route summarization is configured by default.
	Configure manual route summarization	**aggregate** *ip-address* { *mask* | *mask-length* } [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*] *	Choose either as needed; if both are configured, the manual route summarization takes effect.

Advertising a Default Route to an IPv4 MBGP Peer or Peer Group

Follow these steps to advertise a default route to an MBGP peer or peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Advertise a default route to an MBGP peer or peer group	**peer** { *group-name* | *ip-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	Required Not advertised by default

Note

With the **peer default-route-advertise** command executed, the router sends a default route with the next hop being itself to the specified MBGP peer or peer group, regardless of whether the default route is available in the routing table.

Configuring Outbound MBGP Route Filtering

If several filtering policies are configured, they are applied in the following sequence:

- **filter-policy export**
- **peer filter-policy export**
- **peer as-path-acl export**
- **peer ip-prefix export**
- **peer route-policy export**

Only the routes that have passed all the configured policies can be advertised.

Follow these steps to configure BGP route distribution filtering policies:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure the filtering of redistributed routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [**direct | isis** *process-id* **| ospf** *process-id* **| rip** *process-id* **| static**]	At least one of these approaches is required. No outbound route filtering is configured by default
Apply a route policy to advertisements to an IPv4 MBGP peer/peer group	**peer** { *group-name* | *peer-address* } **route-policy** *route-policy-name* **export**	
Reference an ACL to filter advertisements to an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **export**	
Reference an AS path ACL to filer route advertisements to an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **export**	
Reference an IP prefix list to filer route advertisements to an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **export**	

Configuring Inbound MBGP Route Filtering

By configuring MBGP route reception filtering policies, you can filter out unqualified routes from an MBGP peer or peer group.

If several filtering policies are configured, they are applied in the following sequence:

- **filter-policy import**
- **peer filter-policy import**
- **peer as-path-acl import**
- **peer ip-prefix import**
- **peer route-policy import**

Only the routes that have passed all the configured policies can be advertised.

Follow these steps to configure MBGP route reception filtering policies:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—

To do…	Use the command…	Remarks
Filter incoming routes using an ACL or IP prefix list	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	At least one of these approaches is required. No inbound route filtering is configured by default.
Reference a route policy to routes from an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **route-policy** *policy-name* **import**	
Reference an ACL to filter routing information from an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **import**	
Reference an AS path ACL to filter routing information from an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **as-path-acl** *as-path-acl-number* **import**	
Reference an IP prefix list to filter routing information from an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **import**	
Specify the maximum number of routes that can be received from an IPv4 MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **route-limit** *limit* [*percentage*]	Optional The number is unlimited by default.

Caution

Members of a peer group can have different route reception filtering policies from the peer group.

Configuring MBGP Route Dampening

By configuring MBGP route dampening, you can suppress unstable routes from being added to the MBGP routing table or being advertised to MBGP peers.

Follow these steps to configure BGP route dampening:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure BGP route dampening parameters	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	Required Not configured by default

Configuring MBGP Route Attributes

You can modify MBGP route attributes to affect route selection.

Prerequisites

Before configuring this task, you need to configure MBGP basic functions.

Configuring MBGP Route Preferences

You can reference a route policy to set preferences for routes matching it. Routes not matching it use the default preferences.

Follow these steps to configure MBGP route preferences:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure preferences for external, internal, local MBGP routes	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	Optional The default preferences of multicast MBGP eBGP, MBGP iBGP, and local MBGP routes are 255, 255, and 130 respectively.

Configuring the Default Local Preference

Follow these steps to configure the default local preference:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure the default local preference	**default local-preference** *value*	Optional 100 by default.

Configuring the MED Attribute

When other conditions of routes to a destination are identical, the route with the smallest MED is selected.

Follow these steps to configure the MED attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—

To do…		Use the command…	Remarks
Configure the MED attribute	Configure the default MED value	**default med** *med-value*	Optional 0 by default.
	Enable the comparison of the MED of routes from different ASs	**compare-different-as-med**	Optional Not enabled by default
	Enable the comparison of the MED of routes from each AS	**bestroute compare-med**	Optional Not enabled by default
	Enable the comparison of the MED of routes from confederation peers	**bestroute med-confederation**	Optional Not enabled by default

Configuring the Next Hop Attribute

You can use the **peer next-hop-local** command to specify the local router as the next hop of routes sent to a MBGP iBGP peer/peer group. If load balancing is configured, the router specifies itself as the next hop of route advertisements to the multicast iBGP peer/peer group regardless of whether the **peer next-hop-local** command is configured.

In a “third party next hop" network, that is, the local router has two multicast eBGP peers in a broadcast network, the router does not specify itself as the next hop of routing information sent to the eBGP peers unless the **peer next-hop-local** command is configured.

Follow these steps to specify the router as the next hop of routes sent to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Specify the router as the next hop of routes sent to a peer/peer group	**peer** { *group-name* | *ip-address* } **next-hop-local**	Optional By default, the next hop of routes sent to a MBGP eBGP peer/peer group is the advertising router, while that of routes sent to a MBGP iBGP peer/peer group is not.

Configuring the AS-PATH Attribute

In general, MBGP checks whether the AS_PATH attribute of a route from a peer contains the local AS number. If yes, it discards the route to avoid routing loops.

Follow these steps to configure the AS-PATH attribute:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter BGP view		**bgp** *as-number*	—
Enter IPv4 MBGP address family view		**ipv4-family multicast**	—
Configure the AS_PATH attribute	Specify the maximum number of times the local AS number can appear in routes from the peer/peer group	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	Optional By default, the local AS number can not appear in routes from a peer/peer group.
	Disable BGP from considering the AS_PATH during best route selection	**bestroute as-path-neglect**	Optional By default, BGP considers AS_PATH during best route selection.
	Configure updates to a peer/peer group to not keep private AS numbers	**peer** { *group-name* | *ip-address* } **public-as-only**	Optional By default, BGP updates carry private AS numbers.

Tuning and Optimizing MBGP Networks

This task involves resetting MBGP connections and configuring load balancing.

Prerequisites

You need to configure BGP basic functions before configuring this task.

Configuring MBGP Soft Reset

After modifying a route selection policy, you have to reset MBGP connections to make it take effect, causing short time disconnections.

After the route-refresh capability is enabled on all MBGP routers in a network, when a route selection policy is modified on a router, the local router can perform dynamic route updates without tearing down MBGP connections.

If the peer does not support route-refresh, you can save all route updates from the peer. When the route selection policy changes, you can refresh the MBGP routing table and apply the new policy without tearing down MBGP connections.

Soft reset through route-refresh

If the peer is enabled with route-refresh, when the MBGP route selection policy is modified on a router, the router advertises a route-refresh message to its MBGP peers, which resend their routing information to the router after receiving the message. Therefore, the local router can perform dynamic route update and apply the new policy without tearing down MBGP connections.

Follow these steps to configure MBGP soft reset through route-refresh:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enable BGP route refresh for a peer/peer group	**peer** { *group-name* | *ip-address* } **capability-advertise route-refresh**	Optional Enabled by default

Perfom a manual soft reset

If the peer does not support route-refresh, you can use the **peer keep-all-routes** command to save all the route updates from the peer, and then use the **refresh bgp ipv4 multicast** command to soft-reset MBGP connections to refresh the MBGP routing table and apply the new policy without tearing down MBGP connections.

Follow these steps to configure MBGP manual soft reset

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Disable BGP route-refresh and multi-protocol extensions for a peer/peer group	**peer** { *group-name* | *ip-address* } **capability-advertise conventional**	Optional Enabled by default
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Keep all original routes from a peer/peer group regardless of whether they pass the inbound filtering policies	**peer** { *group-name* | *ip-address* } **keep-all-routes**	Required Not kept by default
Return to user view	**return**	—
Soft-reset MBGP connections manually	**refresh bgp ipv4 multicast** { **all** | *ip-address* | **group** *group-name* | **external** | **internal** } { **export** | **import** }	Optional

Configuring the Maximum Number of MBGP Routes for Load Balancing

Follow these steps to configure the number of MBGP routes for load balancing:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure the maximum number of MBGP routes for load balancing	**balance** *number*	Required Not configured by default.

Configuring a Large Scale MBGP Network

Prerequisites

Before configuring this task, you need to make peering nodes accessible to each other at the network layer.

Configuring IPv4 MBGP Peer Groups

In a large-scale network, configuration and maintenance become difficult due to large numbers of MBGP peers. You can configure peer groups to make management easier and improve route distribution efficiency.

Follow these steps to configure an IPv4 MBGP peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Create a BGP peer group	**group** *group-name* [**external** | **internal**]	Required Not created by default.
Add a peer into the peer group	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	Required No peer is added by default.
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Enable the IPv4 unicast peer group	**peer** *group-name* **enable**	Required
Add an IPv4 MBGP peer to the peer group	**peer** *ip-address* **group** *group-name*	Required Not configured by default.

Caution

- To configure an MBGP peer group, you need to enable the corresponding IPv4 BGP unicast peer group in IPv4 MBGP address family view.
- Before adding an MBGP peer to an MBGP peer group, you need to add the corresponding IPv4 unicast peer to the IPv4 BGP peer group.

Configuring MBGP Community

The community attribute can be advertised between MBGP peers in different ASs. Routers in the same community share the same policy.

You can reference a route policy to modify the community attribute for routes sent to a peer. In addition, you can define extended community attributes as needed.

Follow these steps to configure MBGP community:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter BGP view		**bgp** *as-number*	—
Enter IPv4 MBGP address family view		**ipv4-family multicast**	—
Advertise the community attribute to an MBGP peer/peer group	Advertise the community attribute to an MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **advertise-community**	Required Not configured by default.
	Advertise the extended community attribute to an MBGP peer/peer group	**peer** { *group-name* | *ip-address* } **advertise-ext-community**	
Apply a route policy to routes advertised to an MBGP peer/peer group		**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	Required Not configured by default.

Caution

- When configuring MBGP community, you need to reference a route policy to define the specific community attributes, and apply the route policy for route advertisement.
- For route policy configuration, refer to *Route Policy Configuration* in the *IP Routing Volume*.

Configuring an MBGP Route Reflector

To guarantee the connectivity between multicast iBGP peers in an AS, you need to make them fully meshed. But this becomes unpractical when there are large numbers of multicast iBGP peers. Configuring route reflectors can solve this problem.

Follow these steps to configure an MBGP route reflector:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv4 MBGP address family view	**ipv4-family multicast**	—
Configure the router as a route reflector and specify an MBGP peer/peer group as its client	**peer** { *group-name* | *peer-address* } **reflect-client**	Required Not configured by default.
Enable route reflection between clients	**reflect between-clients**	Optional Enabled by default.
Configure the cluster ID of the route reflector	**reflector cluster-id** *cluster-id*	Optional By default, a route reflector uses its router ID as the cluster ID.

Caution

- In general, it is not required that clients of a route reflector be fully meshed. The route reflector forwards routing information between clients. If clients are fully meshed, you can disable route reflection between clients to reduce routing costs.
- In general, a cluster has only one route reflector, and the router ID of the route reflector is used to identify the cluster. You can configure multiple route reflectors to improve network stability. In this case, you need to specify the same cluster ID for these route reflectors to avoid routing loops.

Displaying and Maintaining MBGP

Displaying MBGP

To do…	Use the command…	Remarks
Display the IPv4 MBGP routing table	**display ip multicast routing-table** [**verbose**]	Available in any view
Display the IPv4 MBGP routing information matching the specified destination IP address	**display ip multicast routing-table** *ip-address* [*mask-length* | *mask*] [**longer-match**] [**verbose**]	Available in any view
Display MBGP peer group information	**display bgp multicast group** [*group-name*]	Available in any view
Display the advertised networks	**display bgp multicast network**	Available in any view
Display AS path information	**display bgp multicast paths** [*as-regular-expression*]	Available in any view
Display MBGP peer/peer group information	**display bgp multicast peer** [*ip-address*] [**verbose**]	Available in any view
Display MBGP routing information	**display bgp multicast routing-table** [*ip-address* [{ *mask* | *mask-length* } [**longer-prefixes**]]]	Available in any view
Display MBGP routing information matching the AS path ACL	**display bgp multicast routing-table as-path-acl** *as-path-acl-number*	Available in any view
Display MBGP CIDR routing information	**display bgp multicast routing-table cidr**	Available in any view
Display MBGP routing information matching the specified BGP community	**display bgp multicast routing-table community** [aa:nn&<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**] * [**whole-match**]	Available in any view
Display MBGP routing information matching an MBGP community list	**display bgp multicast routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	Available in any view
Display MBGP dampened routing information	**display bgp multicast routing-table dampened**	Available in any view

To do…	Use the command…	Remarks
Display MBGP dampening parameter information	**display bgp multicast routing-table dampening parameter**	Available in any view
Display MBGP routing information originating from different ASs	**display bgp multicast routing-table different-origin-as**	Available in any view
Display IPv4 MBGP routing flap statistics	**display bgp multicast routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *ip-address* [{ *mask* | *mask-length* } [**longer-match**]]]	Available in any view
Display IPv4 MBGP routing information sent to or received from an MBGP peer	**display bgp multicast routing-table peer** *ip-address* { **advertised-routes** | **received-routes** } [*network-address* [*mask* | *mask-length*] | **statistic**]	Available in any view
Display IPv4 MBGP routing information matching an AS regular expression	**display bgp multicast routing-table regular-expression** *as-regular-expression*	Available in any view
Display IPv4 MBGP routing statistics	**display bgp multicast routing-table statistic**	Available in any view

Resetting MBGP Connections

To do…	Use the command…	Remarks
Reset specified MBGP connections	**reset bgp ipv4 multicast** { **all** | *as-number* | *ip-address* | **group** *group-name* | **external** | **internal** }	Available in user view

Clearing MBGP Information

To do…	Use the command…	Remarks
Clear dampened routing information and release suppressed routes	**reset bgp ipv4 multicast dampening** [*ip-address* [*mask* | *mask-length*]]	Available in user view
Clear MBGP route flap statistics	**reset bgp ipv4 multicast flap-info** [**regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number* | *ip-address* [*mask* | *mask-length*]]	Available in user view

MBGP Configuration Example

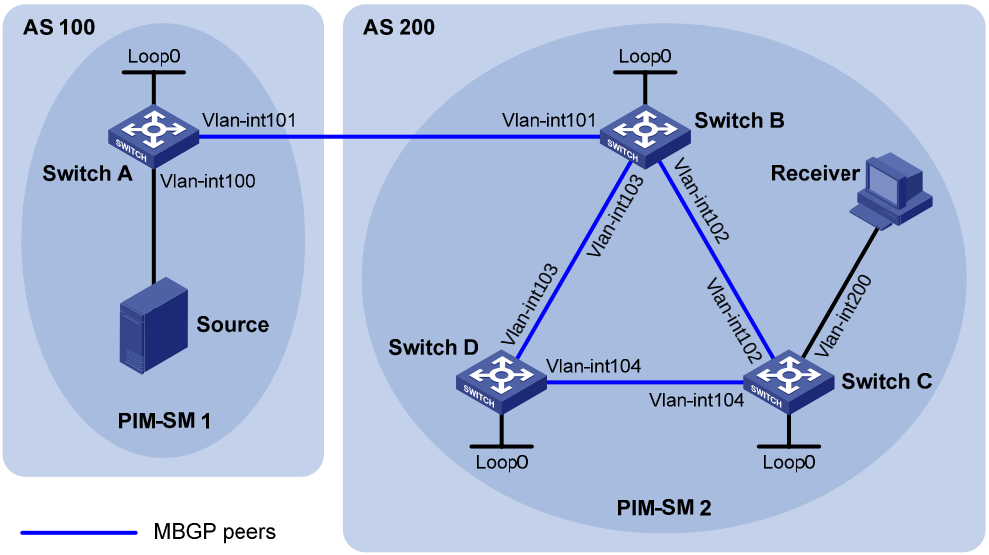
Network requirements

As shown in the following figure:

- PIM-SM 1 is in AS 100 and PIM-SM 2 is in AS 200. OSPF is the IGP in the two ASs, and MBGP runs between the two ASs to exchange multicast route information.
- The multicast source belongs to PIM-SM 1, and the receiver belongs to PIM-SM 2.

- It is required that the respective Loopback 0 of Switch A and Switch B be configured as the C-BSR and C-RP of the respective PIM-SM domains.
- Switch A and Switch B establishes an MSDP peer relationship through MBGP.

Figure 1-1 Network diagram for MBGP configuration



Device	Interface	IP address	Device	Interface	IP address
Source	-	10.110.1.100/24	Switch C	Vlan-int200	10.110.2.1/24
Switch A	Vlan-int100	10.110.1.1/24		Vlan-int102	192.168.2.2/24
	Vlan-int101	192.168.1.1/24		Vlan-int104	192.168.4.1/24
	Loop0	1.1.1.1/32		Loop0	3.3.3.3/32
Switch B	Vlan-int101	192.168.1.2/24	Switch D	Vlan-int103	192.168.3.2/24
	Vlan-int102	192.168.2.1/24		Vlan-int104	192.168.4.2/24
	Vlan-int103	192.168.3.1/24		Loop0	4.4.4.4/32
	Loop0	2.2.2.2/32			

Configuration procedure

1) Configure IP addresses for interfaces as shown in the above figure (omitted).
2) Configure OSPF (omitted).
3) Enable IP multicast routing, PIM-SM and IGMP, and configure a PIM-SM domain border.

Enable IP multicast routing on Switch A, and enable PIM-SM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
```

The configuration on Switch B and Switch D is similar to the configuration on Switch A.

Enable IP multicast routing on Switch C, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 200.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 102
```

```
[SwitchC-Vlan-interface102] pim sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 104
[SwitchC-Vlan-interface104] pim sm
[SwitchC-Vlan-interface104] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim sm
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] quit
```

Configure a PIM domain border on Switch A.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim bsr-boundary
[SwitchA-Vlan-interface101] quit
```

Configure a PIM domain border on Switch B.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit
```

4) Configure Loopback 0 and the position of C-BSR, and C-RP.

Configure Loopback 0 and configure it as the C-BSR and C-RP on Switch A.

```
[SwitchA] interface loopback 0
[SwitchA-LoopBack0] ip address 1.1.1.1 32
[SwitchA-LoopBack0] pim sm
[SwitchA-LoopBack0] quit
[SwitchA] pim
[SwitchA-pim] c-bsr loopback 0
[SwitchA-pim] c-rp loopback 0
[SwitchA-pim] quit
```

Configure Loopback 0 and configure it as the C-BSR and C-RP on Switch B.

```
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] ip address 2.2.2.2 32
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
[SwitchB] pim
[SwitchB-pim] c-bsr loopback 0
[SwitchB-pim] c-rp loopback 0
[SwitchB-pim] quit
```

5) Configure BGP, specify the MBGP peer and enable direct route redistribution.

On Switch A, configure the MBGP peer and enable direct route redistribution.

```
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 192.168.1.2 as-number 200
[SwitchA-bgp] import-route direct
[SwitchA-bgp] ipv4-family multicast
[SwitchA-bgp-af-mul] peer 192.168.1.2 enable
[SwitchA-bgp-af-mul] import-route direct
```

```
[SwitchA-bgp-af-mul] quit
[SwitchA-bgp] quit
```

On Switch B, configure the MBGP peer and enable route redistribution from OSPF.

```
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 192.168.1.1 as-number 100
[SwitchB-bgp] import-route ospf 1
[SwitchB-bgp] ipv4-family multicast
[SwitchB-bgp-af-mul] peer 192.168.1.1 enable
[SwitchB-bgp-af-mul] import-route ospf 1
[SwitchB-bgp-af-mul] quit
[SwitchB-bgp] quit
```

6) Configure MSDP peer

Specify the MSDP peer on Switch A.

```
[SwitchA] msdp
[SwitchA-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchA-msdp] quit
```

Specify the MSDP peer on Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchB-msdp] quit
```

7) Verify the configuration

You can use the **display bgp multicast peer** command to display MBGP peers on a switch. For example, display MBGP peers on Switch B.

```
[SwitchB] display bgp multicast peer

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  192.168.1.1 4   100       56       56     0       0 00:40:54 Established
```

You can use the **display msdp brief** command to display MSDP peers on a switch. For example, display brief information about MSDP peers on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information of VPN-Instance: public net
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1            0            0            0            0
  Peer's Address    State    Up/Down time    AS     SA Count   Reset Count
  192.168.1.1       Up       00:07:17        100    1          0
```

Table of Contents

1 IGMP Snooping Configuration

When configuring IGMP Snooping, go to the following sections for information you are interested in:

- IGMP Snooping Overview
- IGMP Snooping Configuration Task List
- Displaying and Maintaining IGMP Snooping
- IGMP Snooping Configuration Examples
- Troubleshooting IGMP Snooping Configuration

IGMP Snooping Overview

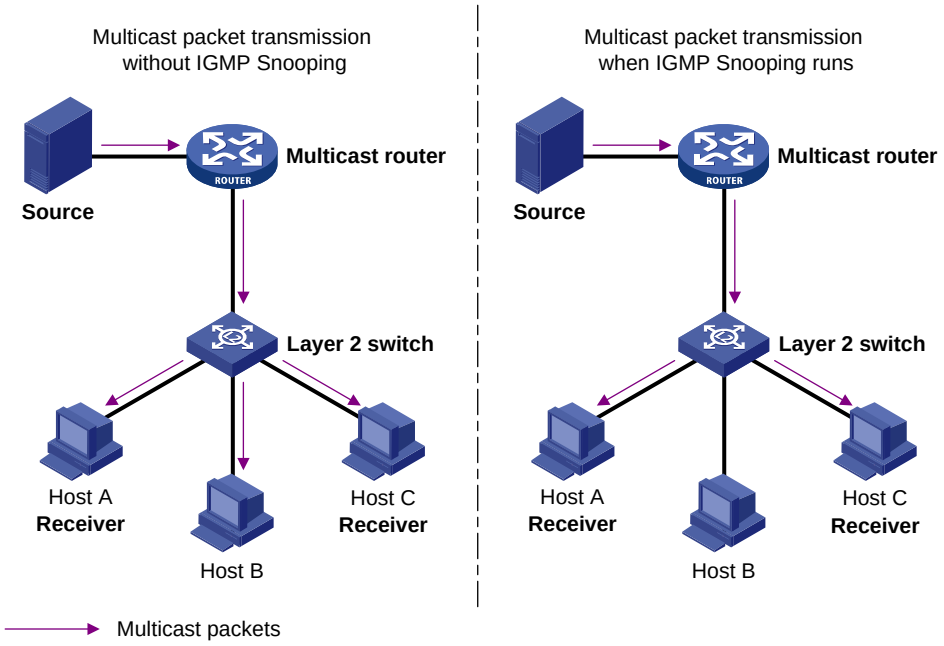
Internet Group Management Protocol Snooping (IGMP Snooping) is a multicast constraining mechanism that runs on Layer 2 devices to manage and control multicast groups.

Principle of IGMP Snooping

By analyzing received IGMP messages, a Layer 2 device running IGMP Snooping establishes mappings between ports and multicast MAC addresses and forwards multicast data based on these mappings.

As shown in Figure 1-1, when IGMP Snooping is not running on the switch, multicast packets are broadcast to all devices at Layer 2. When IGMP Snooping is running on the switch, multicast packets for known multicast groups are multicast to the receivers, rather than broadcast to all hosts, at Layer 2.

Figure 1-1 Before and after IGMP Snooping is enabled on the Layer 2 device



IGMP Snooping forwards multicast data to only the receivers requiring it at Layer 2. It brings the following advantages:

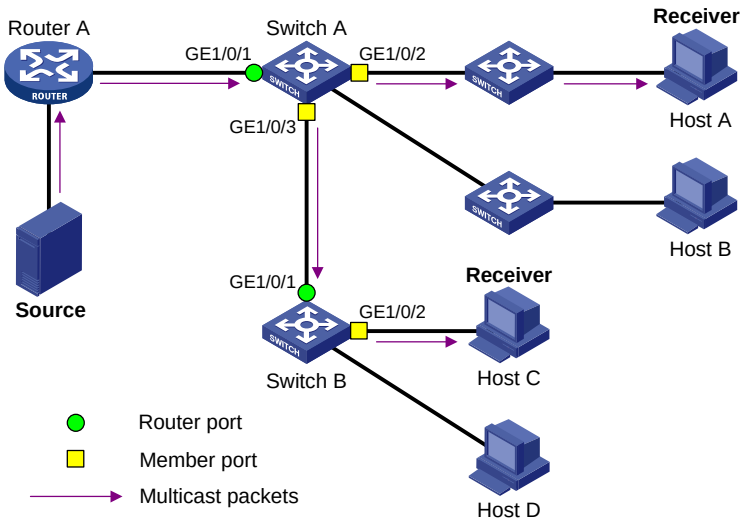
- Reducing Layer 2 broadcast packets, thus saving network bandwidth.
- Enhancing the security of multicast traffic.
- Facilitating the implementation of per-host accounting.

Basic Concepts in IGMP Snooping

IGMP Snooping related ports

As shown in Figure 1-2, Router A connects to the multicast source, IGMP Snooping runs on Switch A and Switch B, Host A and Host C are receiver hosts (namely, multicast group members).

Figure 1-2 IGMP Snooping related ports



Ports involved in IGMP Snooping, as shown in Figure 1-2, are described as follows:

- Router port: A router port is a port on an Ethernet switch that leads switch towards a Layer 3 multicast device (DR or IGMP querier). In the figure, GigabitEthernet 1/0/1 of Switch A and GigabitEthernet 1/0/1 of Switch B are router ports. The switch registers all its local router ports in its router port list.
- Member port: A member port is a port on an Ethernet switch that leads the switch towards multicast group members. In the figure, GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 of Switch A and GigabitEthernet 1/0/2 of Switch B are member ports. The switch registers all the member ports on the local device in its IGMP Snooping forwarding table.

Note

- Whenever mentioned in this document, a router port is a port on the switch that leads the switch to a Layer 3 multicast device, rather than a port on a router.
- Unless otherwise specified, router/member ports mentioned in this document include static and dynamic ports.
- An IGMP-snooping-enabled switch deems that all its ports on which IGMP general queries with the source IP address other than 0.0.0.0 or PIM hello messages are received to be dynamic router ports. For details about PIM hello messages, see *PIM Configuration* of the *IP Multicast Volume*.

Aging timers for dynamic ports in IGMP Snooping and related messages and actions

Table 1-1 Aging timers for dynamic ports in IGMP Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Dynamic router port aging timer	For each dynamic router port, the switch sets a timer initialized to the dynamic router port aging time.	IGMP general query of which the source address is not 0.0.0.0 or PIM hello	The switch removes this port from its router port list.
Dynamic member port aging timer	When a port dynamically joins a multicast group, the switch sets a timer for the port, which is initialized to the dynamic member port aging time.	IGMP membership report	The switch removes this port from the IGMP Snooping forwarding table.

Note

The port aging mechanism of IGMP Snooping works only for dynamic ports; a static port will never age out.

How IGMP Snooping Works

A switch running IGMP Snooping performs different actions when it receives different IGMP messages, as follows:

Caution

The description about adding or deleting a port in this section is only for a dynamic port. Static ports can be added or deleted only through the corresponding configurations. For details, see Configuring Static Ports.

When receiving a general query

The IGMP querier periodically sends IGMP general queries to all hosts and routers (224.0.0.1) on the local subnet to find out whether active multicast group members exist on the subnet.

Upon receiving an IGMP general query, the switch forwards it through all ports in the VLAN except the receiving port and performs the following to the receiving port:

- If the receiving port is a dynamic router port existing in its router port list, the switch resets the aging timer of this dynamic router port.
- If the receiving port is not a dynamic router port existing in its router port list, the switch adds it into its router port list and sets an aging timer for this dynamic router port.

When receiving a membership report

A host sends an IGMP report to the IGMP querier in the following circumstances:

- Upon receiving an IGMP query, a multicast group member host responds with an IGMP report.
- When intended to join a multicast group, a host sends an IGMP report to the IGMP querier to announce that it is interested in the multicast information addressed to that group.

Upon receiving an IGMP report, the switch forwards it through all the router ports in the VLAN, resolves the address of the reported multicast group, and performs the following:

- If no forwarding table entry exists for the reported group, the switch creates an entry, adds the port as a dynamic member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported group, but the port is not included in the outgoing port list for that group, the switch adds the port as a dynamic member port to the outgoing port list, and starts an aging timer for that port.
- If a forwarding table entry exists for the reported group and the port is included in the outgoing port list, which means that this port is already a dynamic member port, the switch resets the aging timer for that port.

Note

A switch does not forward an IGMP report through a non-router port. The reason is as follows: Due to the IGMP report suppression mechanism, if the switch forwards a report message through a member port, all the attached hosts listening to the reported multicast address will suppress their own reports upon receiving this report, and this will prevent the switch from knowing whether the reported multicast group still has active members attached to that port.

For the description of IGMP report suppression mechanism, refer to *IGMP Configuration* in the *IP Multicast Volume*.

When receiving a leave message

When an IGMPv1 host leaves a multicast group, the host does not send an IGMP leave message, so the switch cannot know immediately that the host has left the multicast group. However, as the host stops sending IGMP reports as soon as it leaves a multicast group, the switch deletes the forwarding entry for the dynamic member port corresponding to the host from the forwarding table when its aging timer expires.

When an IGMPv2 or IGMPv3 host leaves a multicast group, the host sends an IGMP leave message to the multicast router.

When the switch receives an IGMP leave message on a dynamic member port, the switch first checks whether a forwarding table entry for the group address in the message exists, and, if one exists, whether the outgoing port list contains the port.

- If the forwarding table entry does not exist or if the outgoing port list does not contain the port, the switch discards the IGMP leave message instead of forwarding it to any port.
- If the forwarding table entry exists and the outgoing port list contains the port, the switch forwards the leave message to all router ports in the native VLAN. Because the switch does not know whether any other hosts attached to the port are still listening to that group address, the switch

does not immediately remove the port from the outgoing port list of the forwarding table entry for that group; instead, it resets the aging timer for the port.

Upon receiving the IGMP leave message from a host, the IGMP querier resolves the multicast group address in the message and sends an IGMP group-specific query to that multicast group through the port that received the leave message. Upon receiving the IGMP group-specific query, the switch forwards it through all its router ports in the VLAN and all member ports for that multicast group, and performs the following to the port on which it received the IGMP leave message:

- If any IGMP report in response to the group-specific query is received on the port (suppose it is a dynamic member port) before its aging timer expires, this means that some host attached to the port is receiving or expecting to receive multicast data for that multicast group. The switch resets the aging timer of the port.
- If no IGMP report in response to the group-specific query is received on the port before its aging timer expires, this means that no hosts attached to the port are still listening to that group address: the switch removes the port from the outgoing port list of the forwarding table entry for that multicast group when the aging timer expires.

Processing of Multicast Protocol Messages

With Layer 3 multicast routing enabled, an IGMP Snooping switch processes multicast protocol messages differently under different conditions, specifically as follows:

1) If only IGMP is enabled, or both IGMP and PIM are enabled on the switch, the switch handles multicast protocol messages in the normal way.
2) In only PIM is enabled on the switch:

- The switch broadcasts IGMP messages as unknown messages in the VLAN.
- Upon receiving a PIM hello message, the switch will maintain the corresponding dynamic router port.

3) When IGMP is disabled on the switch:

- If PIM is disabled, the switch deletes all its dynamic member ports and dynamic router ports.
- If PIM is enabled, the switch deletes only its dynamic member ports without deleting its dynamic router ports.

Note

On a switch with Layer-3 multicast routing enabled, use the **display igmp group port-info** command to view Layer-2 port information.

For details about the **display igmp group port-info** command, refer to *IGMP Commands* in the *IP Multicast Volume*.

---

4) When PIM is disabled on the switch:

- If IGMP is disabled, the switch deletes all its dynamic router ports.
- If IGMP is enabled, the switch maintains all its dynamic member ports and dynamic router ports.

Protocols and Standards

IGMP Snooping is documented in:

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

IGMP Snooping Configuration Task List

Complete these tasks to configure IGMP Snooping:

Task		Remarks
Configuring Basic Functions of IGMP Snooping	Enabling IGMP Snooping	Required
	Configuring the Version of IGMP Snooping	Optional
Configuring IGMP Snooping Port Functions	Configuring Aging Timers for Dynamic Ports	Optional
	Configuring Static Ports	Optional
	Configuring Simulated Joining	Optional
	Configuring Fast Leave Processing	Optional
Configuring IGMP Snooping Querier	Enabling IGMP Snooping Querier	Optional
	Configuring IGMP Queries and Responses	Optional
	Configuring Source IP Address of IGMP Queries	Optional
Configuring an IGMP Snooping Policy	Configuring a Multicast Group Filter	Optional
	Configuring Multicast Source Port Filtering	Optional
	Configuring the Function of Dropping Unknown Multicast Data	Optional
	Configuring IGMP Report Suppression	Optional
	Configuring Maximum Multicast Groups that Can Be Joined on a Port	Optional
	Configuring Multicast Group Replacement	Optional

Note

- Configurations made in IGMP Snooping view are effective for all VLANs, while configurations made in VLAN view are effective only for ports belonging to the current VLAN. For a given VLAN, a configuration made in IGMP Snooping view is effective only if the same configuration is not made in VLAN view.
- Configurations made in IGMP Snooping view are effective for all ports; configurations made in Ethernet port view are effective only for the current port; configurations made in Layer 2 aggregate port view are effect only for the current port; configurations made in port group view are effective only for all the ports in the current port group. For a given port, a configuration made in IGMP Snooping view is effective only if the same configuration is not made in Ethernet port view, Layer 2 aggregate port view or port group view.
- For IGMP Snooping, configurations made on a Layer 2 aggregate port do not interfere with configurations made on its member ports, nor do they take part in aggregation calculations; configurations made on a member port of the aggregate group will not take effect until it leaves the aggregate group.

Configuring Basic Functions of IGMP Snooping

Configuration Prerequisites

Before configuring the basic functions of IGMP Snooping, complete the following task:

- Configure the corresponding VLANs.

Before configuring the basic functions of IGMP Snooping, prepare the following data:

- Version of IGMP Snooping.

Enabling IGMP Snooping

Follow these steps to enable IGMP Snooping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IGMP Snooping globally and enter IGMP-Snooping view	**igmp-snooping**	Required Disabled by default
Return to system view	**quit**	—
Enter VLAN view	**vlan** *vlan-id*	—
Enable IGMP Snooping in the VLAN	**igmp-snooping enable**	Required Disabled by default

Note

- IGMP Snooping must be enabled globally before it can be enabled in a VLAN.
- After enabling IGMP Snooping in a VLAN, you cannot enable IGMP and/or PIM on the corresponding VLAN interface.
- When you enable IGMP Snooping in a specified VLAN, this function takes effect for the ports in this VLAN only.

Configuring the Version of IGMP Snooping

By configuring an IGMP Snooping version, you actually configure the version of IGMP messages that IGMP Snooping can process.

- IGMP Snooping version 2 can process IGMPv1 and IGMPv2 messages, but not IGMPv3 messages, which will be flooded in the VLAN.
- IGMP Snooping version 3 can process IGMPv1, IGMPv2 and IGMPv3 messages.

Follow these steps to configure the version of IGMP Snooping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—

To do...	Use the command...	Remarks
Configure the version of IGMP Snooping	**igmp-snooping version** *version-number*	Optional Version 2 by default

Caution

If you switch IGMP Snooping from version 3 to version 2, the system will clear all IGMP Snooping forwarding entries from dynamic joins, and will:

- Keep forwarding entries for version 3 static (*, G) joins;
- Clear forwarding entries from version 3 static (S, G) joins, which will be restored when IGMP Snooping is switched back to version 3.

For details about static joins, Refer to Configuring Static Ports.

Configuring IGMP Snooping Port Functions

Configuration Prerequisites

Before configuring IGMP Snooping port functions, complete the following tasks:

- Enable IGMP Snooping in the VLAN or enable IGMP on the VLAN interface
- Configure the corresponding port groups.

Before configuring IGMP Snooping port functions, prepare the following data:

- Aging time of dynamic router ports,
- Aging time of dynamic member ports, and
- Multicast group and multicast source addresses

Configuring Aging Timers for Dynamic Ports

If the switch receives no IGMP general queries or PIM hello messages on a dynamic router port, the switch removes the port from the router port list when the aging timer of the port expires.

If the switch receives no IGMP reports for a multicast group on a dynamic member port, the switch removes the port from the outgoing port list of the forwarding table entry for that multicast group when the aging timer of the port for that group expires.

If multicast group memberships change frequently, you can set a relatively small value for the dynamic member port aging timer, and vice versa.

Configuring aging timers for dynamic ports globally

Follow these steps to configure aging timers for dynamic ports globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Configure dynamic router port aging time	**router-aging-time** *interval*	Optional 105 seconds by default

To do...	Use the command...	Remarks
Configure dynamic member port aging time	**host-aging-time** *interval*	Optional 260 seconds by default

Configuring aging timers for dynamic ports in a VLAN

Follow these steps to configure aging timers for dynamic ports in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure dynamic router port aging time	**igmp-snooping router-aging-time** *interval*	Optional 105 seconds by default
Configure dynamic member port aging time	**igmp-snooping host-aging-time** *interval*	Optional 260 seconds by default

Configuring Static Ports

If all the hosts attached to a port are interested in the multicast data addressed to a particular multicast group or the multicast data that a particular multicast source sends to a particular group, you can configure static (*, G) or (S, G) joining on that port, namely configure the port as a group-specific or source-and-group-specific static member port.

You can configure a port of a switch to be a static router port, through which the switch can forward all the multicast traffic it received.

Follow these steps to configure static ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure the port(s) as static member port(s)	**igmp-snooping static-group** *group-address* [**source-ip** *source-address*] **vlan** *vlan-id*	Required No static member ports by default
Configure the port(s) as static router port(s)	**igmp-snooping static-router-port vlan** *vlan-id*	Required No static router ports by default

Note

- A static (S, G) joining can take effect only if a valid multicast source address is specified and IGMP Snooping version 3 is currently running.
- A static member port does not respond to queries from the IGMP querier; when static (*, G) or (S, G) joining is enabled or disabled on a port, the port does not send an unsolicited IGMP report or an IGMP leave message.
- If IGMP is enabled on the virtual interface of a VLAN on a switch that supports both IGMP Snooping and IGMP and you want a port in that VLAN to be a static multicast group or source-group member port, in addition to configuring the port as a static member port, you need to use the **igmp static-group** command to configure the VLAN interface to be a static member of the multicast group or source and group. For details of the **igmp static-group** command, refer to *IGMP Commands* in the *IP Multicast Volume*.
- Static member ports and static router ports never age out. To remove such a port, you need to use the corresponding **undo** command.

Configuring Simulated Joining

Generally, a host running IGMP responds to IGMP queries from the IGMP querier. If a host fails to respond due to some reasons, the multicast router may deem that no member of this multicast group exists on the network segment, and therefore will remove the corresponding forwarding path.

To avoid this situation from happening, you can enable simulated joining on a port of the switch, namely configure the port as a simulated member host for a multicast group. When receiving an IGMP query, the simulated host gives a response. Thus, the switch can continue receiving multicast data.

A simulated host acts like a real host, as follows:

- When a port is configured as a simulated member host, the switch sends an unsolicited IGMP report through that port.
- After a port is configured as a simulated member host, the switch responds to IGMP general queries by sending IGMP reports through that port.
- When the simulated joining function is disabled on a port, the switch sends an IGMP leave message through that port.

Follow these steps to configure simulated joining:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure simulated (*, G) or (S, G) joining	**igmp-snooping host-join** *group-address* [**source-ip** *source-address*] **vlan** *vlan-id*	Required Disabled by default

Note

- Each simulated host is equivalent to an independent host. For example, when receiving an IGMP query, the simulated host corresponding to each configuration responds respectively.
- Unlike a static member port, a port configured as a simulated member host will age out like a dynamic member port.

Configuring Fast Leave Processing

The fast leave processing feature allows the switch to process IGMP leave messages in a fast way. With the fast leave processing feature enabled, when receiving an IGMP leave message on a port, the switch immediately removes that port from the outgoing port list of the forwarding table entry for the indicated group. Then, when receiving IGMP group-specific queries for that multicast group, the switch will not forward them to that port.

In VLANs where only one host is attached to each port, fast leave processing helps improve bandwidth and resource usage.

Configuring fast leave processing globally

Follow these steps to configure fast leave processing globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Enable fast leave processing	**fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Configuring fast leave processing on a port or a group of ports

Follow these steps to configure fast leave processing on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable fast leave processing	**igmp-snooping fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Caution

If fast leave processing is enabled on a port to which more than one host is attached, when one host leaves a multicast group, the other hosts attached to the port and interested in the same multicast group will fail to receive multicast data for that group.

Configuring IGMP Snooping Querier

Configuration Prerequisites

Before configuring IGMP Snooping querier, complete the following task:

- Enable IGMP Snooping in the VLAN.

Before configuring IGMP Snooping querier, prepare the following data:

- IGMP general query interval,
- IGMP last-member query interval,
- Maximum response time to IGMP general queries,
- Source address of IGMP general queries, and
- Source address of IGMP group-specific queries.

Enabling IGMP Snooping Querier

In an IP multicast network running IGMP, a multicast router or Layer 3 multicast switch is responsible for sending IGMP general queries, so that all Layer 3 multicast devices can establish and maintain multicast forwarding entries, thus to forward multicast traffic correctly at the network layer. This router or Layer 3 switch is called IGMP querier.

However, a Layer 2 multicast switch does not support IGMP, and therefore cannot send general queries by default. By enabling IGMP Snooping on a Layer 2 switch in a VLAN where multicast traffic needs to be Layer-2 switched only and no multicast routers are present, the Layer 2 switch will act as the IGMP Snooping querier to send IGMP queries, thus allowing multicast forwarding entries to be established and maintained at the data link layer.

Follow these steps to enable IGMP Snooping querier:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Enable IGMP Snooping querier	**igmp-snooping querier**	Required Disabled by default

Caution

It is meaningless to configure an IGMP Snooping querier in a multicast network running IGMP. Although an IGMP Snooping querier does not take part in IGMP querier elections, it may affect IGMP querier elections because it sends IGMP general queries with a low source IP address.
For details about IGMP querier, see *IGMP Configuration* of the *IP Multicast Volume*.

Configuring IGMP Queries and Responses

You can tune the IGMP general query interval based on actual condition of the network.

Upon receiving an IGMP query (general query or group-specific query), a host starts a timer for each multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time (the host obtains the value of the maximum response time from the Max Response Time field in the IGMP query it received). When the timer value comes down to 0, the host sends an IGMP report to the corresponding multicast group.

An appropriate setting of the maximum response time for IGMP queries allows hosts to respond to queries quickly and avoids bursts of IGMP traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expire simultaneously.

- For IGMP general queries, you can configure the maximum response time to fill their Max Response time field.
- For IGMP group-specific queries, you can configure the IGMP last-member query interval to fill their Max Response time field. Namely, for IGMP group-specific queries, the maximum response time equals to the IGMP last-member query interval.

Configuring IGMP queries and responses globally

Follow these steps to configure IGMP queries and responses globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Configure the maximum response time to IGMP general queries	**max-response-time** *interval*	Optional 10 seconds by default
Configure the IGMP last-member query interval	**last-member-query-interval** *interval*	Optional 1 second by default

Configuring IGMP queries and responses in a VLAN

Follow these steps to configure IGMP queries and responses in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—

To do...	Use the command...	Remarks
Configure IGMP general query interval	**igmp-snooping query-interval** *interval*	Optional 60 seconds by default
Configure the maximum response time to IGMP general queries	**igmp-snooping max-response-time** *interval*	Optional 10 seconds by default
Configure the IGMP last-member query interval	**igmp-snooping last-member-query-interval** *interval*	Optional 1 second by default

Caution

In the configuration, make sure that the IGMP general query interval is larger than the maximum response time for IGMP general queries. Otherwise, multicast group members may be deleted by mistake.

Configuring Source IP Address of IGMP Queries

Upon receiving an IGMP query whose source IP address is 0.0.0.0 on a port, the switch does not enlist that port as a dynamic router port. This may prevent multicast forwarding entries from being correctly created at the data link layer and cause multicast traffic forwarding failure in the end. When a Layer 2 device acts as an IGMP-Snooping querier, to avoid the aforesaid problem, you are commended to configure a non-all-zero IP address as the source IP address of IGMP queries.

Follow these steps to configure source IP address of IGMP queries:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure the source address of IGMP general queries	**igmp-snooping general-query source-ip** { **current-interface** | *ip-address* }	Optional 0.0.0.0 by default
Configure the source IP address of IGMP group-specific queries	**igmp-snooping special-query source-ip** { **current-interface** | *ip-address* }	Optional 0.0.0.0 by default

Caution

The source address of IGMP query messages may affect IGMP querier selection within the segment.

Configuring an IGMP Snooping Policy

Configuration Prerequisites

Before configuring an IGMP Snooping policy, complete the following task:

- Enable IGMP Snooping in the VLAN or enable IGMP on the desired VLAN interface

Before configuring an IGMP Snooping policy, prepare the following data:

- ACL rule for multicast group filtering
- The maximum number of multicast groups that can pass the ports

Configuring a Multicast Group Filter

On an IGMP Snooping–enabled switch, the configuration of a multicast group allows the service provider to define restrictions on multicast programs available to different users.

In an actual application, when a user requests a multicast program, the user's host initiates an IGMP report. Upon receiving this report message, the switch checks the report against the configured ACL rule. If the port on which the report was received can join this multicast group, the switch adds an entry for this port in the IGMP Snooping forwarding table; otherwise the switch drops this report message. Any multicast data that has failed the ACL check will not be sent to this port. In this way, the service provider can control the VOD programs provided for multicast users.

Configuring a multicast group filter globally

Follow these steps to configure a multicast group filter globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Configure a multicast group filter	**group-policy** *acl-number* [**vlan** *vlan-list*]	Required No group filter is configured by default.

Configuring a multicast group filter on a port or a group of ports

Follow these steps to configure a multicast group filter on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure a multicast group filter	**igmp-snooping group-policy** *acl-number* [**vlan** *vlan-list*]	Required No group filter is configured by default.

Configuring Multicast Source Port Filtering

With the multicast source port filtering feature enabled on a port, the port can be connected with multicast receivers only rather than with multicast sources, because the port will block all multicast data packets while it permits multicast protocol packets to pass.

If this feature is disabled on a port, the port can be connected with both multicast sources and multicast receivers.

Configuring multicast source port filtering globally

Follow these steps to configure multicast source port filtering globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Enable multicast source port filtering	**source-deny port** *interface-list*	Required Disabled by default

Configuring multicast source port filtering on a port or a group of ports

Follow these steps to configure multicast source port filtering on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable multicast source port filtering	**igmp-snooping source-deny**	Required Disabled by default

Note

3Com Switch 4800G, when enabled to filter IPv4 multicast data based on the source ports, are automatically enabled to filter IPv6 multicast data based on the source ports.

Configuring the Function of Dropping Unknown Multicast Data

Unknown multicast data refers to multicast data for which no entries exist in the IGMP Snooping forwarding table. When receiving such multicast traffic, the switch floods it in the VLAN, incurring network bandwidth waste and low forwarding efficiency.

With the function of dropping unknown multicast data enabled, the switch forwards unknown multicast data to its router ports instead of flooding it in the VLAN. If no router ports exist, the switch drops the unknown multicast data..

Follow these steps to configure the function of dropping unknown multicast data in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—

To do...	Use the command...	Remarks
Enable the function of dropping unknown multicast data	**igmp-snooping drop-unknown**	Required Disabled by default

Configuring IGMP Report Suppression

When a Layer 2 device receives an IGMP report from a multicast group member, the device forwards the message to the Layer 3 device directly connected with it. Thus, when multiple members of a multicast group are attached to the Layer 2 device, the Layer 3 device directly connected with it will receive duplicate IGMP reports from these members.

With the IGMP report suppression function enabled, within each query cycle, the Layer 2 device forwards only the first IGMP report per multicast group to the Layer 3 device and will not forward the subsequent IGMP reports from the same multicast group to the Layer 3 device. This helps reduce the number of packets being transmitted over the network.

Follow these steps to configure IGMP report suppression:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Enable IGMP report suppression	**report-aggregation**	Optional Enabled by default

Configuring Maximum Multicast Groups that Can Be Joined on a Port

By configuring the maximum number of multicast groups that can be joined on a port, you can limit the number of multicast programs on-demand available to users, thus to regulate traffic on the port.

Follow these steps to configure the maximum number of multicast groups allowed on a port or ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure the maximum number of multicast groups allowed on the port(s)	**igmp-snooping group-limit** *limit* [**vlan** *vlan-list*]	Optional By default, the maximum number of multicast groups allowed on the port(s) is 1024.

Note

- When the number of multicast groups a port has joined reaches the maximum number configured, the system deletes all the forwarding entries persistent to that port from the IGMP Snooping forwarding table, and the hosts on this port need to join the multicast groups again.
- If you have configured static or simulated joins on a port, however, when the number of multicast groups on the port exceeds the configured threshold, the system deletes all the forwarding entries persistent to that port from the IGMP Snooping forwarding table and applies the static or simulated joins again, until the number of multicast groups joined by the port comes back within the configured threshold.

Configuring Multicast Group Replacement

For some special reasons, the number of multicast groups that can be joined on the current switch or port may exceed the number configured for the switch or the port. In addition, in some specific applications, a multicast group newly joined on the switch needs to replace an existing multicast group automatically. A typical example is “channel switching”, namely, by joining a new multicast group, a user automatically switches from the current multicast group to the new one.

To address such situations, you can enable the multicast group replacement function on the switch or certain ports. When the number of multicast groups joined on the switch or a port has joined reaches the limit:

- If the multicast group replacement feature is enabled, the newly joined multicast group automatically replaces an existing multicast group with the lowest address.
- If the multicast group replacement feature is not enabled, new IGMP reports will be automatically discarded.

Configuring multicast group replacement globally

Follow these steps to configure multicast group replacement globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IGMP Snooping view	**igmp-snooping**	—
Enable multicast group replacement	**overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Configuring multicast group replacement on a port or a group of ports

Follow these steps to configure multicast group replacement on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable multicast group replacement	**igmp-snooping overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Caution

Be sure to configure the maximum number of multicast groups allowed on a port (refer to Configuring Maximum Multicast Groups that Can Be Joined on a Port) before enabling multicast group replacement. Otherwise, the multicast group replacement functionality will not take effect.

Displaying and Maintaining IGMP Snooping

To do...	Use the command...	Remarks
View IGMP Snooping multicast group information	**display igmp-snooping group** [**vlan** *vlan-id*] [**slot** *slot-number*] [**verbose**]	Available in any view
View the statistics information of IGMP messages learned by IGMP Snooping	**display igmp-snooping statistics**	Available in any view
Clear IGMP Snooping multicast group information	**reset igmp-snooping group** { *group-address* | **all** } [**vlan** *vlan-id*]	Available in user view
Clear the statistics information of all kinds of IGMP messages learned by IGMP Snooping	**reset igmp-snooping statistics**	Available in user view

Note

- The **reset igmp-snooping group** command works only on an IGMP Snooping–enabled VLAN, but not on a VLAN with IGMP enabled on its VLAN interface.
- The **reset igmp-snooping group** command cannot clear the IGMP Snooping multicast group information for static joins.

IGMP Snooping Configuration Examples

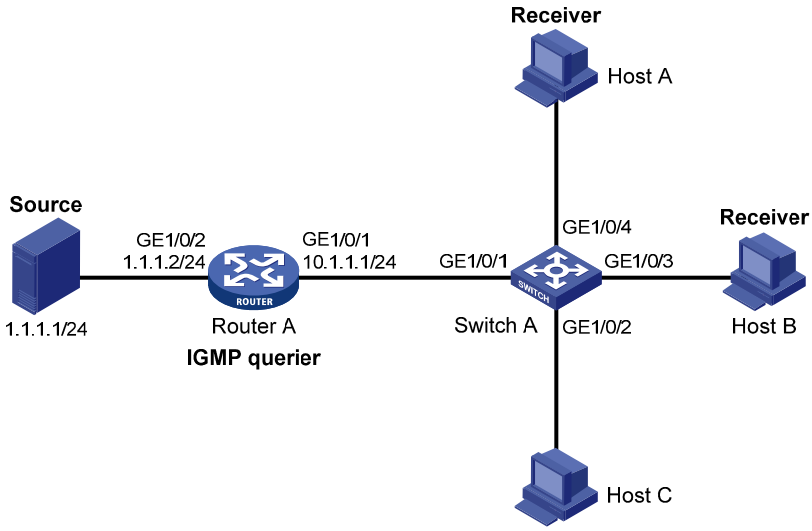
Configuring Group Policy and Simulated Joining

Network requirements

- As shown in Figure 1-3, Router A connects to the multicast source through GigabitEthernet 1/0/2 and to Switch A through GigabitEthernet 1/0/1.
- IGMPv2 is required on Router A, IGMP Snooping version 2 is required on Switch A, and Router A will act as the IGMP querier on the subnet.
- It is required that the receivers, Host A and Host B, attached to Switch A can receive multicast traffic addressed to multicast group 224.1.1.1 only.
- It is required that multicast data for group 224.1.1.1 can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A even if Host A and Host B accidentally, temporarily stop receiving multicast data.

Network diagram

Figure 1-3 Network diagram for group policy simulated joining configuration



Configuration procedure

1) Configure IP addresses

Configure an IP address and subnet mask for each interface as per Figure 1-3. The detailed configuration steps are omitted.

2) Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMP on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
```

```
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

3) Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to this VLAN, and enable IGMP Snooping and the function of dropping unknown multicast traffic in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping drop-unknown
[SwitchA-vlan100] quit
```

Configure a multicast group filter so that the hosts in VLAN 100 can join only the multicast group 224.1.1.1.

```
[SwitchA] acl number 2001
[SwitchA-acl-basic-2001] rule permit source 224.1.1.1 0
[SwitchA-acl-basic-2001] quit
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] group-policy 2001 vlan 100
[SwitchA-igmp-snooping] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 as simulated hosts for multicast group 224.1.1.1.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

4) Verify the configuration

View the detailed IGMP Snooping multicast groups information in VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 1 port.
            GE1/0/1               (D) ( 00:01:30 )
```

```
IP group(s):the following ip group(s) match to one mac group.
  IP group address:224.1.1.1
    (0.0.0.0, 224.1.1.1):
      Attribute:    Host Port
      Host port(s):total 2 port.
        GE1/0/3                 (D) ( 00:03:23 )
        GE1/0/4                 (D) ( 00:04:10 )
MAC group(s):
  MAC group address:0100-5e01-0101
      Host port(s):total 2 port.
        GE1/0/3
        GE1/0/4
```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A has joined multicast group 224.1.1.1.

Static Port Configuration

Network requirements

- As shown in Figure 1-4, Router A connects to a multicast source (Source) through GigabitEthernet 1/0/2, and to Switch A through GigabitEthernet 1/0/1.
- IGMPv2 is to run on Router A, and IGMPv2 Snooping is to run on Switch A, Switch B and Switch C, with Router A acting as the IGMP querier.
- Host A and host C are permanent receivers of multicast group 224.1.1.1. GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C are required to be configured as static member ports for multicast group 224.1.1.1 to enhance the reliability of multicast traffic transmission.
- Suppose STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked under normal conditions, and multicast traffic flows to the receivers attached to Switch C only along the path of Switch A—Switch B—Switch C.
- It is required to configure GigabitEthernet 1/0/3 that connects Switch A to Switch C as a static router port, so that multicast traffic can flow to the receivers nearly uninterruptedly along the path of Switch A—Switch C in the case that the path of Switch A—Switch B—Switch C gets blocked.

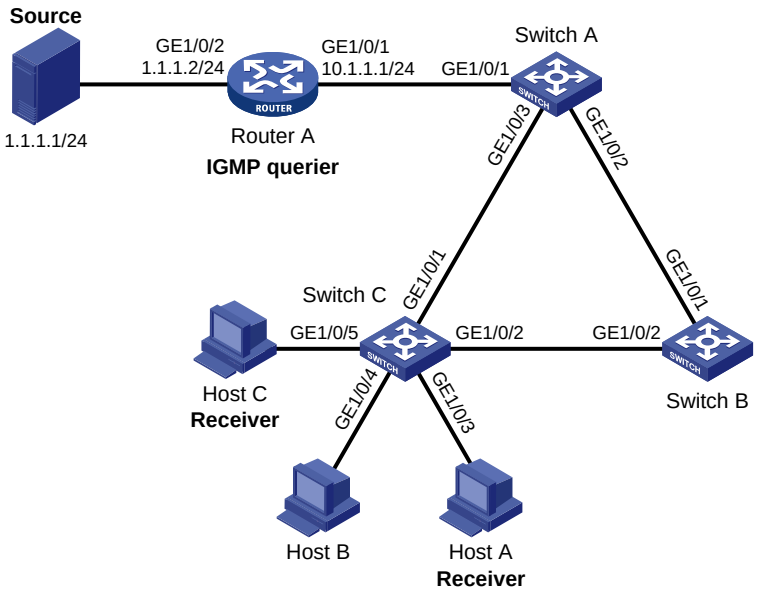
Note

If no static router port is configured, when the path of Switch A—Switch B—Switch C gets blocked, at least one IGMP query-response cycle must be completed before the multicast data can flow to the receivers along the new path of Switch A—Switch C, namely multicast delivery will be interrupted during this process.

For details about the Spanning Tree Protocol (STP), refer to *MSTP Configuration* in the *Access Volume*.

Network diagram

Figure 1-4 Network diagram for static port configuration



Configuration procedure

1) Configure IP addresses

Configure an IP address and subnet mask for each interface as per Figure 1-4. The detailed configuration steps are omitted.

2) Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMP on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

3) Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
[SwitchA-vlan100] igmp-snooping enable
```

```
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 to be a static router port.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping static-router-port vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
```

4) Configure Switch B

Enable IGMP Snooping globally.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
[SwitchB-vlan100] igmp-snooping enable
[SwitchB-vlan100] quit
```

5) Configure Switch C

Enable IGMP Snooping globally.

```
<SwitchC> system-view
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/5
[SwitchC-vlan100] igmp-snooping enable
[SwitchC-vlan100] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 as static member ports for multicast group 224.1.1.1.

```
[SwitchC] interface GigabitEthernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] igmp-snooping static-group 224.1.1.1 vlan 100
[SwitchC-GigabitEthernet1/0/3] quit
[SwitchC] interface GigabitEthernet 1/0/5
[SwitchC-GigabitEthernet1/0/5] igmp-snooping static-group 224.1.1.1 vlan 100
[SwitchC-GigabitEthernet1/0/5] quit
```

6) Verify the configuration

View the detailed IGMP Snooping multicast group information in VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
```

```
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 2 port.
            GE1/0/1               (D) ( 00:01:30 )
            GE1/0/3               (S)
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:224.1.1.1
        (0.0.0.0, 224.1.1.1):
          Attribute:    Host Port
          Host port(s):total 1 port.
            GE1/0/2               (D) ( 00:03:23 )
    MAC group(s):
      MAC group address:0100-5e01-0101
          Host port(s):total 1 port.
            GE1/0/2
```

As shown above, GigabitEthernet 1/0/3 of Switch A has become a static router port.

View the detailed IGMP Snooping multicast group information in VLAN 100 on Switch C.

```
[SwitchC] display igmp-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 1 port.
            GE1/0/2               (D) ( 00:01:23 )
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:224.1.1.1
        (0.0.0.0, 224.1.1.1):
          Attribute:    Host Port
          Host port(s):total 2 port.
            GE1/0/3               (S)
            GE1/0/5               (S)
    MAC group(s):
      MAC group address:0100-5e01-0101
          Host port(s):total 2 port.
            GE1/0/3
            GE1/0/5
```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C have become static member ports for multicast group 224.1.1.1.

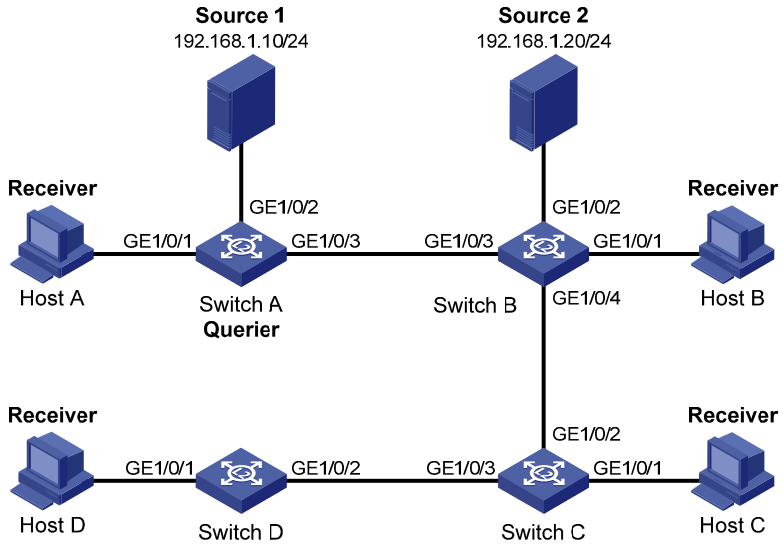
IGMP Snooping Querier Configuration

Network requirements

- As shown in Figure 1-5, in a Layer 2–only network environment, two multicast sources Source 1 and Source 2 send multicast data to multicast groups 224.1.1.1 and 225.1.1.1 respectively, Host A and Host C are receivers of multicast group 224.1.1.1, while Host B and Host D are receivers of multicast group 225.1.1.1.
- All the receivers are running IGMPv2, and all the switches need to run IGMP Snooping version 2. Switch A, which is close to the multicast sources, is chosen as the IGMP-Snooping querier.
- To prevent flooding of unknown multicast traffic within the VLAN, it is required to configure all the switches to drop unknown multicast data packets.
- Because a switch does not enlist a port that has heard an IGMP query with a source IP address of 0.0.0.0 (default) as a dynamic router port, configure a non-all-zero IP address as the source IP address of IGMP queries to ensure normal creation of Layer 2 multicast forwarding entries.

Network diagram

Figure 1-5 Network diagram for IGMP Snooping querier configuration



Configuration procedure

1) Configure switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100 and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable IGMP Snooping and the function of dropping unknown multicast traffic in VLAN 100.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping drop-unknown
```

Enable the IGMP-Snooping querier function in VLAN 100

```
[SwitchA-vlan100] igmp-snooping querier
```

Set the source IP address of IGMP general queries and group-specific queries to 192.168.1.1 in VLAN 100.

```
[SwitchA-vlan100] igmp-snooping general-query source-ip 192.168.1.1
[SwitchA-vlan100] igmp-snooping special-query source-ip 192.168.1.1
[SwitchA-vlan100] quit
```

2) Configure Switch B

Enable IGMP Snooping globally.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 100, and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable IGMP Snooping and the function of dropping unknown multicast traffic in VLAN 100.

```
[SwitchB-vlan100] igmp-snooping enable
[SwitchB-vlan100] igmp-snooping drop-unknown
[SwitchB-vlan100] quit
```

Configurations on Switch C and Switch D are similar to the configuration on Switch B.

3) Verify the configuration

After the IGMP Snooping querier starts to work, all the switches but the querier can receive IGMP general queries. By using the **display igmp-snooping statistics** command, you can view the statistics information about the IGMP messages received. For example:

View the IGMP message statistics on Switch B.

```
[SwitchB] display igmp-snooping statistics
  Received IGMP general queries:3.
  Received IGMPv1 reports:0.
  Received IGMPv2 reports:12.
  Received IGMP leaves:0.
  Received IGMPv2 specific queries:0.
  Sent     IGMPv2 specific queries:0.
  Received IGMPv3 reports:0.
  Received IGMPv3 reports with right and wrong records:0.
  Received IGMPv3 specific queries:0.
  Received IGMPv3 specific sg queries:0.
  Sent     IGMPv3 specific queries:0.
  Sent     IGMPv3 specific sg queries:0.
  Received error IGMP messages:0.
```

Troubleshooting IGMP Snooping Configuration

Switch Fails in Layer 2 Multicast Forwarding

Symptom

A switch fails to implement Layer 2 multicast forwarding.

Analysis

IGMP Snooping is not enabled.

Solution

1) Enter the **display current-configuration** command to view the running status of IGMP Snooping.
2) If IGMP Snooping is not enabled, use the **igmp-snooping** command to enable IGMP Snooping globally, and then use **igmp-snooping enable** command to enable IGMP Snooping in VLAN view.
3) If IGMP Snooping is disabled only for the corresponding VLAN, just use the **igmp-snooping enable** command in VLAN view to enable IGMP Snooping in the corresponding VLAN.

Configured Multicast Group Policy Fails to Take Effect

Symptom

Although a multicast group policy has been configured to allow hosts to join specific multicast groups, the hosts can still receive multicast data addressed to other multicast groups.

Analysis

- The ACL rule is incorrectly configured.
- The multicast group policy is not correctly applied.
- The function of dropping unknown multicast data is not enabled, so unknown multicast data is flooded.

Solution

1) Use the **display acl** command to check the configured ACL rule. Make sure that the ACL rule conforms to the multicast group policy to be implemented.
2) Use the **display this** command in IGMP Snooping view or in the corresponding port view to check whether the correct multicast group policy has been applied. If not, use the **group-policy** or **igmp-snooping group-policy** command to apply the correct multicast group policy.
3) Use the **display current-configuration** command to check whether the function of dropping unknown multicast data is enabled. If not, use the **igmp-snooping drop-unknown** command to enable the function of dropping unknown multicast data.

Table of Contents

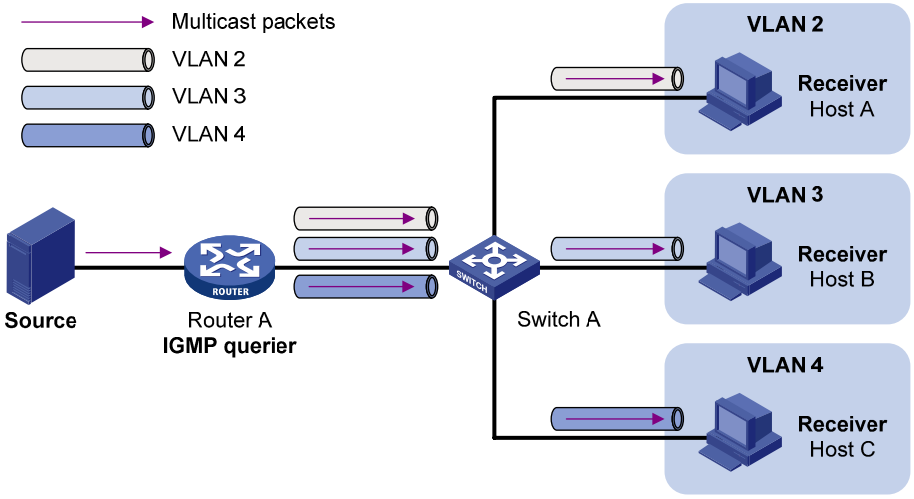
1 Multicast VLAN Configuration

When configuring multicast VLAN, go to these sections for information you are interested in:

Introduction to Multicast VLAN

As shown in Figure 1-1, in the traditional multicast programs-on-demand mode, when hosts, Host A, Host B and Host C, belonging to different VLANs require multicast programs on demand service, the Layer 3 device, Router A, needs to forward a separate copy of the multicast traffic in each user VLAN to the Layer 2 device, Switch A. This results in not only waste of network bandwidth but also extra burden on the Layer 3 device.

Figure 1-1 Multicast transmission without multicast VLAN



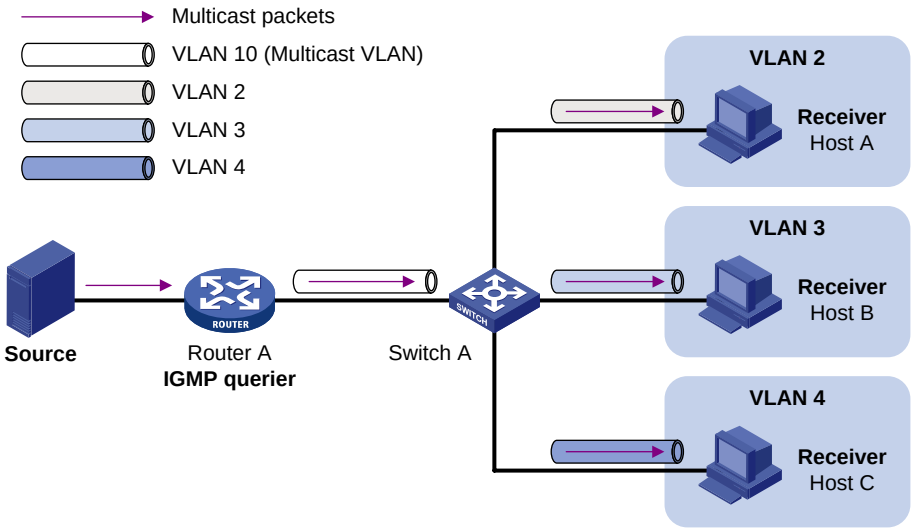
The multicast VLAN feature configured on the Layer 2 device is the solution to this issue. With the multicast VLAN feature, the Layer 3 device needs to replicate the multicast traffic only in the multicast VLAN instead of making a separate copy of the multicast traffic in each user VLAN. This saves the network bandwidth and lessens the burden of the Layer 3 device.

The multicast VLAN feature can be implemented in two approaches, as described below:

Sub-VLAN-based multicast VLAN

As shown in Figure 1-2, Host A, Host B and Host C are in three different user VLANs. On Switch A, configure VLAN 10 as a multicast VLAN, configure all the user VLANs as sub-VLANs of this multicast VLAN, and enable IGMP Snooping in the multicast VLAN.

Figure 1-2 Sub-VLAN-based multicast VLAN

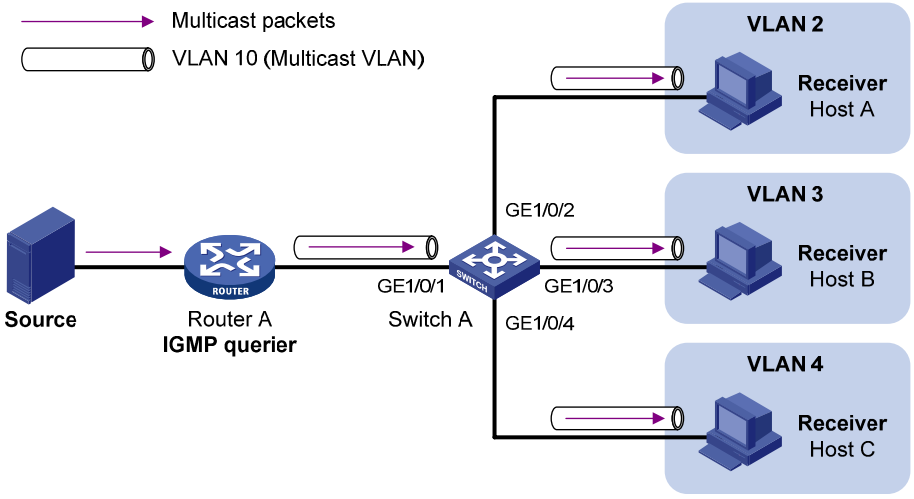


After the configuration, IGMP Snooping manages router ports in the multicast VLAN and member ports in the sub-VLANs. When forwarding multicast data to Switch A, Router A needs to send only one copy of multicast traffic to Switch A in the multicast VLAN, and Switch A distributes the traffic to the multicast VLAN's sub-VLANs that contain receivers.

Port-based multicast VLAN

As shown in Figure 1-3, Host A, Host B and Host C are in three different user VLANs. All the user ports (ports with attached hosts) on Switch A are hybrid ports. On Switch A, configure VLAN 10 as a multicast VLAN, assign all the user ports to this multicast VLAN, and enable IGMP Snooping in the multicast VLAN and all the user VLANs.

Figure 1-3 Port-based multicast VLAN



After the configuration, upon receiving an IGMP message on a user port, Switch A tags the message with the multicast VLAN ID and relays it to the IGMP querier, so that IGMP Snooping can uniformly manage the router ports and member ports in the multicast VLAN. When forwarding multicast data to Switch A, Router A needs to send only one copy of multicast traffic to Switch A in the multicast VLAN, and Switch A distributes the traffic to all the member ports in the multicast VLAN.

Note

- For information about IGMP Snooping, router ports, and member ports, refer to *IGMP Snooping Configuration* in the *IP Multicast Volume*.
- For information about VLAN tags, refer to *VLAN Configuration* in the *Access Volume*.

Multicast VLAN Configuration Task List

Complete the following tasks to configure multicast VLAN:

Task		Remarks
Configuring Sub-VLAN-Based Multicast VLAN		Required Use either approach.
Configuring Port-Based Multicast VLAN	Configuring User Port Attributes	
	Configuring Multicast VLAN Ports	

Note

If you have configured both sub-VLAN-based multicast VLAN and port-based multicast VLAN on a device, the port-based multicast VLAN configuration is given preference.

Configuring Sub-VLAN-Based Multicast VLAN

Configuration Prerequisites

Before configuring sub-VLAN-based multicast VLAN, complete the following tasks:

- Create VLANs as required
- Enable IGMP Snooping in the VLAN to be configured as a multicast VLAN

Configuring Sub-VLAN-Based Multicast VLAN

In this approach, you need to configure a VLAN as a multicast VLAN, and then configure user VLANs as sub-VLANs of the multicast VLAN.

Follow these steps to configure sub-VLAN-based multicast VLAN:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as a multicast VLAN and enter multicast VLAN view	**multicast-vlan** *vlan-id*	Required Not a multicast VLAN by default
Configure the specified VLAN(s) as sub-VLAN(s) of the multicast VLAN	**subvlan** *vlan-list*	Required By default, a multicast VLAN has no sub-VLANs.

Note

- You cannot configure multicast VLAN on a device with IP multicast routing enabled.
- The VLAN to be configured as a multicast VLAN must exist.
- The VLANs to be configured as sub-VLANs of the multicast VLAN must exist and must not be sub-VLANs of another multicast VLAN.
- The total number of sub-VLANs of a multicast VLAN must not exceed 127.

Configuring Port-Based Multicast VLAN

When configuring port-based multicast VLAN, you need to configure the attributes of each user port and then assign the ports to the multicast VLAN.

Note

- A user port can be configured as a multicast VLAN port only if it is of the Ethernet or Layer 2 aggregate port type.
- Configurations made in Ethernet port view are effective only for the current port; configurations made in Layer 2 aggregate port view are effective only for the current port; configurations made in port group view are effective for all the ports in the current port group.

Configuration Prerequisites

Before configuring port-based multicast VLAN, complete the following tasks:

- Create VLANs as required
- Enable IGMP Snooping in the VLAN to be configured as a multicast VLAN
- Enable IGMP Snooping in all the user VLANs

Configuring User Port Attributes

Configure the user ports as hybrid ports that permit packets of the specified user VLAN to pass, and configure the user VLAN to which the user ports belong as the default VLAN.

Configure the user ports to permit packets of the multicast VLAN to pass and untag the packets. Thus, upon receiving multicast packets tagged with the multicast VLAN ID from the upstream device, the Layer 2 device untags the multicast packets and forwards them to its downstream device.

Follow these steps to configure user port attributes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter port view or port group view	**interface** *interface-type interface-number*	Required Use either command
	port-group { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the user port link type as hybrid	**port link-type hybrid**	Required Access by default
Specify the user VLAN that comprises the current user port(s) as the default VLAN	**port hybrid pvid vlan** *vlan-id*	Required VLAN 1 by default
Configure the current user port(s) to permit packets of the specified multicast VLAN(s) to pass and untag the packets	**port hybrid vlan** *vlan-id-list* **untagged**	Required By default, a hybrid port permits only packets of VLAN 1 to pass.

Note

For details about the **port link-type**, **port hybrid pvid vlan**, and **port hybrid vlan** commands, refer to *VLAN Commands* in the *Access Volume*.

Configuring Multicast VLAN Ports

In this approach, you need to configure a VLAN as a multicast VLAN and then assign user ports to this multicast VLAN by either adding the user ports in the multicast VLAN or specifying the multicast VLAN on the user ports. These two configuration methods give the same result.

Configuring multicast VLAN ports in multicast VLAN view

Follow these steps to configure multicast VLAN ports in multicast VLAN view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as a multicast VLAN and enter multicast VLAN view	**multicast-vlan** *vlan-id*	Required Not a multicast VLAN by default
Assign ports to the multicast VLAN	**port** *interface-list*	Required By default, a multicast VLAN has no ports.

Configuring multicast VLAN ports in port view or port group view

Follow these steps to configure multicast VLAN ports in port view or port group view:

To do…	Use this command…	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as a multicast VLAN and enter multicast VLAN view	**multicast-vlan** *vlan-id*	Required Not a multicast VLAN by default.
Return to system view	**quit**	—
Enter port view or port group view	**interface** *interface-type interface-number*	Required Use either command.
	port-group manual *port-group-name*	
Configure the current port(s) as port(s) of the multicast VLAN	**port multicast-vlan** *vlan-id*	Required By default, a user port does not belong to any multicast VLAN.

Note

- You cannot configure multicast VLAN on a device with multicast routing enabled.
- The VLAN to be configured as a multicast VLAN must exist.
- A port can belong to only one multicast VLAN.

Displaying and Maintaining Multicast VLAN

To do…	Use the command…	Remarks
Display information about a multicast VLAN	**display multicast-vlan** [*vlan-id*]	Available in any view

Multicast VLAN Configuration Examples

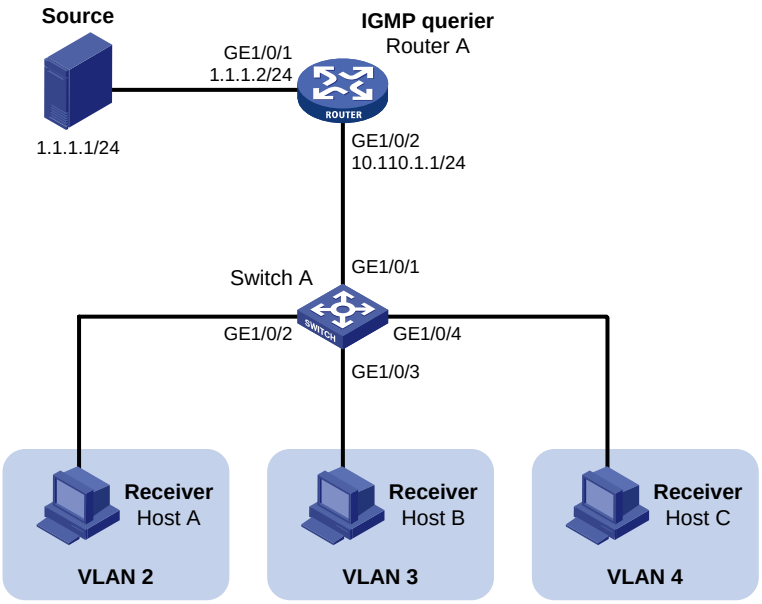
Sub-VLAN-Based Multicast VLAN Configuration

Network requirements

- Router A connects to a multicast source through GigabitEthernet1/0/1 and to Switch A, through GigabitEthernet 1/0/2.
- IGMPv2 is required on Router A, and IGMPv2 Snooping is required on Switch A. Router A is the IGMP querier.
- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 10, GigabitEthernet 1/0/2 through GigabitEthernet1/0/4 belong to VLAN 2 through VLAN 4 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 of Switch A respectively.
- The multicast source sends multicast data to multicast group 224.1.1.1. Host A, Host B, and Host C are receivers of the multicast group.
- Configure the sub-VLAN-based multicast VLAN feature so that Router A just sends multicast data to Switch A through the multicast VLAN and Switch A forwards the traffic to the receivers that belong to different user VLANs.

Network diagram

Figure 1-4 Network diagram for sub-VLAN-based multicast VLAN configuration



Configuration procedure

1) Configure IP addresses

Configure an IP address and subnet mask for each interface as per Figure 1-4. The detailed configuration steps are omitted here.

2) Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface and enable IGMP on the host-side interface GigabitEthernet 1/0/2.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] igmp enable
```

3) Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 2 and assign GigabitEthernet 1/0/2 to this VLAN.

```
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/2
[SwitchA-vlan2] quit
```

The configuration for VLAN 3 and VLAN 4 is similar to the configuration for VLAN 2.

Create VLAN 10, assign GigabitEthernet 1/0/1 to this VLAN and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 10
[SwitchA-vlan10] port gigabitethernet 1/0/1
[SwitchA-vlan10] igmp-snooping enable
[SwitchA-vlan10] quit
```

Configure VLAN 10 as a multicast VLAN and configure VLAN 2 through VLAN 4 as its sub-VLANs.

```
[SwitchA] multicast-vlan 10
[SwitchA-mvlan-10] subvlan 2 to 4
[SwitchA-mvlan-10] quit
```

4) Verify the configuration

Display information about the multicast VLAN.

```
[SwitchA] display multicast-vlan
 Total 1 multicast-vlan(s)

 Multicast vlan 10
   subvlan list:
    vlan 2-4
   port list:
    no port
```

View the IGMP Snooping multicast group information on Switch A.

```
[SwitchA] display igmp-snooping group
  Total 4 IP Group(s).
  Total 4 IP Source(s).
  Total 4 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):2.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 0 port.
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:224.1.1.1
        (0.0.0.0, 224.1.1.1):
          Host port(s):total 1 port.
            GE1/0/2               (D)
    MAC group(s):
      MAC group address:0100-5e01-0101
          Host port(s):total 1 port.
            GE1/0/2

  Vlan(id):3.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
```

```
  Router port(s):total 0 port.
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:224.1.1.1
      (0.0.0.0, 224.1.1.1):
        Host port(s):total 1 port.
          GE1/0/3                (D)
  MAC group(s):
    MAC group address:0100-5e01-0101
        Host port(s):total 1 port.
          GE1/0/3

Vlan(id):4.
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).
  Router port(s):total 0 port.
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:224.1.1.1
      (0.0.0.0, 224.1.1.1):
        Host port(s):total 1 port.
          GE1/0/4                (D)
  MAC group(s):
    MAC group address:0100-5e01-0101
        Host port(s):total 1 port.
          GE1/0/4

Vlan(id):10.
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).
  Router port(s):total 1 port.
          GE1/0/1                (D)
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:224.1.1.1
      (0.0.0.0, 224.1.1.1):
        Host port(s):total 0 port.
  MAC group(s):
    MAC group address:0100-5e01-0101
        Host port(s):total 0 port.
```

As shown above, IGMP Snooping is maintaining the router port in the multicast VLAN (VLAN 10) and the member ports in the sub-VLANs (VLAN 2 through VLAN 4).

Port-Based Multicast VLAN Configuration

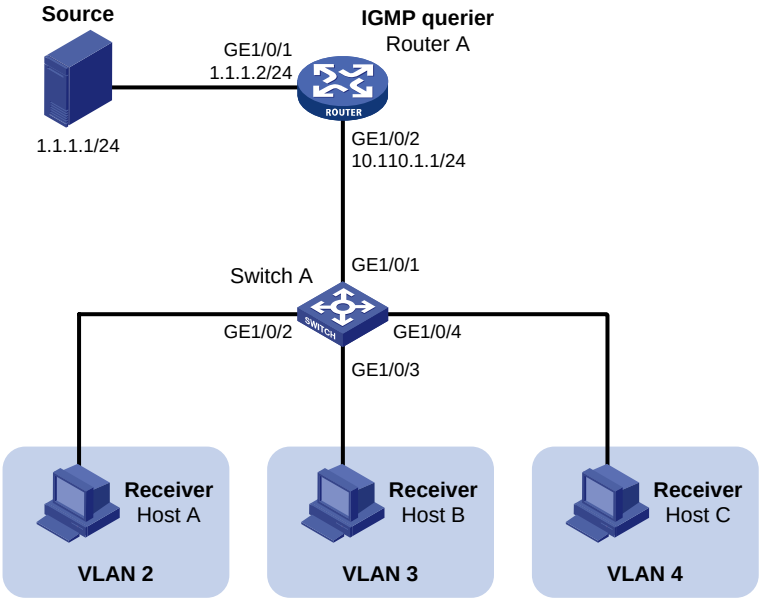
Network requirements

- As shown in Figure 1-5, Router A connects to a multicast source (Source) through GigabitEthernet 1/0/1, and to Switch A through GigabitEthernet 1/0/2.

- IGMPv2 is required on Router A. IGMPv2 Snooping is required on Switch A. Router A acts as the IGMP querier.
- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 10, GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 belong to VLAN 2 through VLAN 4 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet1/0/4 of Switch A respectively.
- The multicast source sends multicast data to multicast group 224.1.1.1. Host A, Host B, and Host C are receivers of the multicast group.
- Configure the port-based multicast VLAN feature so that Router A just sends multicast data to Switch A through the multicast VLAN and Switch A forwards the multicast data to the receivers that belong to different user VLANs.

Network diagram

Figure 1-5 Network diagram for port-based multicast VLAN configuration



Configuration procedure

1) Configure IP addresses

Configure the IP address and subnet mask for each interface as per Figure 1-5. The detailed configuration steps are omitted here.

2) Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMP on the host-side interface GigabitEthernet 1/0/2.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] igmp enable
```

3) Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 10, assign GigabitEthernet 1/0/1 to VLAN 10, and enable IGMP Snooping in this VLAN.

```
[SwitchA] vlan 10
[SwitchA-vlan10] port gigabitethernet 1/0/1
[SwitchA-vlan10] igmp-snooping enable
[SwitchA-vlan10] quit
```

Create VLAN 2 and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 2
[SwitchA-vlan2] igmp-snooping enable
[SwitchA-vlan2] quit
```

The configuration for VLAN 3 and VLAN 4 is similar. The detailed configuration steps are omitted.

Configure GigabitEthernet 1/0/2 as a hybrid port. Configure VLAN 2 as the default VLAN. Configure GigabitEthernet 1/0/2 to permit packets of VLAN 2 and VLAN 10 to pass and untag the packets when forwarding them.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-type hybrid
[SwitchA-GigabitEthernet1/0/2] port hybrid pvid vlan 2
[SwitchA-GigabitEthernet1/0/2] port hybrid vlan 2 untagged
[SwitchA-GigabitEthernet1/0/2] port hybrid vlan 10 untagged
[SwitchA-GigabitEthernet1/0/2] quit
```

The configuration for GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 is similar. The detailed configuration steps are omitted.

Configure VLAN 10 as a multicast VLAN.

```
[SwitchA] multicast-vlan 10
```

Assign GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to VLAN 10.

```
[SwitchA-mvlan-10] port gigabitethernet 1/0/2 to gigabitethernet 1/0/3
[SwitchA-mvlan-10] quit
```

Assign GigabitEthernet 1/0/4 to VLAN 10.

```
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] port multicast-vlan 10
[SwitchA-GigabitEthernet1/0/4] quit
```

4) Verify the configuration

View the multicast VLAN information on Switch A.

```
[SwitchA] display multicast-vlan
 Total 1 multicast-vlan(s)

 Multicast vlan 10
   subvlan list:
    no subvlan
   port list:
    GE1/0/2                 GE1/0/3                 GE1/0/4
```

View the IGMP Snooping multicast group information on Switch A.

```
[SwitchA] display igmp-snooping group
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):10.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 1 port.
            GE1/0/1                (D)
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:224.1.1.1
        (0.0.0.0, 224.1.1.1):
          Host port(s):total 3 port.
            GE1/0/2                (D)
            GE1/0/3                (D)
            GE1/0/4                (D)
    MAC group(s):
      MAC group address:0100-5e01-0101
          Host port(s):total 3 port.
            GE1/0/2
            GE1/0/3
            GE1/0/4
```

As shown above, IGMP Snooping is maintaining the router ports and member ports in VLAN 10.

Table of Contents

1 IPv6 Multicast Routing and Forwarding Configuration

When configuring IPv6 multicast routing and forwarding, go to the following sections for information you are interested in:

- IPv6 Multicast Routing and Forwarding Overview
- Configuration Task List
- Displaying and Maintaining IPv6 Multicast Routing and Forwarding
- Troubleshooting IPv6 Multicast Policy Configuration

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running an IPv6 multicast routing protocol.

IPv6 Multicast Routing and Forwarding Overview

Introduction to IPv6 Multicast Routing and Forwarding

In IPv6 multicast implementations, multicast routing and forwarding are implemented by three types of tables:

- Each IPv6 multicast routing protocol has its own multicast routing table, such as IPv6 PIM routing table.
- The multicast routing information of different IPv6 multicast routing protocols forms a general IPv6 multicast routing table.
- The IPv6 multicast forwarding table is directly used to control the forwarding of IPv6 multicast packets. This is the table that guides IPv6 multicast forwarding.

An IPv6 multicast forwarding table consists of a set of (S, G) entries, each indicating the routing information for delivering multicast data from a multicast source to a multicast group. If a router supports multiple IPv6 multicast protocols, its IPv6 multicast routing table will include routes generated by these protocols. The router chooses the optimal route from the IPv6 multicast routing table based on the configured multicast routing and forwarding policy and installs the route entry into its IPv6 multicast forwarding table.

RPF Check Mechanism

An IPv6 multicast routing protocol relies on the existing IPv6 unicast routing information or IPv6 MBGP routes in creating IPv6 multicast routing entries. When creating IPv6 multicast routing table entries, an IPv6 multicast routing protocol uses the reverse path forwarding (RPF) check mechanism to ensure

IPv6 multicast data delivery along the correct path. In addition, the RPF check mechanism also helps avoid data loops caused by various reasons.

RPF Check process

The basis for an RPF check is an IPv6 unicast route or an IPv6 MBGP route.

- An IPv6 unicast routing table contains the shortest path to each destination subnet;
- An IPv6 MBGP routing table contains IPv6 multicast routing information.

When performing an RPF check, a router searches its IPv6 unicast routing table and IPv6 MBGP routing table at the same time. The specific process is as follows:

1) The router first chooses an optimal route from the IPv6 unicast routing table and IPv6 MBGP routing table respectively:

- The router searches its IPv6 unicast routing table using the IPv6 address of the “packet source” as the destination address and automatically selects the optimal route as the RPF route. The outgoing interface in the corresponding routing entry is the RPF interface and the next hop is the RPF neighbor. The router considers the path along which the IPv6 multicast packet from the RPF neighbor arrived on the RPF interface to be the shortest path that leads back to the source.
- The router automatically chooses an optimal IPv6 MBGP route by searching its MBGP routing table, using the IPv6 address of the “packet source” as the destination address. The outgoing interface in the corresponding routing entry is the RPF interface and the next hop is the RPF neighbor.

2) Then, the router selects one from these two optimal routes as the RPF route. The selection process is as follows:

- If configured to use the longest match principle, the router selects the longest match route from the two; if these two routes have the same prefix length, the router selects the route with a higher priority; if these two routes have the same priority, the router selects the IPv6 MBGP route as the RPF route.
- If not configured to use the longest match principle, the router selects the route with a higher priority; if these two routes have the same priority, the router selects the IPv6 MBGP route as the RPF route.

Note

The above-mentioned “packet source” can mean different things in different situations:

- For a packet traveling along the shortest path tree (SPT) from the multicast source to the receivers or the rendezvous point (RP), the “packet source” for RPF check is the multicast source.
- For a packet traveling along the rendezvous point tree (RPT) from the RP to the receivers, the “packet source” for RPF check is the RP.
- For a bootstrap message from the bootstrap router (BSR), the “packet source” for RPF check is the BSR.

For details about the concepts of SPT, RPT, RP and BSR, refer to *IPv6 PIM Configuration* in the *IP Multicast Volume*.

Implementation of the RPF check in IPv6 multicast

Implementing an RPF check on each received IPv6 multicast data packet would bring a big burden to the router. The use of an IPv6 multicast forwarding table is the solution to this issue. When creating an IPv6 multicast routing entry and an IPv6 multicast forwarding entry for an IPv6 multicast packet, the router sets the RPF interface of the packet as the incoming interface of the (S, G) entry. Upon receiving an (S, G) IPv6 multicast packet, the router first searches its IPv6 multicast forwarding table:

1) If the corresponding (S, G) entry does not exist in the IPv6 multicast forwarding table, the packet is subject to an RPF check. The router creates an IPv6 multicast routing entry based on the relevant routing information and installs the entry into the IPv6 multicast forwarding table, with the RPF interface as the incoming interface.

- If the interface on which the packet actually arrived is the RPF interface, the RPF check succeeds and the router forwards the packet to all the outgoing interfaces.
- If the interface on which the packet actually arrived is not the RPF interface, the RPF check fails and the router discards the packet.

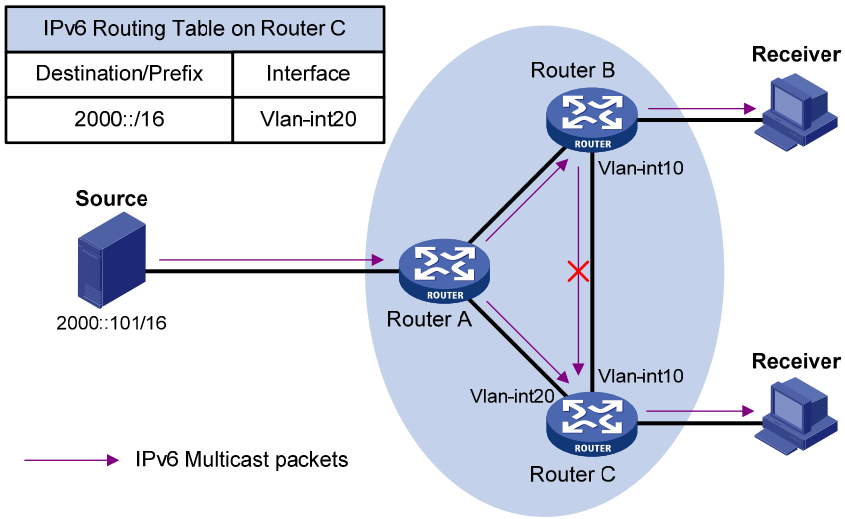
2) If the corresponding (S, G) entry exists, and the interface on which the packet actually arrived is the incoming interface, the router forwards the packet to all the outgoing interfaces.

3) If the corresponding (S, G) entry exists, but the interface on which the packet actually arrived is not the incoming interface in the IPv6 multicast forwarding table, the IPv6 multicast packet is subject to an RPF check.

- If the RPF interface is the incoming interface of the (S, G) entry, this means the (S, G) entry is correct but the packet arrived from a wrong path. The packet is to be discarded.
- If the RPF interface is not the incoming interface, this means the (S, G) entry has expired, and router replaces the incoming interface with the RPF interface. If the interface on which the packet arrived in the RPF interface, the router forwards the packet to all the outgoing interfaces; otherwise it discards the packet.

Assume that IPv6 unicast routes are available in the network, IPv6 MBGP is not configured, and IPv6 multicast packets travel along the SPT from the multicast source to the receivers, as shown in Figure 1-1. The IPv6 multicast forwarding table on Router C contains the (S, G) entry, with Vlan-interface 20 as the RPF interface.

Figure 1-1 RPF check process



- When an IPv6 multicast packet arrives on Vlan-interface 20 of Router C, as the interface is the incoming interface of the (S, G) entry, the router forwards the packet to all outgoing interfaces.

- When an IPv6 multicast packet arrives on Vlan-interface 10 of Router C, as the interface is not the incoming interface of the (S, G) entry, the router performs an RPF check on the packet: The router searches its IPv6 unicast routing table and finds that the outgoing interface to Source (the RPF interface) is Vlan-interface 20. This means the (S, G) entry is correct and packet arrived along a wrong path. The RPF check fails and the packet is discarded.

Configuration Task List

Complete these tasks to configure IPv6 multicast routing and forwarding:

Task		Remarks
Enabling IPv6 Multicast Routing		Required
Configuring IPv6 Multicast Routing and Forwarding	Configuring an IPv6 Multicast Routing Policy	Optional
	Configuring an IPv6 Multicast Forwarding Range	Optional
	Configuring the IPv6 Multicast Forwarding Table Size	Optional

Enabling IPv6 Multicast Routing

Before configuring any Layer 3 IPv6 multicast functionality, you must enable IPv6 multicast routing.

Follow these steps to enable IPv6 multicast routing:

To do…	Use the Command…	Remarks
Enter system view	**system-view**	—
Enable IPv6 multicast routing	**multicast ipv6 routing-enable**	Required Disabled by default

Configuring IPv6 Multicast Routing and Forwarding

Configuration Prerequisites

Before configuring IPv6 multicast routing and forwarding, complete the following tasks:

- Configure an IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure IPv6 PIM-DM or IPv6 PIM-SM.

Before configuring IPv6 multicast routing and forwarding, prepare the following data:

- Minimum hop limit value required for an IPv6 multicast packet to be forwarded
- Maximum number of downstream nodes for a single entry in the IPv6 multicast forwarding table
- Maximum number of entries in the IPv6 multicast forwarding table

Configuring an IPv6 Multicast Routing Policy

You can configure the router to determine the RPF route based on the longest match principle. For details about RPF route selection, refer to RPF Check process.

By configuring per-source or per-source-and-group load splitting, you can optimize the traffic delivery when multiple IPv6 multicast data streams are handled.

Follow these steps to configure an IPv6 multicast routing policy:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the device to select the RPF route based on the longest match	**multicast ipv6 longest-match**	Optional The route with the highest priority is selected as the RPF route by default
Configure IPv6 multicast load splitting	**multicast ipv6 load-splitting** {**source** | **source-group** }	Optional Disabled by default

Configuring an IPv6 Multicast Forwarding Range

IPv6 multicast packets do not travel infinitely in a network. The IPv6 multicast data of each IPv6 multicast group must be transmitted within a definite scope. Presently, you can define an IPv6 multicast forwarding range by:

- Specifying boundary interfaces, which form a closed IPv6 multicast forwarding area, or
- Setting the minimum hop limit value required for an IPv6 multicast packet to be forwarded.

Note

Setting the minimum hop limit value is not supported on 3Com Switch 4800G.

You can configure the forwarding boundary for a specific IPv6 multicast group on all interfaces that support IPv6 multicast forwarding. A multicast forwarding boundary sets the boundary condition for the IPv6 multicast groups in the specified range. If the destination address of an IPv6 multicast packet matches the set boundary condition, the packet will not be forwarded. Once an IPv6 multicast boundary is configured on an interface, this interface can no longer forward IPv6 multicast packets (including those sent from the local device) or receive IPv6 multicast packets.

Follow these steps to configure an IPv6 multicast forwarding range:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an IPv6 multicast forwarding boundary	**multicast ipv6 boundary** *ipv6-group-address prefix-length*	Required No forwarding boundary by default

Configuring the IPv6 Multicast Forwarding Table Size

The router maintains the corresponding forwarding entry for each IPv6 multicast packet it receives. Excessive IPv6 multicast routing entries, however, can exhaust the router's memory and thus result in lower router performance. You can set a limit on the number of entries in the IPv6 multicast forwarding

table based on the actual networking situation and the performance requirements. If the configured maximum number of IPv6 multicast forwarding table entries is smaller than the current value, the entries in excess will not be immediately deleted; instead they will be deleted by the IPv6 multicast routing protocol running on the router. The router will no longer install new IPv6 multicast forwarding entries until the number of existing IPv6 multicast forwarding entries comes down below the configured value.

When forwarding IPv6 multicast traffic, the router replicates a copy of the IPv6 multicast traffic for each downstream node and forwards the traffic, and thus each of these downstream nodes forms a branch of the IPv6 multicast distribution tree. You can configure the maximum number of downstream nodes (namely, the maximum number of outgoing interfaces) for a single entry in the IPv6 multicast forwarding table to lessen burden on the router for replicating IPv6 multicast traffic. If the configured maximum number of downstream nodes for a single IPv6 multicast forwarding entry is smaller than the current number, the downstream nodes in excess will not be deleted immediately; instead they must be deleted by the IPv6 multicast routing protocol. The router will no longer install new IPv6 multicast forwarding entries for newly added downstream nodes until the number of existing downstream nodes comes down below the configured value.

Follow these steps to configure the IPv6 multicast forwarding table size:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the maximum number of entries in the IPv6 multicast forwarding table	**multicast ipv6 forwarding-table route-limit** *limit*	Optional 512 by default.
Configure the maximum number of downstream nodes for a single IPv6 multicast forwarding entry	**multicast ipv6 forwarding-table downstream-limit** *limit*	Optional 128 by default.

Displaying and Maintaining IPv6 Multicast Routing and Forwarding

To do...	Use the command...	Remarks
Display the IPv6 multicast boundary information	**display multicast ipv6 boundary** [*ipv6-group-address* [*prefix-length*] | **interface** *interface-type interface-number*]	Available in any view
Display the information of the IPv6 multicast forwarding table	**display multicast ipv6 forwarding-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } } | **statistics** | **slot** *slot-number*] * [**port-info**]	Available in any view
Display the information of the IPv6 multicast routing table	**display multicast ipv6 routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **exclude** | **include** | **match** } { *interface-type interface-number* | **register** } }] *	Available in any view

To do...	Use the command...	Remarks
Display the RPF route information of the specified IPv6 multicast source	**display multicast ipv6 rpf-info** *ipv6-source-address* [*ipv6-group-address*]	Available in any view
Clear forwarding entries from the IPv6 multicast forwarding table	**reset multicast ipv6 forwarding-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }	Available in user view
Clear routing entries from the IPv6 multicast routing table	**reset multicast ipv6 routing-table** { { *ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*] | **incoming-interface** { *interface-type interface-number* | **register** } } * | **all** }	Available in user view

Caution

- The **reset** command clears the information in the IPv6 multicast routing table or the multicast forwarding table, and thus may cause transmission failure of IPv6 multicast information.
- When a routing entry is deleted from the IPv6 multicast routing table, the corresponding forwarding entry will also be deleted from the IPv6 multicast forwarding table.
- When a forwarding entry is deleted from the IPv6 multicast forwarding table, the corresponding routing entry will also be deleted from the IPv6 multicast routing table.

Troubleshooting IPv6 Multicast Policy Configuration

Abnormal Termination of IPv6 Multicast Data

Symptom

- A host sends an MLD report announcing its joining an IPv6 multicast group (G). However, there is no member information about the IPv6 multicast group (G) on the immediate router. The intermediate router can receive IPv6 multicast packets successfully, but the packets cannot reach the stub network.
- The interface of the intermediate router receives the IPv6 multicast packets, but there is no corresponding (S, G) entry in the IPv6 PIM routing table.

Analysis

- The **multicast ipv6 boundary** command is used to filter IPv6 multicast packets received on an interface. If an IPv6 multicast packet fails to match the IPv6 ACL rule of this command, IPv6 PIM will create no routing entry.
- In addition, the **source-policy** command in IPv6 PIM is used to filter received IPv6 multicast packets. If an IPv6 multicast packet fails to match the IPv6 ACL rule of this command, IPv6 PIM will not create a routing entry, either.

Solution

- Use the **display current-configuration** command to display the IPv6 ACL rule configured on the multicast forwarding boundary. Change the IPv6 ACL rule used in the **multicast ipv6 boundary**

command so that the source address of the IPv6 multicast packets and the IPv6 multicast group address can both match the IPv6 ACL rule.

- Check the configuration of the multicast filter. Use the **display current-configuration** command to view the configuration of the IPv6 multicast filter, and change the IPv6 ACL rule used in the **source-policy** command so that the source address of the IPv6 multicast packets and the IPv6 multicast group address can both match the IPv6 ACL rule.

Table of Contents

1 MLD Configuration

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running the MLD protocol.

When configuring MLD, go to the following sections for information you are interested in:

MLD Overview

The Multicast Listener Discovery protocol (MLD) is used by an IPv6 router to discover the presence of multicast listeners on the directly attached subnets. Multicast listeners are nodes wishing to receive IPv6 multicast packets.

Through MLD, the router can learn whether there are any IPv6 multicast listeners on the directly connected subnets, put corresponding records in the database, and maintain timers related to IPv6 multicast addresses.

Routers running MLD use an IPv6 unicast link-local address as the source address to send MLD messages. MLD messages are Internet Control Message Protocol for IPv6 (ICMPv6) messages. All MLD messages are confined to the local subnet, with a hop count of 1.

MLD Versions

So far, two MLD versions are available:

- MLDv1 (defined in RFC 2710), which is derived from IGMPv2.
- MLDv2 (defined in RFC 3810), which is derived from IGMPv3.

All MLD versions support the Any-Source Multicast (ASM) model. In addition, MLDv2 can be directly deployed to implement the Source-Specific Multicast (SSM) model, while MLDv1 needs to work with the MLD SSM mapping function to implement SSM service.

Note

For more information about the ASM and SSM models, see *Multicast Overview* in the *IP Multicast Volume*.

How MLDv1 Works

MLDv1 implements IPv6 multicast listener management based on the query/response mechanism.

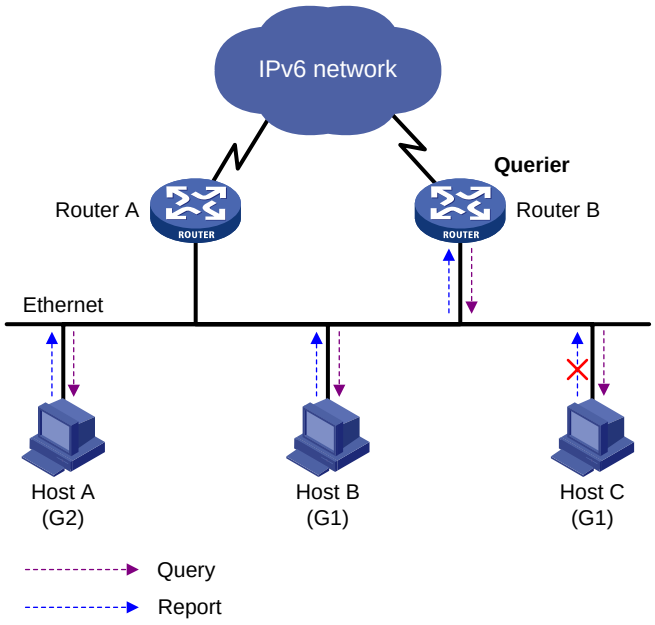
MLD querier election

Of multiple IPv6 multicast routers on the same subnet, all the routers can hear MLD listener report messages (often referred to as reports) from hosts, but only one router is needed for sending MLD query messages (often referred to as queries). So, a querier election mechanism is required to determine which router will act as the MLD querier on the subnet.

1) Initially, every MLD router assumes itself as the querier and sends MLD general query messages (often referred to as general queries) to all hosts and routers on the local subnet (the destination address is FF02::1).
2) Upon hearing a general query, every MLD router compares the source IPv6 address of the query message with its own interface address. After comparison, the router with the lowest IPv6 address wins the querier election and all other routers become non-queriers.
3) All the non-queriers start a timer, known as “other querier present timer”. If a router receives an MLD query from the querier before the timer expires, it resets this timer; otherwise, it assumes the querier to have timed out and initiates a new querier election process.

Joining an IPv6 multicast group

Figure 1-1 MLD queries and reports



Assume that Host B and Host C are expected to receive IPv6 multicast data addressed to IPv6 multicast group G1, while Host A is expected to receive IPv6 multicast data addressed to G2, as shown

in Figure 1-1. The following describes how the hosts join the IPv6 multicast groups and the MLD querier (Router B in the figure) maintains the IPv6 multicast group memberships:

1) The hosts send unsolicited MLD reports to the addresses of the IPv6 multicast groups that they want to join, without having to wait for the MLD queries from the MLD querier.
2) The MLD querier periodically multicasts MLD queries (with the destination address of FF02::1) to all hosts and routers on the local subnet.
3) Upon receiving a query message, Host B or Host C (the delay timer of whichever expires first) sends an MLD report to the IPv6 multicast group address of G1, to announce its membership for G1. Assume it is Host B that sends the report message. Upon hearing the report from Host B, Host C, which is on the same subnet with Host B, suppresses its own report for G1, because the MLD routers (Router A and Router B) already know that at least one host on the local subnet is interested in G1. This mechanism, known as MLD report suppression, helps reduce traffic on the local subnet.
4) At the same time, because Host A is interested in G2, it sends a report to the IPv6 multicast group address of G2.
5) Through the above-mentioned query/report process, the MLD routers learn that members of G1 and G2 are attached to the local subnet, and the IPv6 multicast routing protocol (IPv6 PIM for example) running on the routers generates (*, G1) and (*, G2) multicast forwarding entries, which will be the basis for subsequent IPv6 multicast forwarding, where * represents any IPv6 multicast source.
6) When the IPv6 multicast data addressed to G1 or G2 reaches an MLD router, because the (*, G1) and (*, G2) multicast forwarding entries exist on the MLD router, the router forwards the IPv6 multicast data to the local subnet, and then the receivers on the subnet receive the data.

Leaving an IPv6 multicast group

When a host leaves a multicast group:

1) This host sends an MLD done message to all IPv6 multicast routers (the destination address is FF02::2) on the local subnet.
2) Upon receiving the MLD done message, the querier sends a configurable number of multicast-address-specific queries to the group being left. The destination address field and group address field of the message are both filled with the address of the IPv6 multicast group being queried.
3) One of the remaining members, if any on the subnet, of the group being queried should send a report within the time of the maximum response delay set in the query messages.
4) If the querier receives a report for the group within the maximum response delay time, it will maintain the memberships of the IPv6 multicast group; otherwise, the querier will assume that no hosts on the subnet are still interested in IPv6 multicast traffic addressed to that group and will stop maintaining the memberships of the group.

How MLDv2 Works

Compared with MLDv1, MLDv2 provides the following new features:

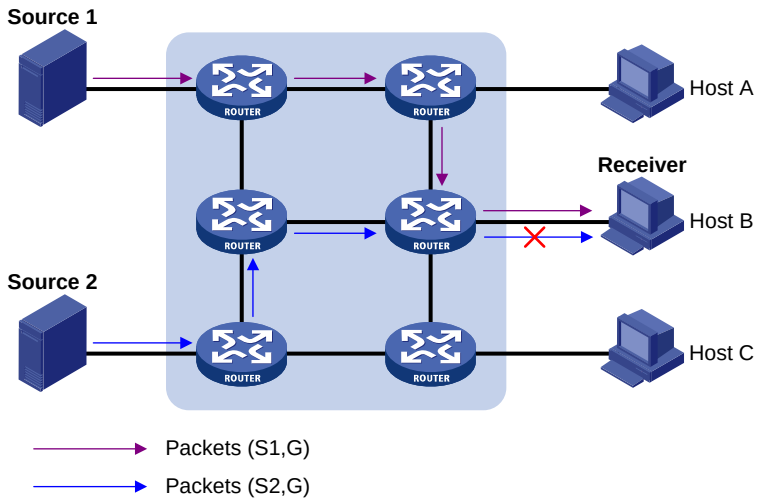
IPv6 multicast group filtering

MLDv2 has introduced IPv6 multicast source filtering modes (Include and Exclude), so that a host can specify a list of IPv6 multicast sources it expect or does not expect IPv6 multicast data from when it joins an IPv6 multicast group:

- If it expects IPv6 multicast data from specific IPv6 multicast sources like S1, S2, …, it sends a report with the Filter-Mode denoted as “Include Sources (S1, S2, …).
- If it does not expect IPv6 multicast data from specific IPv6 multicast sources like S1, S2, …, it sends a report with the Filter-Mode denoted as “Exclude Sources (S1, S2, …).

As shown in Figure 1-2, the network comprises two IPv6 multicast sources, Source 1 (S1) and Source 2 (S2), both of which can send IPv6 multicast data to IPv6 multicast group G. Host B is interested only in the IPv6 multicast data that Source 1 sends to G but not in the data from Source 2.

Figure 1-2 Flow paths of multicast-address-and-source-specific multicast traffic



In the case of MLDv1, Host B cannot select IPv6 multicast sources when it joins IPv6 multicast group G. Therefore, IPv6 multicast streams from both Source 1 and Source 2 will flow to Host B whether it needs them or not.

When MLDv2 is running on the hosts and routers, Host B can explicitly express its interest in the IPv6 multicast data Source 1 sends to G (denoted as (S1, G)), rather than the IPv6 multicast data Source 2 sends to G (denoted as (S2, G)). Thus, only IPv6 multicast data from Source 1 will be delivered to Host B.

MLD state

A multicast router running MLDv2 maintains the multicast address state per multicast address per attached subnet. The multicast address state consists of the following:

- Filter mode: The router keeps tracing the Include or Exclude state.
- List of sources: The router keeps tracing the newly added or deleted IPv6 multicast source.
- Timers: Filter timer (the time the router waits before switching to the Include mode after an IPv6 multicast address times out), source timer (for source recording), and so on.

Receiver host state listening

By listening to the state of receiver hosts, a multicast router running MLDv2 records and maintains information of hosts joining the source group on the attached subnet.

MLD Message Types

The following descriptions are based on MLDv2 messages.

MLD query message

An MLD querier learns the multicast listening state of neighbor interfaces by sending MLD query messages. Figure 1-3 shows the format of an MLD query message. The dark blue area in the figure shows the format of an MLDv1 message.

Figure 1-3 Format of MLDv2 query message

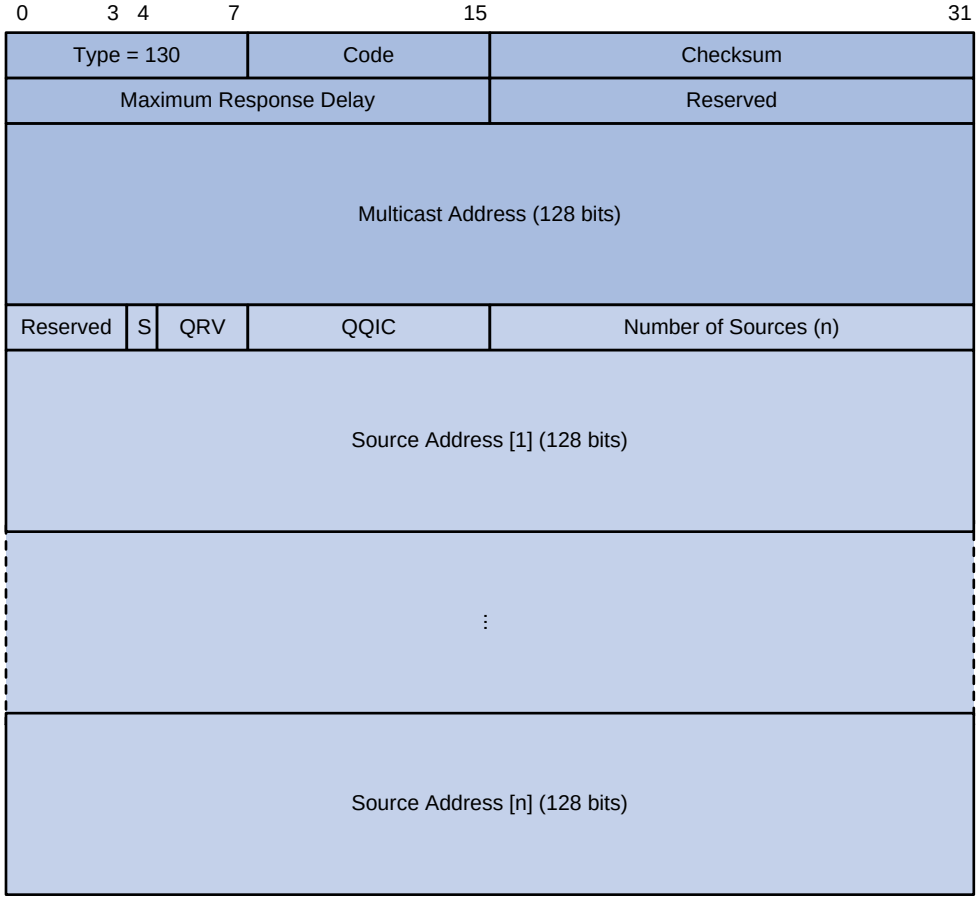


Table 1-1 describes the fields in Figure 1-3.

Table 1-1 Description on fields in an MLDv2 query message

Field	Description
Type = 130	Message type. For a query message, this field is set to 130.
Code	Initialized to zero
Checksum	Standard IPv6 checksum
Maximum Response Delay	Maximum response delay allowed before a host sends a report message
Reserved	Reserved field and initialized to zero
Multicast Address	• This field is set to 0 in a general query message. • It is set to a specific IPv6 multicast address in a multicast-address-specific query message or multicast-address-and-source-specific query message.
S	Flag indicating whether a router updates the timer for suppression after receiving a query message.
QRV	Querier's Robustness Variable

Field	Description
QQIC	Querier's Query Interval Code
Number of Sources	• This field is set to 0 in a general query message or a multicast-address-specific query message. • This field represents the number of source addresses in a multicast-address-and-source-specific query message
Source Address(i)	IPv6 multicast source address in a multicast-address-specific query message (i = 1, 2, .., n, where n represents the number of multicast source addresses.)

MLD report message

A host sends an MLD report message to report the current multicast listening state Figure 1-4 shows the format of an MLD report message.

Figure 1-4 Format of MLDv2 report message

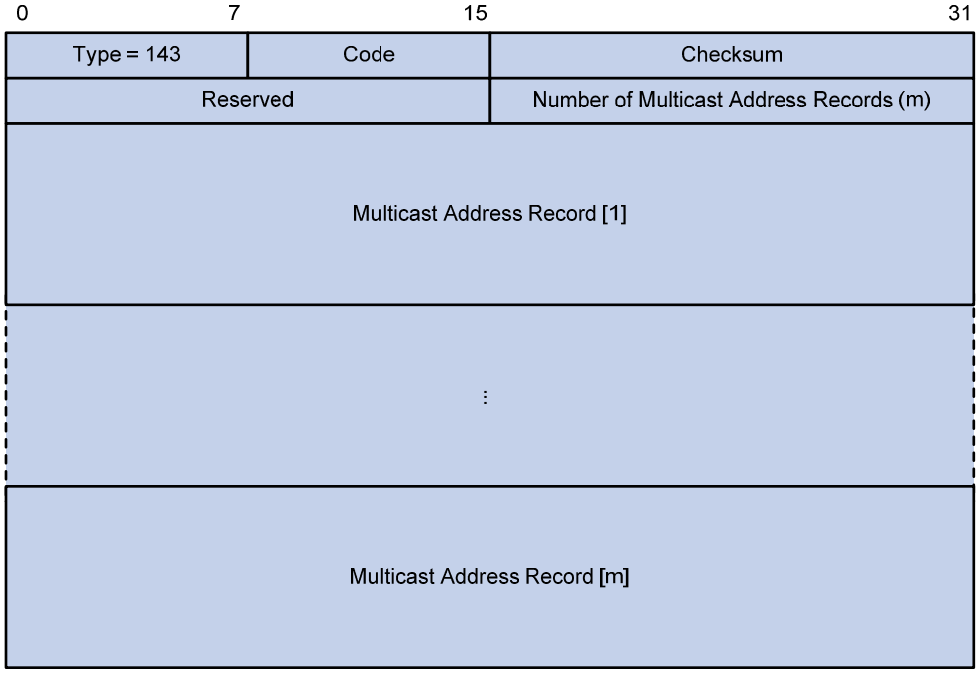


Table 1-2 describes the fields in Figure 1-4.

Table 1-2 Description on fields in an MLDv2 report message

Field	Description
Type = 143	Message type. For a report message, this field is set to 143.
Reserved	The Reserved fields are set to 0 on transmission and ignored on reception.
Checksum	Standard IPv6 checksum
Number of Multicast Address Records	This field indicates how many IPv6 multicast address records are present in this report message.

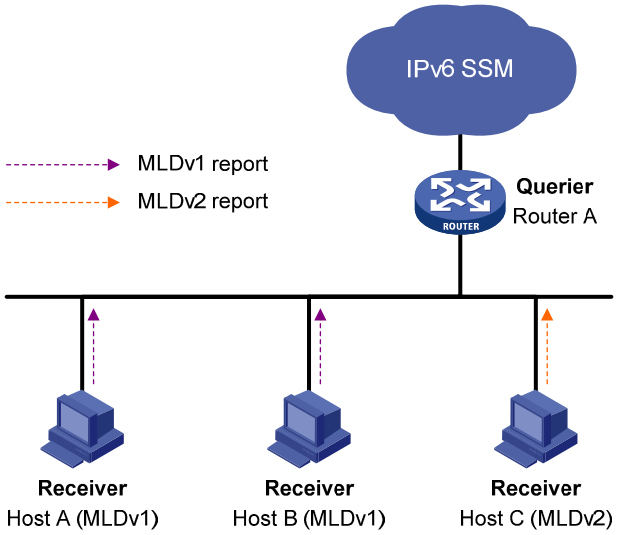
Field	Description
Multicast Address Record(i)	This field represents information of each IPv6 multicast address the host listens to on the interface from which the report message is sent, including record type, IPv6 multicast address, and IPv6 multicast source address on the sender (i= 1, 2, ... m, where m represents the number of IPv6 multicast address records).

MLD SSM Mapping

The MLD SSM mapping feature allows you to configure static MLD SSM mappings on the last hop router to provide SSM support for receiver hosts running MLDv1. The SSM model assumes that the last hop router is aware of the desired IPv6 multicast sources when receivers join IPv6 multicast groups.

- When a host running MLDv2 joins a multicast group, it can explicitly specify one or more multicast sources in its MLDv2 report.
- A host running MLDv1, however, cannot specify multicast source addresses in its MLDv1 report. In this case, you need to configure the MLD SSM mapping feature to translate the (*, G) information in the MLDv1 report into (G, INCLUDE, (S1, S2...)) information.

Figure 1-5 Network diagram for MLD SSM mapping



As shown in Figure 1-5, on an IPv6 SSM network, Host A and Host B are running MLDv1 and Host C is running MLDv2. To provide SSM service for all the hosts while it is infeasible to run MLDv2 on Host A and Host B, you need to configure the MLD SSM mapping feature on Router A.

With the MLD SSM mapping feature configured, when Router A receives an MLDv1 report, it checks the IPv6 multicast group address G carried in the message:

- If G is not in the IPv6 SSM group range, Router A cannot provide the SSM service but the ASM service.
- If G is in the IPv6 SSM group range but no MLD SSM mappings corresponding to the IPv6 multicast group G have been configured on Router A, Router A cannot provide SSM service and drops the packet.
- If G is in the IPv6 SSM group range, and the MLD SSM mappings have been configured on Router A for multicast group G, Router A translates the (*, G) information in the MLD report into (G,

INCLUDE, (S1, S2...)) information based on the configured MLD SSM mappings and provides SSM service accordingly.

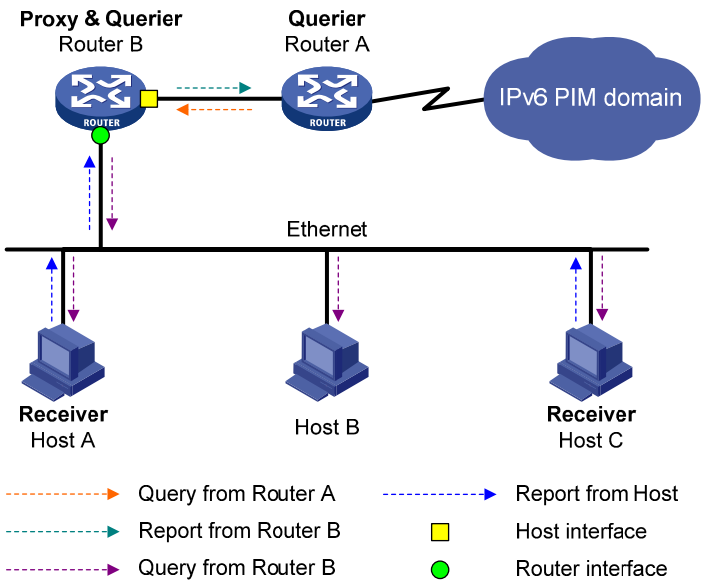
Note

- The MLD SSM mapping feature does not process MLDv2 reports.
- For more information about the IPv6 SSM group range, refer to *IPv6 PIM Configuration* in the *IP Multicast Volume*.

MLD Proxying

In some simple tree-shaped topologies, it is not necessary to configure complex IPv6 multicast routing protocols, such as IPv6 PIM, on the boundary device. Instead, you can configure MLD proxying on the boundary device. With MLD proxying configured, the device serves as a proxy for the downstream hosts to send MLD messages, maintain group memberships, and implement IPv6 multicast forwarding based on the memberships. In this case, the MLD proxy device is a host but no longer an IPv6 PIM neighbor to the upstream device.

Figure 1-6 Network diagram for MLD proxying



As shown in Figure 1-6, two types of interfaces are defined on a MLD proxy device:

- Upstream interface: Also referred to as the proxy interface. A proxy interface is an interface on which MLD proxying is configured. It is in the direction toward the root of the multicast forwarding tree. An upstream interface acts as a host running MLD; therefore, it is also called host interface.
- Downstream interface: An interface that is running MLD and not in the direction toward the root of the multicast forwarding tree. A downstream interface acts as a router running MLD; therefore, it is also called router interface.

A device with MLD proxying configured maintains a group membership database, which stores the group memberships on all the downstream interfaces in this database. Each entry comprises the

multicast address, filter mode, and source list. Such an entry is a collection of members in the same multicast group on each downstream interface.

A proxy device performs host functions on the upstream interface based on the database. It responds to the queries according to the information in the database or sends join/leave messages when the database changes. On the other hand, the proxy device performs router functions on the downstream interfaces by participating in the querier election, sending queries, and maintaining memberships based on the reports.

Protocols and Standards

MLD-related specifications are described in the following documents:

- RFC 2710: Multicast Listener Discovery (MLD) for IPv6
- RFC 3810: Multicast Listener Discovery Version 2 (MLDv2) for IPv6
- RFC 4605: Internet Group Management Protocol (IGMP)/Multicast Listener Discovery (MLD)-Based Multicast Forwarding (“IGMP/MLD Proxying”)

Configuration Task List

Task		Remarks
Configuring Basic Functions of MLD	Enabling MLD	Required
	Configuring the MLD Version	Option
	Configuring Static Joining	Optional
	Configuring an IPv6 Multicast Group Filter	Optional
	Configuring the Maximum Number of IPv6 Multicast Groups on an Interface	Optional
Adjusting MLD Performance	Configuring MLD Message Options	Optional
	Configuring MLD Query and Response Parameters	Optional
	Configuring MLD Fast Leave Processing	Optional
Configuring MLD SSM Mapping	Enabling MLD SSM Mapping	Optional
	Configuring MLD SSM Mappings	Optional
Configuring MLD Proxying	Enabling MLD Proxying	Optional
	Configuring IPv6 Multicast Forwarding on a Downstream Interface	Optional

Note

- Configurations performed in MLD view are globally effective, while configurations performed in interface view are effective on the current interface only.
- If no configuration is performed in interface view, the global configurations performed in MLD view will apply to that interface. Configurations performed in interface view take precedence over those performed in MLD view.

Configuring Basic Functions of MLD

Configuration Prerequisites

Before configuring the basic functions of MLD, complete the following tasks:

- Configure any IPv6 unicast routing protocol so that all devices in the domain can be interoperable at the network layer.
- Configure IPv6 PIM-DM or IPv6 PIM-SM.

In addition, prepare the following data:

- MLD version
- IPv6 multicast group address and IPv6 multicast source address for static group member configuration
- ACL rule for IPv6 multicast group filtering
- The maximum number of IPv6 multicast groups that can be joined on an interface

Enabling MLD

Enable MLD on the interface on which IPv6 multicast group memberships are to be created and maintained.

Follow these steps to enable MLD:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable IPv6 multicast routing	**multicast ipv6 routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable MLD	**mld enable**	Required Disabled by default

Note

For details about the **multicast ipv6 routing-table** command, see *IPv6 Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring the MLD Version

Because MLD message types and formats vary with MLD versions, the same MLD version should be configured for all routers on the same subnet before MLD can work properly.

Configuring an MLD version globally

Follow these steps to configure an MLD version globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter MLD view	**mld**	—
Configure an MLD version globally	**version** *version-number*	Optional MLDv1 by default

Configuring an MLD version on an interface

Follow these steps to configure an MLD version on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an MLD version on the interface	**mld version** *version-number*	Optional MLDv1 by default

Configuring Static Joining

After an interface is configured as a static member of an IPv6 multicast group or an IPv6 multicast source and group, it will act as a virtual member of the IPv6 multicast group to receive IPv6 multicast data addressed to that IPv6 multicast group for the purpose of testing IPv6 multicast data forwarding.

Follow these steps to configure a static member of an IPv6 multicast group or an IPv6 multicast source and group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a static member of an IPv6 multicast group or an IPv6 multicast source and group	**mld static-group** *ipv6-group-address* [**source** *ipv6-source-address*]	Required By default, an interface is not a static member of any IPv6 multicast group or IPv6 multicast source and group.

Note

- Before you can configure an interface of an IPv6 PIM-SM device as a static member of an IPv6 multicast group or an IPv6 multicast source and group, if the interface is IPv6 PIM-SM enabled, it must be an IPv6 PIM-SM DR; if this interface is MLD enabled but not IPv6 PIM-SM enabled, it must be an MLD querier. For more information about IPv6 PIM-SM and a DR, refer to *IPv6 PIM Configuration* in the *IP Multicast Volume*.
- As a static member of an IPv6 multicast group or an IPv6 multicast source and group, an interface does not respond to the queries from the MLD querier, nor does it send an unsolicited MLD membership report or a MLD done message when it joins or leaves an IPv6 multicast group or an IPv6 source and group. In other words, the interface will not become a real member of the IPv6 multicast group or the IPv6 multicast and source group.

Configuring an IPv6 Multicast Group Filter

To restrict the hosts on the network attached to an interface from joining certain IPv6 multicast groups, you can set an IPv6 ACL rule on the interface as a packet filter to limit the range of multicast groups that the interface serves.

Follow these steps to configure an IPv6 multicast group filter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an IPv6 multicast group filter	**mld group-policy** *acl6-number* [*version-number*]	Required By default, no IPv6 multicast filter is configured.

Configuring the Maximum Number of IPv6 Multicast Groups on an Interface

You can configure the allowed maximum number of the IPv6 multicast groups on an interface to flexibly control the number of IPv6 multicast groups the interface can join.

Follow these steps to configure the maximum number of IPv6 multicast groups an interface can join:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the maximum number of IPv6 multicast groups that can be joined on the current interface	**mld group-limit** *limit*	Required 1024 by default

Adjusting MLD Performance

Note

For the configuration tasks described in this section,

- Configurations performed in MLD view are globally effective, while configurations performed in interface view are effective on the current interface only.
- If the same function or parameter is configured in both PIM view and interface view, the configuration performed in interface view is given priority, regardless of the configuration sequence.

Configuration Prerequisites

Before adjusting MLD performance, complete the following tasks:

- Configure any IPv6 unicast routing protocol so that all devices in the domain can be interoperable at the network layer.
- Configure basic functions of MLD.

In addition, prepare the following data:

- Startup query interval
- Startup query count
- MLD query interval
- MLD querier robustness variable
- Maximum response delay of MLD general query messages
- MLD last listener query interval
- MLD other querier present interval

Configuring MLD Message Options

MLD queries include multicast-address-specific queries and multicast-address-and-source-specific queries, and IPv6 multicast groups change dynamically, so a device cannot maintain the information for all IPv6 multicast sources and groups. Therefore, a router may receive IPv6 multicast packets addressed to IPv6 multicast groups that have no members on the local subnet. In this case, the Router-Alert option carried in the IPv6 multicast packets is useful for the router to make a decision whether to deliver the IPv6 multicast packets to the upper-layer protocol for processing. For details about the Router-Alert option, refer to RFC 2113.

An MLD message is processed differently depending whether it carries the Router-Alert option in the IPv6 header:

- By default, in consideration of compatibility, the device does not check the Router-Alert option, that is, it processes all received MLD messages. In this case, the device passes MLD messages to the upper layer protocol for processing, no matter whether the MLD messages carry the Router-Alert option or not.
- To enhance the device performance, avoid unnecessary costs, and ensure the protocol security, you can configure the device to discard MLD messages without the Router-Alert option.

Configuring the Router-Alert option for MLD messages globally

Follow these steps to configure the Router-Alert option for MLD messages globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter MLD view	**mld**	—
Configure the interface to discard any MLD message without the Router-Alert option	**require-router-alert**	Optional By default, the device does not check MLD messages for the Router-Alert option.
Enable the insertion of the Router-Alert option into MLD messages	**send-router-alert**	Optional By default, MLD messages carry the Router-Alert option.

Configuring the Router-Alert option on an interface

Follow these steps to configure the Router-Alert option on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the interface to discard any MLD message without the Router-Alert option	**mld require-router-alert**	Optional By default, the device does not check MLD messages for the Router-Alert option.
Enable the insertion of the Router-Alert option into MLD messages	**mld send-router-alert**	Optional By default, MLD messages carry the Router-Alert option.

Configuring MLD Query and Response Parameters

On startup, the MLD querier sends “startup query count” MLD general queries at the “startup query interval”, which is 1/4 of the “MLD query interval”.

The MLD querier periodically sends MLD general queries at the “MLD query interval” to determine whether any IPv6 multicast group member exists on the network. You can modify the query interval based on the actual condition of the network.

Upon receiving an MLD done message, the MLD querier sends “last listener query count” MLD multicast-address-specific queries at the “MLD last listener query interval”. MLD is robust to “robustness variable minus 1” packet losses on a network. Therefore, a greater value of the robustness variable makes the MLD querier “more robust”, but results in a longer IPv6 multicast group timeout time.

Upon receiving an MLD query (general query or multicast-address-specific query) message, a host starts a timer for each IPv6 multicast group it has joined. The timer is initialized to a random value in the range of 0 to the maximum response delay (the host obtains the maximum response delay from the Maximum Response Delay field in the MLD query message it received). When the timer value drops to 0, the host sends an MLD membership report message to the corresponding IPv6 multicast group.

Proper setting of the maximum response delay of MLD query messages not only allows hosts to respond to MLD query messages quickly, but also avoids bursts of MLD traffic on the network caused by reports simultaneously sent by a large number of hosts when corresponding timers expire simultaneously.

- For MLD general queries, you can configure the maximum response delay to fill their Maximum Response Delay field.
- For MLD multicast-address-specific query messages, you can configure the last listener query interval to fill their Maximum Response Delay field. That is to say, the maximum response time of MLD general query messages equals the last listener query interval.

When multiple multicast routers exist on the same subnet, the MLD querier is responsible for sending MLD query messages. If a non-querier router receives no MLD query from the querier within the “other querier present interval”, it will assume that the querier has failed and a new querier election process is launched. Otherwise, the non-querier will reset “other querier present timer”.

Configuring MLD query and response parameters globally

Follow these steps to configure MLD query and response parameters globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter MLD view	**mld**	—
Configure the startup query interval	**startup-query-interval** *interval*	Optional For the system default, see “Note” below.
Configure the startup query count	**startup-query-count** *value*	Optional For the system default, see “Note” below.
Configure the MLD query interval	**timer query** *interval*	Optional 125 seconds by default.
Configure the MLD querier robustness variable	**robust-count** *robust-value*	Optional 2 times by default
Configure the maximum response delay for MLD general query messages	**max-response-time** *interval*	Optional 10 seconds by default
Configure the MLD last listener query interval	**last-listener-query-interval** *interval*	Optional 1 second by default
Configure the MLD other querier present interval	**timer other-querier-present** *interval*	Optional For the system default, see “Note” below.

Configuring MLD query and response parameters on an interface

Follow these steps to configure MLD query and response parameters on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the startup query interval	**mld startup-query-interval** *interval*	Optional For the system default, see “Note” below.
Configure the startup query count	**mld startup-query-count** *value*	Optional For the system default, see “Note” below.
Configure the MLD query interval	**mld timer query** *interval*	Optional 125 seconds by default.
Configure the MLD querier robustness variable	**mld robust-count** *robust-value*	Optional 2 times by default
Configure the maximum response delay for MLD general query messages	**mld max-response-time** *interval*	Optional 10 seconds by default
Configure the MLD last listener query interval	**mld last-listener-query-interval** *interval*	Optional 1 second by default
Configure the MLD other querier present interval	**mld timer other-querier-present** *interval*	Optional For the system default, see “Note” below.

Note

- If not statically configured, the startup query interval is 1/4 of the “MLD query interval”. By default, the MLD query interval is 125 seconds, so the startup query interval = 125 / 4 = 31.25 (seconds).
- If not statically configured, the startup query count is set to the MLD querier robustness variable. By default, the MLD querier robustness variable is 2, so the startup query count is also 2.
- If not statically configured, the other querier present interval is determined by the formula: Other querier present interval (in seconds) = [MLD query interval] times [MLD querier robustness variable] plus [maximum response delay for MLD general query] divided by two. The default values of these three parameters are 125, 2, and 10 respectively, so the other querier present interval = 125 x 2 + 10 / 2 = 255 (seconds).
- If statically configured, the startup query interval, the startup query count, and the other querier present interval take the configured values.

Caution

- Make sure that the other querier present interval is greater than the MLD query interval; otherwise the MLD querier may frequently change.
- Make sure that the MLD query interval is greater than the maximum response delay for MLD general queries; otherwise, multicast group members may be wrongly removed.

Configuring MLD Fast Leave Processing

In some applications, such as ADSL dial-up networking, only one multicast receiver host is attached to a port of the MLD querier. To allow fast response to the MLD done messages of the host when it switches frequently from one IPv6 multicast group to another, you can enable MLD fast leave processing on the MLD querier.

With fast leave processing enabled, after receiving an MLD done message from a host, the MLD querier sends a leave notification to the upstream immediately without first sending MLD multicast-address-specific queries. In this way, the leave latency is reduced on one hand, and the network bandwidth is saved on the other hand.

Follow these steps to configure MLD fast leave processing globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter MLD view	**mld**	—
Configure MLD fast leave processing	**fast-leave** [**group-policy** *acl6-number*]	Required Disabled by default.

Configuring MLD SSM Mapping

Due to some possible restrictions, some receiver hosts on an SSM network may run MLDv1. To provide SSM service support for these receiver hosts, you need to configure the MLD SSM mapping feature on the last hop router.

Configuration Prerequisites

Before configuring the MLD SSM mapping feature, complete the following tasks:

- Configure any IPv6 unicast routing protocol so that all devices in the domain can be interoperable at the network layer.
- Configure MLD basic functions

Enabling MLD SSM Mapping

Follow these steps to enable the MLD SSM mapping feature:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the MLD SSM mapping feature	**mld ssm-mapping enable**	Required Disabled by default

Note

To ensure SSM service for all hosts on a subnet, regardless of the MLD version running on the hosts, enable MLDv2 on the interface that forwards IPv6 multicast traffic onto the subnet.

Configuring MLD SSM Mappings

By performing this configuration multiple times, you can map an IPv6 multicast group to different IPv6 multicast sources.

Follow these steps to configure an MLD SSM mapping:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter MLD view	**mld**	—
Configure an MLD SSM mapping	**ssm-mapping** *ipv6-group-address prefix-length ipv6-source-address*	Required No MLD mappings are configured by default.

Caution

If MLDv2 is enabled on a VLAN interface of an 3Com Switch 4800G, and if a port in that VLAN is configured as a simulated host, the simulated host will send MLDv2 reports even if you did not specify an IPv6 multicast source when configuring simulated joining with the **mld-snooping host-join** command. In this case, the corresponding IPv6 multicast group will not be created based on the configured MLD SSM mappings. For details about the **mld-snooping host-join** command, refer to *MLD Snooping Commands* in the *IP Multicast Volume*.

Configuring MLD Proxying

Configuration Prerequisites

Before configuring the MLD proxying feature, complete the following tasks:

- Configure any IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Enable IPv6 multicast routing.

Enabling MLD Proxying

You can enable MLD proxying on the interface in the direction toward the root of the multicast forwarding tree to make the device serve as an MLD proxy.

Follow these steps to enable MLD proxying:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the MLD proxying feature	**mld proxying enable**	Required Disabled by default

Note

- Each device can have only one interface serving as the MLD proxy interface.
- You cannot enable MLD on interfaces with MLD proxying enabled. Moreover, only the **mld require-router-alert**, **mld send-router-alert**, and **mld version** commands can take effect on such interfaces.
- You cannot enable other IPv6 multicast routing protocols (such as IPv6 PIM-DM or IPv6-SM) on interfaces with MLD proxying enabled, or vice versa. However, the **source-lifetime**, **source-policy**, and **ssm-policy** commands configured in IPv6 PIM view can still take effect..
- You cannot enable MLD proxying on a VLAN interface with MLD Snooping enabled, or vice versa.

Configuring IPv6 Multicast Forwarding on a Downstream Interface

Typically, only queriers are able to forward IPv6 multicast traffic while non-queriers have no forwarding capabilities, to avoid duplicate multicast flows. It is the same on MLD proxy devices. Only the downstream interfaces acting as a querier can forward IPv6 multicast traffic to downstream hosts.

However, when a downstream interface of a proxy device fails to win the querier election, you need to enable IPv6 multicast forwarding on this interface.

Follow these steps to enable IPv6 multicast forwarding on a downstream interface

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable IPv6 multicast forwarding on a non-querier downstream interface	**mld proxying forwarding**	Required Disabled by default

Caution

On a multi-access network with more than one MLD proxy devices, you cannot enable IPv6 multicast forwarding on any other non-querier downstream interface after one of the downstream interfaces of these MLD proxy devices has been elected as the querier. Otherwise, duplicate multicast flows may be received on the multi-access network.

Displaying and Maintaining MLD Configuration

To do…	Use the command…	Remarks
View MLD multicast group information	**display mld group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]	Available in any view
View Layer 2 port information about MLD multicast groups	**display mld group port-info** [**vlan** *vlan-id*] [**slot** *slot-number*] [**verbose**]	Available in any view
View MLD configuration and running information on the specified interface or all MLD-enabled interfaces	**display mld interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View the information of the MLD proxying groups	**display mld proxying group** [*group-address*] [**verbose**]	Available in any view
View the information of the MLD routing table	**display mld routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*]] *	Available in any view
View MLD SSM mappings	**display mld ssm-mapping** *ipv6-group-address*	Available in any view
View the IPv6 multicast group information created based on the configured MLD SSM mappings	**display mld ssm-mapping group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**verbose**]	Available in any view
Clear MLD multicast group information	**reset mld group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }	Available in user view
Clear Layer 2 port information about MLD multicast groups	**reset mld group port-info** { **all** | *ipv6-group-address* } [**vlan** *vlan-id*]	Available in user view

To do…	Use the command…	Remarks
Clear MLD SSM mappings	**reset mld ssm-mapping group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }	Available in user view

Note

- You cannot use the **reset mld group** command to clear the MLD multicast group information of static joins.
- The **reset mld group port-info** command cannot clear Layer 2 port information about MLD multicast groups of static joins.

Caution

The **reset mld group** command cause an interruption of receivers' reception of multicast data.

MLD Configuration Examples

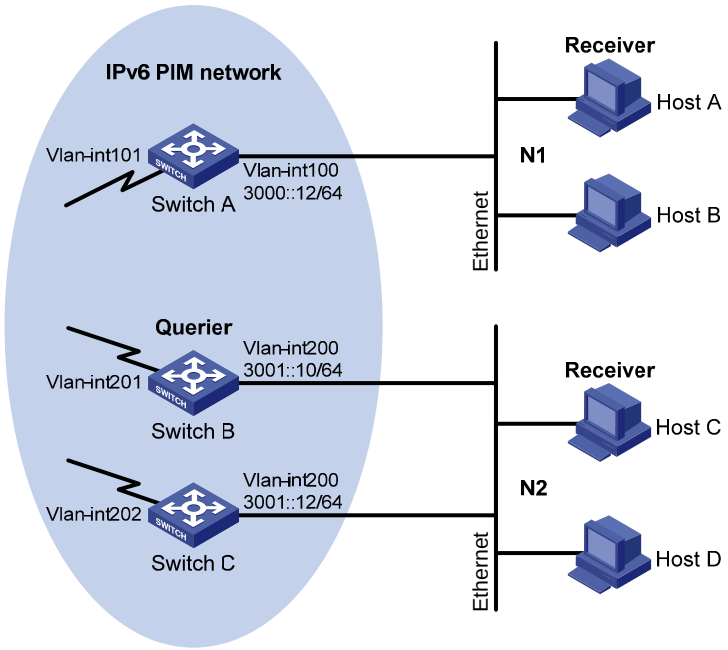
Basic MLD Functions Configuration Example

Network requirements

- Receivers receive VOD information in the multicast mode. Receivers of different organizations form stub networks N1 and N2, and Host A and Host C are multicast receivers in N1 and N2 respectively.
- Switch A in the IPv6 PIM network connects to N1, and Switch B and Switch C connect to N2.
- Switch A connects to N1 through VLAN-interface 100, and to other devices in the IPv6 PIM network through VLAN-interface 101.
- Switch B and Switch C connects to N2 through their own VLAN-interface 200, and to other devices in the IPv6 PIM network through VLAN-interface 201 and VLAN-interface 202 respectively.
- MLDv1 is required between Switch A and N1. MLDv1 is also required between the other two switches (Switch B and Switch C) and N2, Switch B serves as the MLD querier in N2 because its IP address is lower.

Network diagram

Figure 1-7 Network diagram for basic MLD functions configuration



Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses and IPv6 unicast routing

Enable IPv6 forwarding on each switch and configure an IP address and prefix length for each interface as shown in Figure 1-7. The detailed configuration steps are not discussed in this document.

Configure OSPFv3 for interoperation between the switches. Ensure the network-layer interoperation among the switches on the IPv6 PIM network and dynamic update of routing information between the switches through a unicast routing protocol. The detailed configuration steps are omitted here.

2) Enable the IPv6 multicast routing, and enable IPv6 PIM-DM and MLD.

Enable IPv6 multicast routing on Switch A, enable IPv6 PIM-DM on each interface, and enable MLD on VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] pim ipv6 dm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 dm
[SwitchA-Vlan-interface101] quit
```

Enable IPv6 multicast routing on Switch B, enable IPv6 PIM-DM on each interface, and enable MLD on VLAN-interface 200.

```
<SwitchB> system-view
[SwitchB] multicast ipv6 routing-enable
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mld enable
```

```
[SwitchB-Vlan-interface200] pim ipv6 dm
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 201
[SwitchB-Vlan-interface201] pim ipv6 dm
[SwitchB-Vlan-interface201] quit
```

Enable IPv6 multicast routing on Switch C, enable IPv6 PIM-DM on each interface, and enable MLD on VLAN-interface 200.

```
<SwitchC> system-view
[SwitchC] multicast ipv6 routing-enable
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] mld enable
[SwitchC-Vlan-interface200] pim ipv6 dm
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] pim ipv6 dm
[SwitchC-Vlan-interface202] quit
```

3) Verify the configuration

Carry out the **display mld interface** command to display the MLD configuration and running information on each switch interface. Example:

Display MLD information on VLAN-interface 200 of Switch B.

```
[SwitchB] display mld interface vlan-interface 200
 Vlan-interface200(FE80::200:5EFF:FE66:5100):
   MLD is enabled
   Current MLD version is 1
   Value of query interval for MLD(in seconds): 125
   Value of other querier present interval for MLD(in seconds): 255
   Value of maximum query response time for MLD(in seconds): 10
   Querier for MLD: FE80::200:5EFF:FE66:5100 (this router)
  Total 1 MLD Group reported
```

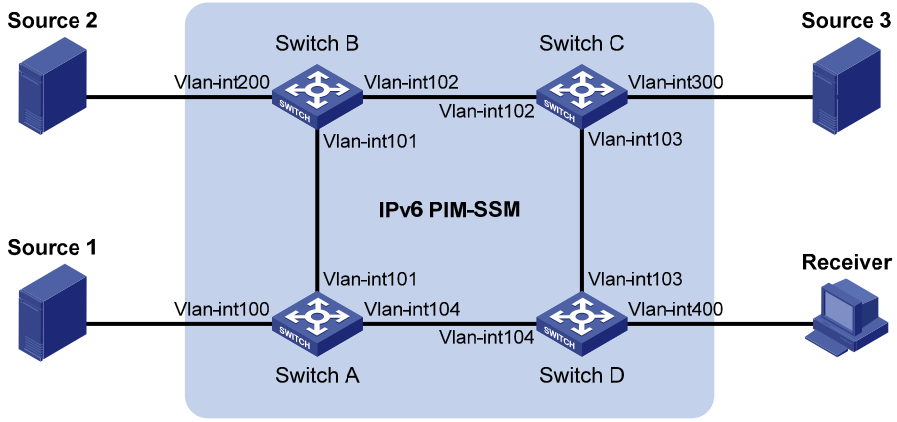
MLD SSM Mapping Configuration Example

Network requirements

- On the IPv6 PIM-SSM network shown in Figure 1-8, the receiver host receives VOD information through multicast. The receiver runs MLDv1, so it cannot specify the expected IPv6 multicast sources in its reports.
- It is required that MLD SSM mapping be configured on Switch D so that the receiver host will receive IPv6 multicast data from Source 1 and Source 3 only.

Network diagram

Figure 1-8 Network diagram for MLD SSM mapping configuration



Device	Interface	IP address	Device	Interface	IP address
Source 1	—	1001::1/64	Source 3	—	3001::1/64
Source 2	—	2001::1/64	Receiver	—	4001::1/64
Switch A	Vlan-int100	1001::2/64	Switch C	Vlan-int300	3001::2/64
	Vlan-int101	1002::1/64		Vlan-int103	3002::1/64
	Vlan-int104	1003::1/64		Vlan-int102	2002::2/64
Switch B	Vlan-int200	2001::2/64	Switch D	Vlan-int400	4001::2/64
	Vlan-int101	1002::2/64		Vlan-int103	3002::2/64
	Vlan-int102	2002::1/64		Vlan-int104	1003::2/64

Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses and IPv6 unicast routing

Enable IPv6 forwarding on each switch and configure an IPv6 address and prefix length for each interface as shown in Figure 1-8. The detailed configuration steps are omitted.

Configure OSPFv3 for interoperability among the switches. Ensure the network-layer interoperation on the IPv6 PIM-SSM network and dynamic update of routing information among the switches through a unicast routing protocol. The detailed configuration steps are omitted here.

2) Enable IPv6 multicast routing, enable IPv6 PIM-SM on each interface and enable MLD and MLD SSM mapping on the host-side interface.

Enable IPv6 multicast routing on Switch D, enable IPv6 PIM-SM on each interface, and enable MLD (version 2) and MLD SSM mapping on VLAN-interface 400.

```
<SwitchD> system-view
[SwitchD] multicast ipv6 routing-enable
[SwitchD] interface vlan-interface 400
[SwitchD-Vlan-interface400] mld enable
[SwitchD-Vlan-interface400] mld version 2
[SwitchD-Vlan-interface400] mld ssm-mapping enable
[SwitchD-Vlan-interface400] pim ipv6 sm
[SwitchD-Vlan-interface400] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim ipv6 sm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 104
```

```
[SwitchD-Vlan-interface104] pim ipv6 sm
[SwitchD-Vlan-interface104] quit
```

Enable IPv6 multicast routing on Switch A, and enable IPv6 PIM-SM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim ipv6 sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 104
[SwitchA-Vlan-interface104] pim ipv6 sm
[SwitchA-Vlan-interface104] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A.

3) Configure the IPv6 SSM group range

Configure the IPv6 SSM group range FF3E::/64 on Switch D.

```
[SwitchD] acl ipv6 number 2000
[SwitchD-acl6-basic-2000] rule permit source ff3e:: 64
[SwitchD-acl6-basic-2000] quit
[SwitchD] pim ipv6
[SwitchD-pim6] ssm-policy 2000
[SwitchD-pim6] quit
```

The configuration on Switch A, Switch B and Switch C is similar to that on Switch D.

4) Configure MLD SSM mappings

Configure MLD SSM mappings on Switch D.

```
[SwitchD] mld
[SwitchD-mld] ssm-mapping ff3e::101 128 1001::1
[SwitchD-mld] ssm-mapping ff3e::101 128 3001::1
[SwitchD-mld] quit
```

5) Verify the configuration

Use the **display mld ssm-mapping** command to view MLD SSM mappings on the switch.

View the MLD SSM mapping information for IPv6 multicast group FF3E::101 on Switch D.

```
[SwitchD] display mld ssm-mapping ff3e::101
Group: FF3E::101
Source list:
       1001::1
       3001::1
```

Use the **display mld ssm-mapping group** command to view information of the MLD multicast groups created based on the configured MLD SSM mappings.

View the IPv6 multicast group information created based on the configured MLD SSM mappings on Switch D.

```
[SwitchD] display mld ssm-mapping group
Total 1 MLD SSM-mapping Group(s).
```

```
Interface group report information
 Vlan-interface400 (4001::2):
  Total 1 MLD SSM-mapping Group reported
   Group Address: FF3E::101
    Last Reporter: 4001::1
    Uptime: 00:02:04
    Expires: off
```

Use the **display pim ipv6 routing-table** command to view the IPv6 PIM routing table information on each switch.

View the IPv6 PIM routing table information on Switch D.

```
[SwitchD] display pim ipv6 routing-table
Total 0 (*, G) entry; 2 (S, G) entry

 (1001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface104
         Upstream neighbor: 1003::1
         RPF prime neighbor: 1003::1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: Vlan-interface400
                 Protocol: mld, UpTime: 00:13:25, Expires: never

 (3001::1, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:13:25
     Upstream interface: Vlan-interface103
         Upstream neighbor: 3002::1
         RPF prime neighbor: 3002::1
     Downstream interface(s) information:
       Total number of downstreams: 1
         1: Vlan-interface400
                 Protocol: mld, UpTime: 00:13:25, Expires: never
```

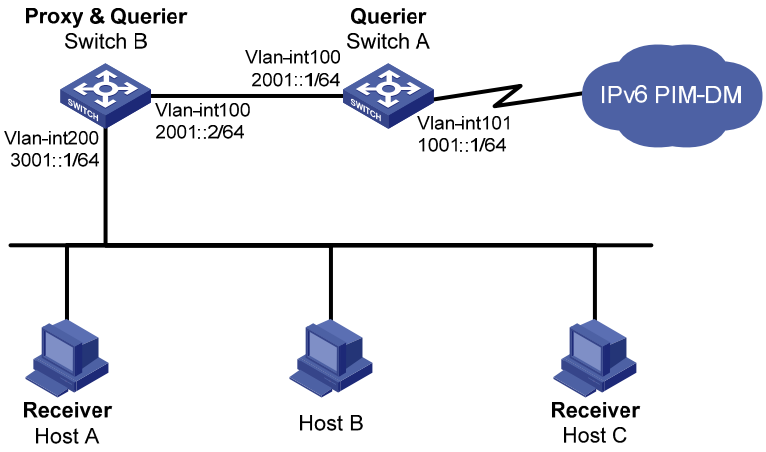
MLD Proxying Configuration Example

Network requirements

- IPv6 PIM-DM is required to run on the core network. Host A and Host C in the stub network receive VOD information destined to multicast group FF3E::101.
- It is required to configure the MLD proxying feature on Switch B so that Switch B can maintain group memberships and forward IPv6 multicast traffic without running IPv6 PIM-DM.

Network diagram

Figure 1-9 Network diagram for MLD proxying configuration



Configuration procedure

1) Enable IPv6 forwarding and configure the IPv6 addresses

Enable IPv6 forwarding on each switch and configure the IPv6 address and prefix length of each interface as per Figure 1-9. The detailed configuration steps are omitted here.

2) Enable IPv6 multicast routing, IPv6 PIM-DM, MLD, and MLD proxying respectively.

Enable IPv6 multicast routing on Switch A, IPv6 PIM-DM on VLAN-interface 101, and MLD on VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 dm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] pim ipv6 dm
[SwitchA-Vlan-interface100] quit
```

Enable IPv6 multicast routing on Switch B, MLD proxying on VLAN-interface 100, and MLD on VLAN-interface 200.

```
<SwitchB> system-view
[SwitchB] multicast ipv6 routing-enable
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] mld proxying enable
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mld enable
[SwitchB-Vlan-interface200] quit
```

3) Verify the installation

Use the **display mld interface** command to view the MLD configuration and operation information on an interface. For example,

View MLD configuration and operation information on VLAN-interface 100 of Switch B.

```
[SwitchB] display mld interface vlan-interface 100 verbose
 Vlan-interface100(2001::2):
   MLD proxy is enabled
   Current MLD version is 1
   Multicast routing on this interface: enabled
   Require-router-alert: disabled
```

Use the **display mld group** command to view MLD multicast group information. For example,

View the MLD multicast group information on Switch A.

```
[SwitchA] display mld group
Total 1 MLD Group(s).
Interface group report information
 Vlan-interface100(2001::1):
  Total 1 MLD Groups reported
   Group Address   Last Reporter   Uptime      Expires
   ff3e::101       2001::2         00:02:04    00:01:15
```

As shown above, the MLD reports sent from the hosts are forwarded to Switch A through the proxy interface, VLAN-interface 100 of Switch B.

Troubleshooting MLD

No Member Information on the Receiver-Side Router

Symptom

When a host sends a message for joining IPv6 multicast group G, there is no member information of multicast group G on the immediate router.

Analysis

- The correctness of networking and interface connections and whether the protocol layer of the interface is up directly affect the generation of IPv6 group member information.
- IPv6 multicast routing must be enabled on the router and MLD must be enabled on the interface connecting to the host.
- If the MLD version on the router interface is lower than that on the host, the router will not be able to recognize the MLD report from the host.
- If the **mld group-policy** command has been configured on an interface, the interface cannot receive report messages that fail to pass filtering.

Solution

1) Check that the networking, interface connections, and IP address configuration are correct. Check the interface information with the **display mld interface** command. If there is no information output, the interface is in an abnormal state. This is usually because you have configured the **shutdown** command on the interface, the interface is not properly connected, or the IPv6 address configuration is not correctly done.
2) Check that the IPv6 multicast routing is enabled. Carry out the **display current-configuration** command to check whether the **multicast ipv6 routing-enable** command has been executed. If not, carry out the **multicast ipv6 routing-enable** command in system view to enable IPv6 multicast routing. In addition, enable MLD on the corresponding interface.

3) Check the MLD version on the interface. You can use the **display mld interface** command to check whether the MLD version on the interface is lower than that on the host.
4) Check that no ACL rule has been configured to restrict the host from joining IPv6 multicast group G. Carry out the **display current-configuration interface** command to check whether the **mld group-policy** command has been executed. If an IPv6 ACL is configured to restrict the host from joining IPv6 multicast group G, the ACL must be modified to allow IPv6 multicast group G to receive report messages.

Inconsistent Memberships on Routers on the Same Subnet

Symptom

Different memberships are maintained on different MLD routers on the same subnet.

Analysis

- A router running MLD maintains multiple parameters for each interface, and these parameters influence one another, forming very complicated relationships. Inconsistent MLD interface parameter configurations for routers on the same subnet will surely result in inconsistent MLD memberships.
- Two MLD versions are currently available. Although routers running different MLD versions are compatible with hosts, all routers on the same subnet must run the same MLD version. Inconsistent MLD versions running on routers on the same subnet will also lead to inconsistent MLD memberships.

Solution

1) Check MLD configurations. Carry out the **display current-configuration** command to display the MLD configuration information on the interface.
2) Carry out the **display mld interface** command on all routers on the same subnet to check the MLD timers for consistent configurations.
3) Use the **display mld interface** command to check that the routers are running the same MLD version.

Table of Contents

1 IPv6 PIM Configuration

When configuring IPv6 PIM, go to these sections for information you are interested in:

Note

The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running IPv6 PIM.

IPv6 PIM Overview

Protocol Independent Multicast for IPv6 (IPv6 PIM) provides IPv6 multicast forwarding by leveraging static routes or IPv6 unicast routing tables generated by any IPv6 unicast routing protocol, such as RIPng, OSPFv3, IS-ISv6, or BGP4+. IPv6 PIM uses an IPv6 unicast routing table to perform reverse path forwarding (RPF) check to implement IPv6 multicast forwarding. Independent of the IPv6 unicast routing protocols running on the device, IPv6 multicast routing can be implemented as long as the corresponding IPv6 multicast routing entries are created through IPv6 unicast routes. IPv6 PIM uses the reverse path forwarding (RPF) mechanism to implement IPv6 multicast forwarding. When an IPv6 multicast packet arrives on an interface of the device, it is subject to an RPF check. If the RPF check succeeds, the device creates the corresponding routing entry and forwards the packet; if the RPF check fails, the device discards the packet. For more information about RPF, refer to *IPv6 Multicast Routing and Forwarding Configuration* in the *IP Multicast Volume*.

Based on the implementation mechanism, IPv6 PIM falls into two modes:

- Protocol Independent Multicast–Dense Mode for IPv6 (IPv6 PIM-DM), and
- Protocol Independent Multicast–Sparse Mode for IPv6 (IPv6 PIM-SM).

Note

To facilitate description, a network comprising IPv6 PIM–supporting routers is referred to as an “IPv6 PIM domain” in this document.

Introduction to IPv6 PIM-DM

IPv6 PIM-DM is a type of dense mode IPv6 multicast protocol. It uses the “push mode” for IPv6 multicast forwarding, and is suitable for small-sized networks with densely distributed IPv6 multicast members.

The basic implementation of IPv6 PIM-DM is as follows:

- IPv6 PIM-DM assumes that at least one IPv6 multicast group member exists on each subnet of a network, and therefore IPv6 multicast data is flooded to all nodes on the network. Then, branches without IPv6 multicast forwarding are pruned from the forwarding tree, leaving only those branches that contain receivers. This “flood and prune” process takes place periodically, that is, pruned branches resume IPv6 multicast forwarding when the pruned state times out and then data is re-flooded down these branches, and then are pruned again.
- When a new receiver on a previously pruned branch joins an IPv6 multicast group, to reduce the join latency, IPv6 PIM-DM uses the graft mechanism to resume IPv6 multicast data forwarding to that branch.

Generally speaking, the IPv6 multicast forwarding path is a source tree, namely a forwarding tree with the IPv6 multicast source as its “root” and IPv6 multicast group members as its “leaves”. Because the source tree is the shortest path from the IPv6 multicast source to the receivers, it is also called shortest path tree (SPT).

How IPv6 PIM-DM Works

The working mechanism of IPv6 PIM-DM is summarized as follows:

- Neighbor discovery
- SPT establishment
- Graft
- Assert

Neighbor discovery

In an IPv6 PIM domain, a PIM router discovers IPv6 PIM neighbors, maintains IPv6 PIM neighboring relationships with other routers, and builds and maintains SPTs by periodically multicasting IPv6 PIM hello messages (hereinafter referred to as “hello messages”) to all other IPv6 PIM routers.

Note

Every IPv6 PIM enabled interface on a router sends hello messages periodically, and thus learns the IPv6 PIM neighboring information pertinent to the interface.

SPT establishment

The process of constructing an SPT is the “flood and prune” process.

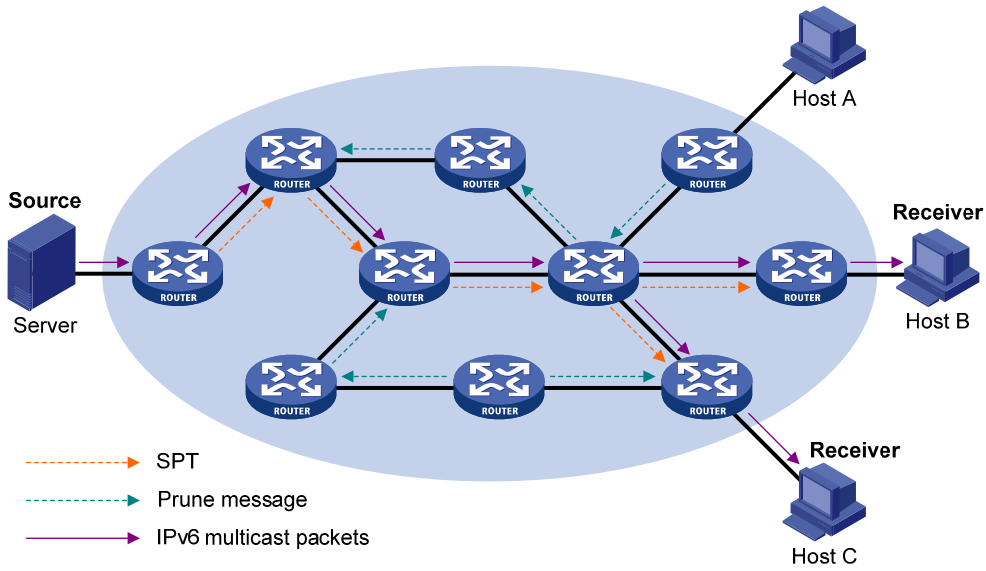
1) In an IPv6 PIM-DM domain, an IPv6 multicast source first floods IPv6 multicast packets when it sends IPv6 multicast data to IPv6 multicast group G: The packet is subject to an RPF check. If the packet passes the RPF check, the router creates an (S, G) entry and forwards the packet to all downstream nodes in the network. In the flooding process, an (S, G) entry is created on all the routers in the IPv6 PIM-DM domain.
2) Then, nodes without downstream receivers are pruned: A router having no down stream receivers sends a prune message to the upstream node to notify the upstream node to delete the corresponding interface from the outgoing interface list in the (S, G) entry and stop forwarding subsequent packets addressed to that IPv6 multicast group down to this node.

Note

- An (S, G) entry contains the multicast source address S, IPv6 multicast group address G, outgoing interface list, and incoming interface.
- For a given IPv6 multicast stream, the interface that receives the IPv6 multicast stream is referred to as “upstream”, and the interfaces that forward the IPv6 multicast stream are referred to as “downstream”.

A prune process is first initiated by a leaf router. As shown in Figure 1-1, a router without any receiver attached to it (the router connected with Host A, for example) sends a prune message, and this prune process goes on until only necessary branches are left in the IPv6 PIM-DM domain. These branches constitute the SPT.

Figure 1-1 SPT establishment in an IPv6 PIM-DM domain



The “flood and prune” process takes place periodically. A pruned state timeout mechanism is provided. A pruned branch restarts multicast forwarding when the pruned state times out and then is pruned again when it no longer has any multicast receiver.

Note

Pruning has a similar implementation in IPv6 PIM-SM.

Graft

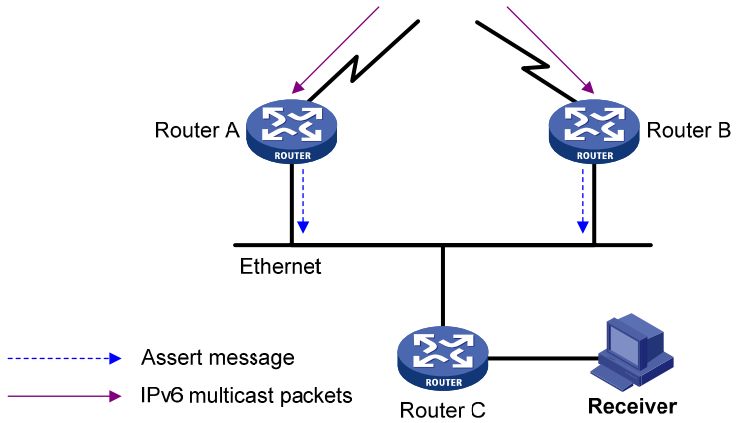
When a host attached to a pruned node joins an IPv6 multicast group, to reduce the join latency, IPv6 PIM-DM uses the graft mechanism to resume IPv6 multicast data forwarding to that branch. The process is as follows:

1) The node that needs to receive IPv6 multicast data sends a graft message toward its upstream node, as a request to join the SPT again.
2) Upon receiving this graft message, the upstream node puts the interface on which the graft was received into the forwarding state and responds with a graft-ack message to the graft sender.
3) If the node that sent a graft message does not receive a graft-ack message from its upstream node, it will keep sending graft messages at a configurable interval until it receives an acknowledgment from its upstream node.

Assert

The assert mechanism is used to shutoff duplicate IPv6 multicast flows onto the same multi-access network, where more than one multicast routers exists, by electing a unique IPv6 multicast forwarder on the multi-access network.

Figure 1-2 Assert mechanism



As shown in Figure 1-2, after Router A and Router B receive an (S, G) IPv6 multicast packet from the upstream node, they both forward the packet to the local subnet. As a result, the downstream node Router C receives two identical multicast packets, and both Router A and Router B, on their own local interface, receive a duplicate IPv6 multicast packet forwarded by the other. Upon detecting this condition, both routers send an assert message to all IPv6 PIM routers through the interface on which the packet was received. The assert message contains the following information: the multicast source

address (S), the multicast group address (G), and the preference and metric of the IPv6 unicast route to the source. By comparing these parameters, either Router A or Router B becomes the unique forwarder of the subsequent (S, G) IPv6 multicast packets on the multi-access subnet. The comparison process is as follows:

1) The router with a higher IPv6 unicast route preference to the source wins;
2) If both routers have the same IPv6 unicast route preference to the source, the router with a smaller metric to the source wins;
3) If there is a tie in the route metric to the source, the router with a higher IPv6 link-local address wins.

Introduction to IPv6 PIM-SM

IPv6 PIM-DM uses the “flood and prune” principle to build SPTs for IPv6 multicast data distribution. Although an SPT has the shortest path, it is built with a low efficiency. Therefore the PIM-DM mode is not suitable for large-and medium-sized networks.

IPv6 PIM-SM is a type of sparse mode IPv6 multicast protocol. It uses the “pull mode” for IPv6 multicast forwarding, and is suitable for large- and medium-sized networks with sparsely and widely distributed IPv6 multicast group members.

The basic implementation of IPv6 PIM-SM is as follows:

- IPv6 PIM-SM assumes that no hosts need to receive IPv6 multicast data. In the IPv6 PIM-SM mode, routers must specifically request a particular IPv6 multicast stream before the data is forwarded to them. The core task for IPv6 PIM-SM to implement IPv6 multicast forwarding is to build and maintain rendezvous point trees (RPTs). An RPT is rooted at a router in the IPv6 PIM domain as the common node, or rendezvous point (RP), through which the IPv6 multicast data travels along the RPT and reaches the receivers.
- When a receiver is interested in the IPv6 multicast data addressed to a specific IPv6 multicast group, the router connected to this receiver sends a join message to the RP corresponding to that IPv6 multicast group. The path along which the message goes hop by hop to the RP forms a branch of the RPT.
- When an IPv6 multicast source sends IPv6 multicast streams to an IPv6 multicast group, the source-side designated router (DR) first registers the multicast source with the RP by sending register messages to the RP by unicast until it receives a register-stop message from the RP. The arrival a register message at the RP triggers the establishment of an SPT. Then, the IPv6 multicast source sends subsequent IPv6 multicast packets along the SPT to the RP. Upon reaching the RP, the IPv6 multicast packet is duplicated and delivered to the receivers along the RPT.

Note

IPv6 multicast traffic is duplicated only where the distribution tree branches, and this process automatically repeats until the IPv6 multicast traffic reaches the receivers.

How IPv6 PIM-SM Works

The working mechanism of IPv6 PIM-SM is summarized as follows:

- Neighbor discovery
- DR election
- RP discovery
- Embedded RP
- RPT establishment
- IPv6 Multicast source registration
- Switchover to SPT
- Assert

Neighbor discovery

IPv6 PIM-SM uses the similar neighbor discovery mechanism as IPv6 PIM-DM does. Refer to Neighbor discovery.

DR election

IPv6 PIM-SM also uses hello messages to elect a DR for a multi-access network (such as a LAN). The elected DR will be the only IPv6 multicast forwarder on this multi-access network.

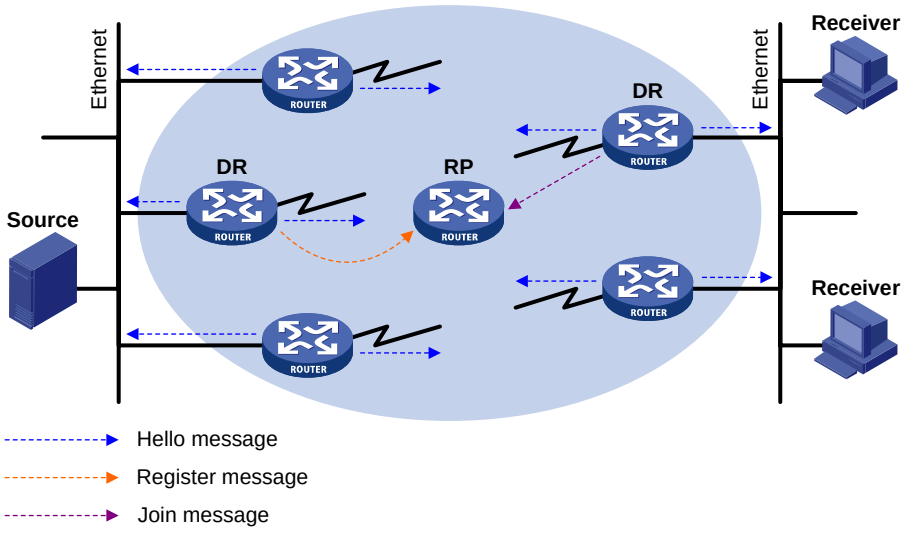
In the case of a multi-access network, a DR must be elected, no matter this network connects to IPv6 multicast sources or to receivers. The DR at the receiver side sends join messages to the RP; the DR at the IPv6 multicast source side sends register messages to the RP.

Note

- A DR is elected on a multi-access subnet by means of comparison of the priorities and IPv6 link-local addresses carried in hello messages.
- MLD must be enabled on a device that acts as a receiver-side DR before receivers attached to this device can join IPv6 multicast groups through this DR.

For details about MLD, refer to *MLD Configuration* in the *IP Multicast Volume*.

Figure 1-3 DR election



As shown in Figure 1-3, the DR election process is as follows:

1) Routers on the multi-access network send hello messages to one another. The hello messages contain the router priority for DR election. The router with the highest DR priority will become the DR.
2) In the case of a tie in the router priority, or if any router in the network does not support carrying the DR-election priority in hello messages, The router with the highest IPv6 link-local address will win the DR election.

When the DR works abnormally, a timeout in receiving hello message triggers a new DR election process among the other routers.

RP discovery

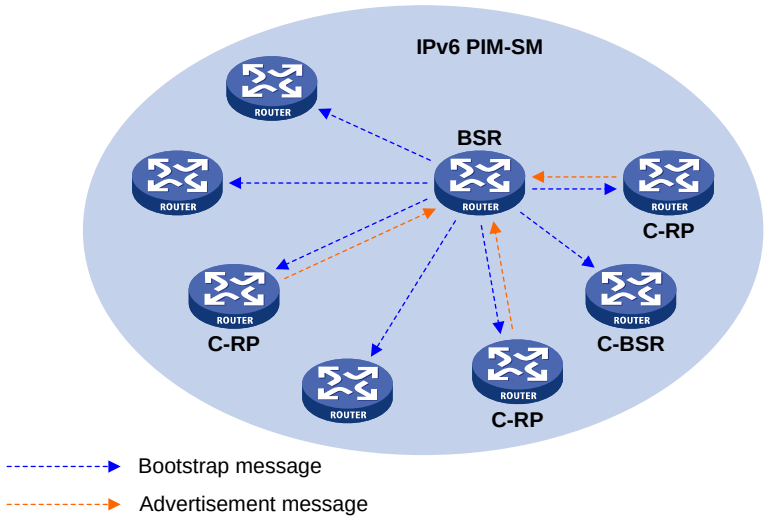
The RP is the core of an IPv6 PIM-SM domain. For a small-sized, simple network, one RP is enough for forwarding IPv6 multicast information throughout the network, and the position of the RP can be statically specified on each router in the IPv6 PIM-SM domain. In most cases, however, an IPv6 PIM-SM network covers a wide area and a huge amount of IPv6 multicast traffic needs to be forwarded through the RP. To lessen the RP burden and optimize the topological structure of the RPT, multiple candidate RPs (C-RPs) can be configured in an IPv6 PIM-SM domain, among which an RP is dynamically elected through the bootstrap mechanism. Each elected RP serves a different multicast group range. For this purpose, a bootstrap router (BSR) must be configured. The BSR serves as the administrative core of the IPv6 PIM-SM domain. An IPv6 PIM-SM domain can have only one BSR, but can have multiple candidate-BSRs (C-BSRs). Once the BSR fails, a new BSR is automatically elected from the C-BSRs to avoid service interruption.

Note

- An RP can serve IPv6 multiple multicast groups or all IPv6 multicast groups. Only one RP can serve a given IPv6 multicast group at a time.
- A device can server as a C-RP and a C-BSR at the same time.

As shown in the figure below, each C-RP periodically unicasts its advertisement messages (C-RP-Adv messages) to the BSR. A C-RP-Adv message contains the address of the advertising C-RP and the IPv6 multicast group range it serves. The BSR collects these advertisement messages and chooses the appropriate C-RP information for each multicast group to form an RP-set, which is a database of mappings between IPv6 multicast groups and RPs. The BSR then encapsulates the RP-set in the bootstrap messages it periodically originates and floods the bootstrap messages to the entire IPv6 PIM-SM domain.

Figure 1-4 BSR and C-RPs



Based on the information in the RP-sets, all routers in the network can calculate the location of the corresponding RPs based on the following rules:

1) The C-RP with the highest priority wins.
2) If all the C-RPs have the same priority, their hash values are calculated through the hashing algorithm. The C-RP with the largest hash value wins.
3) If all the C-RPs have the same priority and hash value, the C-RP has the highest IP address wins.

The hashing algorithm used for RP calculation is: Value (G, M, C_i) = (1103515245 * ((1103515245 * (G & M) + 12345) XOR C_i) + 12345) mod 2^{31}. The table below gives the meanings of the values in this algorithm.

Table 1-1 Values in the hashing algorithm

Value	Description
Value	Hash value
G	The digest from the exclusive-or (XOR) operation between the 32-bit segments of the IPv6 multicast group address. For example, if the IPv6 multicast address is FF0E:C20:1A3:63::101, G = 0xFF0E0C20 XOR 0x01A30063 XOR 0x00000000 XOR 0x00000101
M	Hash mask length
C_i	The digest from the exclusive-or (XOR) operation between the 32-bit segments of the C-RP IPv6 address. For example, if the IPv6 address of the C-RP is 3FFE:B00:C18:1::10, C_i = 0x3FFE0B00 XOR 0x0C180001 XOR 0x00000000 XOR 0x00000010
&	Logical operator of “and”
XOR	Logical operator of “exclusive-or”
mod	Modulo operator, which gives the remainder of an integer division

Embedded RP

The Embedded RP mechanism allows a router to resolve the RP address from an IPv6 multicast address so that the IPv6 multicast group is mapped to an RP, which can take the place of the statically

configured RP or the RP dynamically calculated based on the BSR mechanism. The DR does not need to know the RP address beforehand. The specific process is as follows.

- At the receiver side:

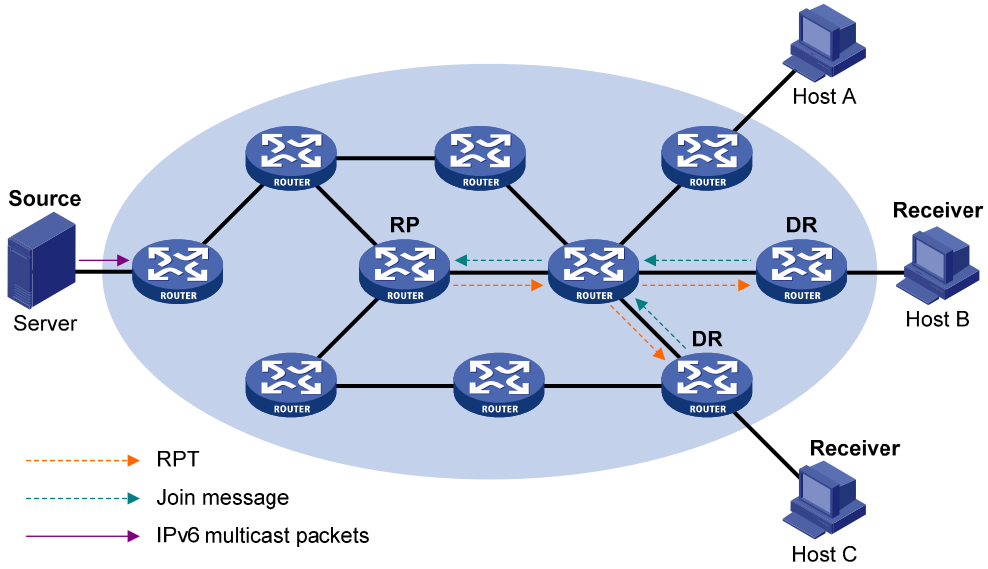
1) A receiver host initiates an MLD report to announce its joining an IPv6 multicast group.
2) Upon receiving the MLD report, the receiver-side DR resolves the RP address embedded in the IPv6 multicast address, and sends a join message to the RP.

- At the IPv6 multicast source side:

1) The IPv6 multicast source sends IPv6 multicast traffic to the IPv6 multicast group.
2) The source-side DR resolves the RP address embedded in the IPv6 multicast address, and sends a register message to the RP.

RPT establishment

Figure 1-5 RPT establishment in an IPv6 PIM-SM domain



As shown in Figure 1-5, the process of building an RPT is as follows:

1) When a receiver joins IPv6 multicast group G, it uses an MLD report message to inform the directly connected DR.
2) Upon getting the IPv6 multicast group G's receiver information, the DR sends a join message, which is hop by hop forwarded to the RP corresponding to the multicast group.
3) The routers along the path from the DR to the RP form an RPT branch. Each router on this branch generates a (*, G) entry in its forwarding table. The * means any IPv6 multicast source. The RP is the root, while the DRs are the leaves, of the RPT.

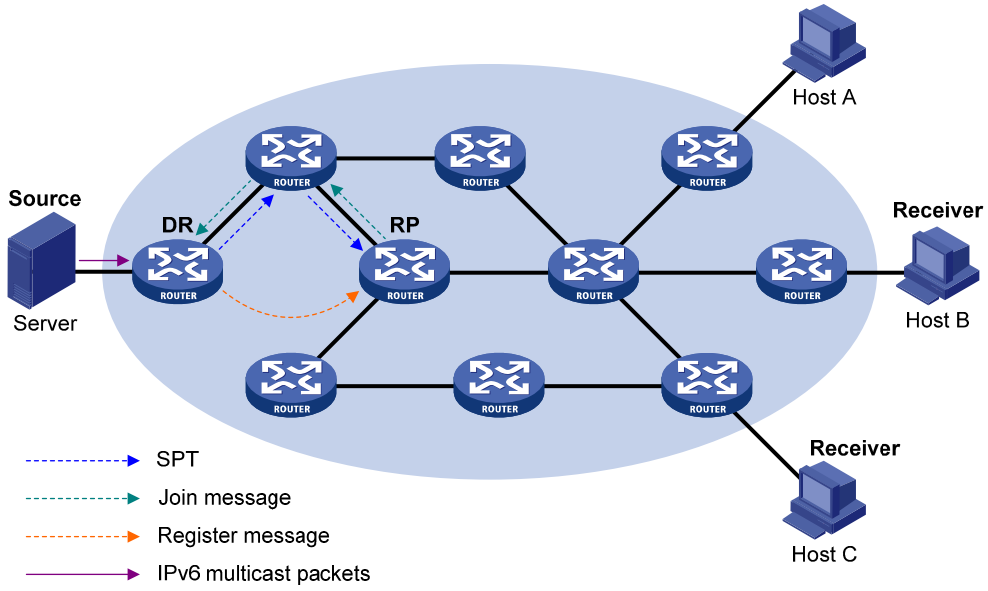
The IPv6 multicast data addressed to the IPv6 multicast group G flows through the RP, reaches the corresponding DR along the established RPT, and finally is delivered to the receiver.

When a receiver is no longer interested in the IPv6 multicast data addressed to a multicast group G, the directly connected DR sends a prune message, which goes hop by hop along the RPT to the RP. Upon receiving the prune message, the upstream node deletes the interface connected with this downstream node from the outgoing interface list and checks whether it has receivers for that IPv6 multicast group. If not, the router continues to forward the prune message to its upstream router.

Multicast source registration

The purpose of IPv6 multicast source registration is to inform the RP about the existence of the IPv6 multicast source.

Figure 1-6 IPv6 multicast source registration



As shown in Figure 1-6, the IPv6 multicast source registers with the RP as follows:

1) When the IPv6 multicast source S sends the first IPv6 multicast packet to IPv6 multicast group G, the DR directly connected with the multicast source, upon receiving the multicast packet, encapsulates the packet in a register message, and sends the message to the corresponding RP by unicast.
2) When the RP receives the register message, it extracts the multicast packet from the register message and forwards the multicast IPv6 multicast packet down the RPT, and sends an (S, G) join message hop by hop toward the IPv6 multicast source. Thus, the routers along the path from the RP to the IPv6 multicast source form an SPT branch. Each router on this branch generates an (S, G) entry in its forwarding table. The DR at the IPv6 multicast source side is the root, while the RP is the leaf, of the SPT.
3) The subsequent IPv6 multicast data from the IPv6 multicast source travels along the established SPT to the RP, and then the RP forwards the data along the RPT to the receivers. When the IPv6 multicast traffic arrives at the RP along the SPT, the RP sends a register-stop message to the source-side DR by unicast to stop the source registration process.

Note

The RP is configured to initiate an SPT switchover as described in this section. Otherwise, the DR at the IPv6 multicast source side keeps encapsulating IPv6 multicast data in register messages and the registration process will not stop unless no outgoing interfaces exist in the (S, G) entry on the RP.

Switchover to SPT

In a IPv6 PIM-SM domain, a IPv6 multicast group corresponds to one RP and RPT. Before the SPT switchover takes place, the DR at the IPv6 multicast source side encapsulates all IPv6 multicast data destined to the multicast group in register messages and sends these messages to the RP. Upon receiving these register messages, the RP abstracts the IPv6 multicast data and sends the IPv6 multicast data down the RPT to the DRs at the receiver side. The RP acts as a transfer station for all IPv6 multicast packets. The whole process involves three issues as follows:

- The DR at the source side and the RP need to implement complicated encapsulation and decapsulation of IPv6 multicast packets.
- IPv6 Multicast packets are delivered along a path that is not necessarily the shortest one.
- When the IPv6 multicast traffic increases, a great burden is added to the RP, increasing the risk of failure.

To solve the issues, IPv6 PIM-SM allows an RP or the DR at the receiver side to initiate an SPT switchover process:

1) The RP initiates an SPT switchover process

Upon receiving the first IPv6 multicast packet, the RP sends an (S, G) join message hop by hop toward the IPv6 multicast source to establish an SPT between the DR at the source side and the RP. The subsequent IPv6 multicast data from the multicast source travel along the established SPT to the RP.

Note

For details about the SPT switchover initiated by the RP, refer to Multicast source registration.

---

2) The receiver-side DR initiates an SPT switchover process

Upon discovering that the traffic rate exceeds a configurable threshold, the receiver-side DR initiates an SPT switchover process, as follows:

- First, the receiver-side DR sends an (S, G) join message hop by hop toword the multicast source S. When the join message reaches the source-side DR, all the routers on the path have installed the (S, G) entry in their forwarding table, and thus an SPT branch is established.
- When subsequent IPv6 multicast packets arrive at the router at the junction of the RPT and SPT, the router drops those transmitted along the RPT and sends an RP-bit prune message containing the RP bit hop by hop to the RP. Upon receiving this prune message, the RP sends a prune message toward the IPv6 multicast source (suppose only one receiver exists), thus to implement SPT switchover.

IPv6 PIM-SM builds SPTs through SPT switchover more economically than IPv6 PIM-DM does through the “flood and prune” mechanism.

Assert

IPv6 PIM-SM uses the similar assert mechanism as IPv6 PIM-DM does. Refer to Assert.

SSM Model Implementation in IPv6 PIM

The source-specific multicast (SSM) model and the any-source multicast (ASM) model are two opposite models. Presently, the ASM model includes the IPv6 PIM-DM and IPv6 PIM-SM modes. The SSM model can be implemented by leveraging part of the IPv6 PIM-SM technique.

The SSM model provides a solution for source-specific multicast. It maintains the relationships between hosts and routers through MLDv2. IPv6 PIM-DM implements IPv6 multicast forwarding by building SPTs rooted at the IPv6 multicast source through the “flood and prune” mechanism. Although an SPT has the shortest path, it is built in a low efficiency. Therefore the IPv6 PIM-DM mod is not suitable for large- and medium-sized networks.

In actual application, part of the IPv6 PIM-SM technique is adopted to implement the SSM model. In the SSM model, receivers know exactly where an IPv6 multicast source is located by means of advertisements, consultancy, and so on. Therefore, no RP is needed, no RPT is required, and is no source registration process is needed for the purpose of discovering IPv6 multicast sources in other IPv6 PIM domains.

Compared with the ASM model, the SSM model only needs the support of MLDv2 and some subsets of IPv6 PIM-SM. The operation mechanism of the SSM model in an IPv6 PIM domain can be summarized as follows:

- Neighbor discovery
- DR election
- SPT building

Neighbor discovery

IPv6 PIM-SSM uses the same neighbor discovery mechanism as in IPv6 PIM-SM. Refer to Neighbor discovery.

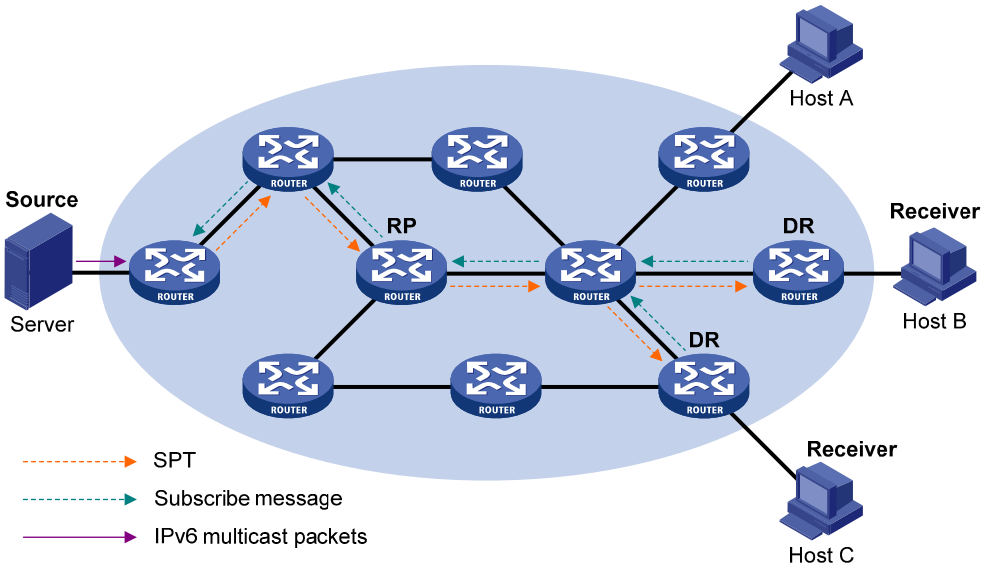
DR election

IPv6 PIM-SSM uses the same DR election mechanism as in IPv6 PIM-SM. Refer to DR election.

SPT building

Whether to build an RPT for IPv6 PIM-SM or an SPT for IPv6 PIM-SSM depends on whether the IPv6 multicast group the receiver is to join falls in the IPv6 SSM group range (the IPv6 SSM group range reserved by IANA is FF3x::/32, where x represents any legal address scope).

Figure 1-7 Building an SPT in IPv6 PIM-SSM



As shown in Figure 1-7, Hosts B and C are IPv6 multicast information receivers. They send an MLDv2 report message to the respective DRs to announce that they are interested in the information of the specific IPv6 multicast source S and that sent to the IPv6 multicast group G.

The DR that has received the report first checks whether the IPv6 group address in this message falls in the IPv6 SSM group range:

- If so, the IPv6 PIM-SSM model is built: the DR sends a channel subscription message hop by hop toward the IPv6 multicast source S. An (S, G) entry is created on all routers on the path from the DR to the source. Thus, an SPT is built in the network, with the source S as its root and receivers as its leaves. This SPT is the transmission channel in IPv6 PIM-SSM.
- If not, the IPv6 PIM-SM process is followed: the DR needs to send a (*, G) join message to the RP, and an IPv6 multicast source registration process is needed.

Note

In IPv6 PIM-SSM, the “channel” concept is used to refer to an IPv6 multicast group, and the “channel subscription” concept is used to refer to a join message.

Protocols and Standards

IPv6 PIM–related specifications are as follows:

- RFC 4601: Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)
- RFC 3973: Protocol Independent Multicast-Dense Mode(PIM-DM):Protocol Specification(Revised)
- RFC 3956: Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address
- RFC 4607: Source-Specific Multicast for IP
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- draft-ietf-ssm-overview-05: An Overview of Source-Specific Multicast (SSM)

Configuring IPv6 PIM-DM

IPv6 PIM-DM Configuration Task List

Complete these tasks to configure IPv6 PIM-DM:

Task	Remarks
Enabling IPv6 PIM-DM	Required
Enabling State-Refresh Capability	Optional
Configuring State Refresh Parameters	Optional
Configuring IPv6 PIM-DM Graft Retry Period	Optional
Configuring IPv6 PIM Common Features	Optional

Configuration Prerequisites

Before configuring IPv6 PIM-DM, complete the following task:

- Configure any IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring IPv6 PIM-DM, prepare the following data:

- The interval between state refresh messages
- Minimum time to wait before receiving a new refresh message
- Hop limit value of state-refresh messages
- Graft retry period

Enabling IPv6 PIM-DM

With IPv6 PIM-DM enabled, a router sends hello messages periodically to discover IPv6 PIM neighbors and processes messages from the IPv6 PIM neighbors. When deploying an IPv6 PIM-DM domain, you are recommended to enable IPv6 PIM-DM on all non-border interfaces of routers.

Follow these steps to enable IPv6 PIM-DM:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IPv6 multicast routing	**multicast ipv6 routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable IPv6 PIM-DM	**pim ipv6 dm**	Required Disabled by default

Caution

- All the interfaces of the same device must work in the same IPv6 PIM mode.
- IPv6 PIM-DM cannot be used for IPv6 multicast groups in the IPv6 SSM group range.

Note

For details about the **multicast ipv6 routing-enable** command, see *IPv6 Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Enabling State-Refresh Capability

A multi-access subnet can have the state-refresh capability only if the state-refresh capability is enabled on all IPv6 PIM routers on the subnet.

Follow these steps to enable the state-refresh capability:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable the state-refresh capability	**pim ipv6 state-refresh-capable**	Optional Enabled by default

Configuring State Refresh Parameters

To avoid the resource-consuming reflooding of unwanted traffic caused by timeout of pruned interfaces, the router directly connected with the IPv6 multicast source periodically sends an (S, G) state-refresh message, which is forwarded hop by hop along the initial flooding path of the IPv6 PIM-DM domain, to refresh the prune timer state of all the routers on the path.

A router may receive multiple state-refresh messages within a short time, of which some may be duplicated messages. To keep a router from receiving such duplicated messages, you can configure the time the router must wait before receiving the next state-refresh message. If a new state-refresh message is received within the waiting time, the router will discard it; if this timer times out, the router will accept a new state-refresh message, refresh its own IPv6 PIM-DM state, and reset the waiting timer.

The hop limit value of a state-refresh message decrements by 1 whenever it passes a router before it is forwarded to the downstream node until the hop limit value comes down to 0. In a small network, a state-refresh message may cycle in the network. To effectively control the propagation scope of state-refresh messages, you need to configure an appropriate hop limit value based on the network size.

It is recommended to perform the following configurations on all routers in the IPv6 PIM domain.

Follow these steps to configure state-refresh parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the interval between state-refresh messages	**state-refresh-interval** *interval*	Optional 60 seconds by default
Configure the time to wait before receiving a new state-refresh message	**state-refresh-rate-limit** *interval*	Optional 30 seconds by default
Configure the hop limit value of state-refresh messages	**state-refresh-hoplimit** *hoplimit-value*	Optional 255 by default

Configuring IPv6 PIM-DM Graft Retry Period

In IPv6 PIM-DM, graft is the only type of message that uses the acknowledgment mechanism. In an IPv6 PIM-DM domain, if a router does not receive a graft-ack message from the upstream router within the specified time after it sends a graft message, the router keeps sending new graft messages at a configurable interval, namely graft retry period, until it receives a graft-ack from the upstream router.

Follow these steps to configure IPv6 PIM-DM graft retry period:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure graft retry period	**pim ipv6 timer graft-retry** *interval*	Optional 3 seconds by default

Note

For the configuration of other timers in IPv6 PIM-DM, refer to Configuring IPv6 PIM Common Timers.

Configuring IPv6 PIM-SM

IPv6 PIM-SM Configuration Task List

Complete these tasks to configure IPv6 PIM-SM:

Task		Remarks
Enabling IPv6 PIM-SM		Required
Configuring an RP	Configuring a static RP	Optional
	Configuring a C-RP	Optional
	Enabling embedded RP	Optional
	Configuring C-RP timers globally	Optional

Task		Remarks
Configuring a BSR	Configuring a C-BSR	Optional
	Configuring an IPv6 PIM domain border	Optional
	Configuring C-BSR parameters globally	Optional
	Configuring C-BSR timers	Optional
Configuring IPv6 Multicast Source Registration		Optional
Disabling SPT Switchover		Optional
Configuring IPv6 PIM Common Features		Optional

Configuration Prerequisites

Before configuring IPv6 PIM-SM, complete the following task:

- Configure any IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring IPv6 PIM-SM, prepare the following data:

- The IP address of a static RP and an ACL rule defining the range of IPv6 multicast groups to be served by the static RP
- C-RP priority and an ACL rule defining the range of IPv6 multicast groups to be served by each C-RP
- A legal C-RP address range and an ACL rule defining the range of IPv6 multicast groups to be served
- C-RP-Adv interval
- C-RP timeout
- C-BSR priority
- Hash mask length
- An IPv6 ACL rule defining a legal BSR address range
- BS period
- BS timeout
- An IPv6 ACL rule for register message filtering
- Register suppression time
- Register probe time
- The IPv6 ACL rule and sequencing rule for SPT switchover

Enabling IPv6 PIM-SM

With IPv6 PIM-SM enabled, a router sends hello messages periodically to discover IPv6 PIM neighbors and processes messages from the IPv6 PIM neighbors. When deploying an IPv6 PIM-SM domain, you are recommended to enable IPv6 PIM-SM on all non-border interfaces of the routers.

Follow these steps to enable IPv6 PIM-SM:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IPv6 multicast routing	**multicast ipv6 routing-enable**	Required Disable by default

To do...	Use the command...	Remarks
Enter interface view	**interface** *interface-type interface-number*	—
Enable IPv6 PIM-SM	**pim ipv6 sm**	Required Disabled by default

Caution

All the interfaces of the same device must work in the same IPv6 PIM mode.

Note

For details about the **multicast ipv6 routing-enable** command, see *IPv6 Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring an RP

An RP can be manually configured or dynamically elected through the BSR mechanism. For a large IPv6 PIM network, static RP configuration is a tedious job. Generally, static RP configuration is just a backup means for the dynamic RP election mechanism to enhance the robustness and operation manageability of a multicast network.

Configuring a static RP

If there is only one dynamic RP in a network, manually configuring a static RP can avoid communication interruption due to single-point failures and avoid frequent message exchange between C-RPs and the BSR.

Perform the following configuration on all the routers in the IPv6 PIM-SM domain.

Follow these steps to configure a static RP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure a static RP	**static-rp** *ipv6-rp-address* [*acl6-number*] [**preferred**]	Required No static RP by default

Caution

To enable a static RP to work normally, you must perform this configuration on all routers in the IPv6 PIM-SM domain and specify the same RP address.

Configuring a C-RP

In an IPv6 PIM-SM domain, you can configure routers that intend to become the RP as C-RPs. The BSR collects the C-RP information by receiving the C-RP-Adv messages from C-RPs or auto-RP announcements from other routers and organizes the information into an RP-Set, which is flooded throughout the entire network. Then, the other routers in the network calculate the mappings between specific group ranges and the corresponding RPs based on the RP-Set. We recommend that you configure C-RPs on backbone routers.

To guard against C-RP spoofing, you need to configure a legal C-RP address range and the range of IPv6 multicast groups to be served on the BSR. In addition, because every C-BSR has a chance to become the BSR, you need to configure the same filtering policy on all C-BSRs in the IPv6 PIM-SM domain.

Follow these steps to configure a C-RP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure an interface to be a C-RP	**c-rp** *ipv6-address* [**group-policy** *acl6-number* | **priority** *priority* | **holdtime** *hold-interval* | **advertisement-interval** *adv-interval*] *	Required No C-RPs are configured by default.
Configure a legal C-RP address range and the range of IPv6 multicast groups to be served	**crp-policy** *acl6-number*	Optional No restrictions by default

Note

- When configuring a C-RP, ensure a relatively large bandwidth between this C-RP and the other devices in the IPv6 PIM-SM domain.
- An RP can serve multiple IPv6 multicast groups or all IPv6 multicast groups. Only one RP can forward IPv6 multicast traffic for an IPv6 multicast group at a moment.

Enabling embedded RP

With the Embedded RP feature enabled, the router can resolve the RP address directly from the IPv6 multicast group address of an IPv6 multicast packets. This RP can replace the statically configured RP or the RP dynamically calculated based on the BSR mechanism. Thus, the DR does not need to know the RP address beforehand.

Follow these steps to enable embedded RP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—

To do...	Use the command...	Remarks
Enable embedded RP	**embedded-rp** [*acl6-number*]	Optional By default, embedded RP is enabled for IPv6 multicast groups in the default embedded RP address scopes.

Note

The default embedded RP address scopes are FF7x::/12 and FFFx::/12. Here “x” refers to any legal address scope. For details of the scope field, see *Multicast Overview* of the *IP Multicast Volume*.

Configuring C-RP timers globally

To enable the BSR to distribute the RP-Set information within the IPv6 PIM-SM domain, C-RPs must periodically send C-RP-Adv messages to the BSR. The BSR learns the RP-Set information from the received messages, and encapsulates its own IPv6 address together with the RP-Set information in its bootstrap messages. The BSR then floods the bootstrap messages to all IPv6 routers in the network.

Each C-RP encapsulates a timeout value in its C-RP-Adv messages. Upon receiving a C-RP-Adv message, the BSR obtains this timeout value and starts a C-RP timeout timer. If the BSR fails to hear a subsequent C-RP-Adv message from the C-RP when the timer times out, the BSR assumes the C-RP to have expired or become unreachable.

The C-RP timers need to be configured on C-RP routers.

Follow these steps to configure C-RP timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the C-RP-Adv interval	**c-rp advertisement-interval** *interval*	Optional 60 seconds by default
Configure C-RP timeout time	**c-rp holdtime** *interval*	Optional 150 seconds by default

Note

For the configuration of other timers in IPv6 PIM-SM, refer to Configuring IPv6 PIM Common Timers.

Configuring a BSR

An IPv6 PIM-SM domain can have only one BSR, but must have at least one C-BSR. Any router can be configured as a C-BSR. Elected from C-BSRs, the BSR is responsible for collecting and advertising RP information in the IPv6 PIM-SM domain.

Configuring a C-BSR

C-BSRs should be configured on routers in the backbone network. When configuring a router as a C-BSR, make sure to specify the IPv6 address of an IPv6 PIM-SM enabled interface on the router. The BSR election process is summarized as follows:

- Initially, every C-BSR assumes itself to be the BSR of this IPv6 PIM-SM domain, and uses its interface IPv6 address as the BSR address to send bootstrap messages.
- When a C-BSR receives the bootstrap message of another C-BSR, it first compares its own priority with the other C-BSR's priority carried in the message. The C-BSR with a higher priority wins. If there is a tie in the priority, the C-BSR with a higher IPv6 address wins. The loser uses the winner's BSR address to replace its own BSR address and no longer assumes itself to be the BSR, while the winner keeps its own BSR address and continues assuming itself to be the BSR.

Configuring a legal range of BSR addresses enables filtering of bootstrap messages based on the address range, thus to prevent a maliciously configured host from masquerading as a BSR. The same configuration needs to be made on all routers in the IPv6 PIM-SM domain. The following are typical BSR spoofing cases and the corresponding preventive measures:

1) Some maliciously configured hosts can forge bootstrap messages to fool routers and change RP mappings. Such attacks often occur on border routers. Because a BSR is inside the network whereas hosts are outside the network, you can protect a BSR against attacks from external hosts by enabling the border routers to perform neighbor checks and RPF checks on bootstrap messages and discard unwanted messages.
2) When a router in the network is controlled by an attacker or when an illegal router is present in the network, the attacker can configure this router as a C-BSR and make it win BSR election to control the right of advertising RP information in the network. After being configured as a C-BSR, a router automatically floods the network with bootstrap messages. As a bootstrap message has a hop limit value of 1, the whole network will not be affected as long as the neighbor router discards these bootstrap messages. Therefore, with a legal BSR address range configured on all routers in the entire network, all these routers will discard bootstrap messages from out of the legal address range.

The above-mentioned preventive measures can partially protect the security of BSRs in a network. However, if a legal BSR is controlled by an attacker, the above-mentioned problem will also occur.

Follow these steps to complete basic BSR configuration:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure an interface as a C-BSR	**c-bsr** *ipv6-address* [*hash-length* [*priority*]]	Required No C-BSRs are configured by default.
Configure a legal BSR address range	**bsr-policy** *acl6-number*	Optional No restrictions by default

Note

Since a large amount of information needs to be exchanged between a BSR and the other devices in the IPv6 PIM-SM domain, a relatively large bandwidth should be provided between the C-BSR and the other devices in the IPv6 PIM-SM domain.

Configuring an IPv6 PIM domain border

As the administrative core of an IPv6 PIM-SM domain, the BSR sends the collected RP-Set information in the form of bootstrap messages to all routers in the IPv6 PIM-SM domain.

An IPv6 PIM domain border is a bootstrap message boundary. Each BSR has its specific service scope. A number of IPv6 PIM domain border interfaces partition a network into different IPv6 PIM-SM domains. Bootstrap messages cannot cross a domain border in either direction.

Perform the following configuration on routers that can become an IPv6 PIM domain border.

Follow these steps to configure an IPv6 PIM border domain:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configuring an IPv6 PIM domain border	**pim ipv6 bsr-boundary**	Required No IPv6 PIM domain border is configured by default

Configuring C-BSR parameters globally

In each IPv6 PIM-SM domain, a unique BSR is elected from C-BSRs. The C-RPs in the IPv6 PIM-SM domain send advertisement messages to the BSR. The BSR summarizes the advertisement messages to form an RP-set and advertises it to all routers in the IPv6 PIM-SM domain. All the routers use the same Hash algorithm to get the RP address corresponding to specific IPv6 multicast groups.

Perform the following configuration on C-BSR routers.

Follow these steps to configure C-BSR parameters globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the Hash mask length	**c-bsr hash-length** *hash-length*	Optional 126 by default
Configure the C-BSR priority	**c-bsr priority** *priority*	Optional 0 by default

Configuring C-BSR timers

The BSR election winner multicasts its own IPv6 address and RP-Set information throughout the region that it serves through bootstrap messages. The BSR floods bootstrap messages throughout the network at the interval of BS (BSR state) period. Any C-BSR that receives a bootstrap message retains the RP-set for the length of BS timeout, during which no BSR election takes place. If the BSR state times out and no bootstrap message is received from the BSR, a new BSR election process is triggered among the C-BSRs.

Perform the following configuration on C-BSR routers.

Follow these steps to configure C-BSR timers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the BS period	**c-bsr interval** *interval*	Optional For the default value, see the note below.
Configure the BS timeout	**c-bsr holdtime** *interval*	Optional For the default value, see the note below.

Note

About the BS period:

- By default, the BS period is determined by this formula: BS period = (BS timeout – 10) / 2. The default BS timeout is 130 seconds, so the default BS period = (130 – 10) / 2 = 60 (seconds).
- If this parameter is manually configured, the system will use the configured value.

About the BS timeout:

- By default, the BS timeout value is determined by this formula: BS timeout = BS period × 2 + 10. The default BS period is 60 seconds, so the default BS timeout = 60 × 2 + 10 = 130 (seconds).
- If this parameter is manually configured, the system will use the configured value.

Caution

In configuration, make sure that the BS period is smaller than the BS timeout value.

Configuring IPv6 Multicast Source Registration

Within an IPv6 PIM-SM domain, the source-side DR sends register messages to the RP, and these register messages have different IPv6 multicast source or IPv6 multicast group addresses. You can configure a filtering rule to filter register messages so that the RP can serve specific IPv6 multicast groups. If an (S, G) entry is denied by the filtering rule, or the action for this entry is not defined in the filtering rule, the RP will send a register-stop message to the DR to stop the registration process for the IPv6 multicast data.

In view of information integrity of register messages in the transmission process, you can configure the device to calculate the checksum based on the entire register messages. However, to reduce the workload of encapsulating data in register messages and for the sake of interoperability, this method of checksum calculation is not recommended.

When receivers stop receiving data addressed to a certain IPv6 multicast group through the RP (that is, the RP stops serving the receivers of that IPv6 multicast group), or when the RP formally starts receiving IPv6 multicast data from the IPv6 multicast source, the RP sends a register-stop message to the source-side DR. Upon receiving this message, the DR stops sending register messages encapsulated with IPv6 multicast data and starts a register-stop timer. When the register-stop timer expires, the DR sends a null register message (a register message without encapsulated multicast data) to the RP. If the DR receives a register-stop message during the register probe time, it will reset its register-stop timer; otherwise, the DR starts sending register messages with encapsulated data again when the register-stop timer expires.

The Register-Stop Timer is set to a random value chosen uniformly from the interval (0.5 times register_suppression_time, 1.5 times register_suppression_time) minus register_probe_time.

Configure a filtering rule for register messages on all C-RP routers and configure them to calculate the checksum based on the entire register messages. Configure the register suppression time and the register probe time on all routers that may become source-side DRs.

Follow these steps to configure register-related parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure a filtering rule for register messages	**register-policy** *acl6-number*	Optional No register filtering rule by default
Configure the device to calculate the checksum based on the entire register messages	**register-whole-checksum**	Optional Based on the header of register messages by default
Configure the register suppression time	**register-suppression-timeout** *interval*	Optional 60 seconds by default
Configure the register probe time	**probe-interval** *interval*	Optional 5 seconds by default

Disabling SPT Switchover

If an 3Com Switch 4800G acts as an RP or the receiver-side DR, it initiates an SPT switchover process (by default) upon receiving the first IPv6 multicast packet along the RPT. You can disable the switchover from RPT to SPT.

Perform the following operations to disable the SPT switchover:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—

To do...	Use the command...	Remarks
Apply the ACL for initiating an SPT switchover	**spt-switch-threshold infinity** [**group-policy** *acl6-number* [**order** *order-value*]]	Optional By default, the device switches to the SPT immediately after it receives the first IPv6 multicast packet

Configuring IPv6 PIM-SSM

Note

The IPv6 PIM-SSM model needs the support of MLDv2. Therefore, be sure to enable MLDv2 on IPv6 PIM routers with receivers attached to them.

IPv6 PIM-SSM Configuration Task List

Complete these tasks to configure IPv6 PIM-SSM:

Task	Remarks
Enabling IPv6 PIM-SM	Required
Configuring the IPv6 SSM Group Range	Optional
Configuring IPv6 PIM Common Features	Optional

Configuration Prerequisites

Before configuring IPv6 PIM-SSM, complete the following task:

- Configure any IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring IPv6 PIM-SSM, prepare the following data:

- The IPv6 SSM group range

Enabling IPv6 PIM-SM

The SSM model is implemented based on some subsets of IPv6 PIM-SM. Therefore, a router is IPv6 PIM-SSM capable after you enable IPv6 PIM-SM on it.

When deploying an IPv6 PIM-SM domain, you are recommended to enable IPv6 PIM-SM on all non-border interfaces of routers.

Follow these steps to enable IPv6 PIM-SSM:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable IPv6 multicast routing	**multicast ipv6 routing-enable**	Required Disable by default
Enter interface view	**interface** *interface-type interface-number*	—
Enable IPv6 PIM-SM	**pim ipv6 sm**	Required Disabled by default

Caution

All the interfaces of the same device must work in the same IPv6 PIM mode.

Note

For details about the **multicast ipv6 routing-enable** command, see *IPv6 Multicast Routing and Forwarding Commands* in the *IP Multicast Volume*.

Configuring the IPv6 SSM Group Range

As for whether the information from an IPv6 multicast source is delivered to the receivers based on the IPv6 PIM-SSM model or the IPv6 PIM-SM model, this depends on whether the group address in the (S, G) channel subscribed by the receivers falls in the IPv6 SSM group range. All IPv6 PIM-SM-enabled interfaces assume that IPv6 multicast groups within this address range are using the IPv6 SSM model.

Perform the following configuration on all routers in the IPv6 PIM-SM domain.

Follow these steps to configure the IPv6 SSM group range:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the IPv6 SSM group range	**ssm-policy** *acl6-number*	Optional FF3x::/32 by default, here “x” refers to any legal group scope.

Caution

- Make sure that the same IPv6 SSM group range is configured on all routers in the entire domain. Otherwise, IPv6 multicast data cannot be delivered through the IPv6 SSM model.
- When a member of an IPv6 multicast group in the IPv6 SSM group range sends an MLDv1 report message, the device does not trigger a (*, G) join.

Configuring IPv6 PIM Common Features

Note

For the functions or parameters that can be configured in both IPv6 PIM view and interface view described in this section:

- Configurations performed in IPv6 PIM view are effective to all interfaces, while configurations performed in interface view are effective to the current interface only.
- If the same function or parameter is configured in both IPv6 PIM view and interface view, the configuration made in interface view has preference over the configuration made in PIM view, regardless of the configuration sequence.

IPv6 PIM Common Feature Configuration Task List

Complete these tasks to configure IPv6 PIM common features:

Task	Remarks
Configuring an IPv6 Multicast Data Filter	Optional
Configuring a Hello Message Filter	Optional
Configuring IPv6 PIM Hello Options	Optional
Configuring IPv6 PIM Common Timers	Optional
Configuring Join/Prune Message Sizes	Optional

Configuration Prerequisites

Before configuring IPv6 PIM common features, complete the following tasks:

- Configure any IPv6 unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure IPv6 PIM-DM (or IPv6 PIM-SM or IPv6 PIM-SSM).

Before configuring IPv6 PIM common features, prepare the following data:

- An IPv6 ACL rule for filtering IPv6 multicast data
- An IPv6 ACL rule defining a legal source address range for hello messages
- Priority for DR election (global value/interface level value)

- IPv6 PIM neighbor timeout time (global value/interface value)
- Prune delay (global value/interface level value)
- Prune override interval (global value/interface level value)
- Hello interval (global value/interface level value)
- Maximum delay between hello message (interface level value)
- Assert timeout time (global value/interface value)
- Join/prune interval (global value/interface level value)
- Join/prune timeout (global value/interface value)
- IPv6 multicast source lifetime
- Maximum size of join/prune messages
- Maximum number of (S, G) entries in a join/prune message

Configuring an IPv6 Multicast Data Filter

No matter in an IPv6 PIM-DM domain or an IPv6 PIM-SM domain, routers can check passing-by IPv6 multicast data based on the configured filtering rules and determine whether to continue forwarding the IPv6 multicast data. In other words, IPv6 PIM routers can act as IPv6 multicast data filters. These filters can help implement traffic control on one hand, and control the information available to downstream receivers to enhance data security on the other hand.

Follow these steps to configure an IPv6 multicast data filter:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure an IPv6 multicast group filter	**source-policy** *acl6-number*	Required No IPv6 multicast data filter by default

Note

- Generally, a smaller distance from the filter to the IPv6 multicast source results in a more remarkable filtering effect.
- This filter works not only on independent IPv6 multicast data but also on IPv6 multicast data encapsulated in register messages.

Configuring a Hello Message Filter

Along with the wide applications of IPv6 PIM, the security requirement for the protocol is becoming more and more demanding. The establishment of correct IPv6 PIM neighboring relationships is a prerequisite for secure application of IPv6 PIM. You can configure a legal source address range for hello messages on interfaces of routers to ensure the correct IPv6 PIM neighboring relationships and thus to guide against IPv6 PIM message attacks.

Follow these steps to configure a hello message filter:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure a hello message filter	**pim ipv6 neighbor-policy** *acl6-number*	Required No hello message filter by default.

Note

With the hello message filter configured, if hello messages of an existing IPv6 PIM neighbor fail to pass the filter, the IPv6 PIM neighbor will be removed automatically when it times out.

Configuring IPv6 PIM Hello Options

No matter in an IPv6 PIM-DM domain or an IPv6 PIM-SM domain, the hello messages sent among routers contain many configurable options, including:

- DR_Priority (for IPv6 PIM-SM only): priority for DR election. The higher the priority is, the easier it is for the router to win DR election. You can configure this parameter on all the routers in a multi-access network directly connected to IPv6 multicast sources or receivers.
- Holdtime: the timeout time of IPv6 PIM neighbor reachability state. When this timer times out, if the router has received no hello message from an IPv6 PIM neighbor, it assumes that this neighbor has expired or become unreachable.
- LAN_Prune_Delay: the delay of prune messages on a multi-access network. This option consists of Lan-delay (namely, prune delay), Override-interval, and neighbor tracking flag. If the LAN-delay or override-interval values of different IPv6 PIM routers on a multi-access subnet are different, the largest value will take effect. If you want to enable neighbor tracking, the neighbor tracking feature should be enabled on all IPv6 PIM routers on a multi-access subnet.

The LAN-delay setting will cause the upstream routers to delay processing received prune messages. If the LAN-delay setting is too small, it may cause the upstream router to stop forwarding IPv6 multicast packets before a downstream router sends a prune override message. Therefore, be cautious when configuring this parameter.

The override-interval sets the length of time a downstream router is allowed to wait before sending a prune override message. When a router receives a prune message from a downstream router, it does not perform the prune action immediately; instead, it maintains the current forwarding state for a period of LAN-delay plus override-interval. If the downstream router needs to continue receiving IPv6 multicast data, it must send a prune override message within the prune override interval; otherwise, the upstream route will perform the prune action when the period of LAN-delay plus override-interval time out.

A hello message sent from an IPv6 PIM router contains a generation ID option. The generation ID is a random value for the interface on which the hello message is sent. Normally, the generation ID of an IPv6 PIM router does not change unless the status of the router changes (for example, when IPv6 PIM is just enabled on the interface or the device is restarted). When the router starts or restarts sending hello messages, it generates a new generation ID. If an IPv6 PIM router finds that the generation ID in a hello message from the upstream router has changed, it assumes that the status of the upstream

neighbor is lost or the upstream neighbor has changed. In this case, it triggers a join message for state update.

If you disable join suppression (namely, enable neighbor tracking), the join suppression feature should be disabled on all IPv6 PIM routers on a multi-access subnet; otherwise, the upstream router will fail to explicitly track which downstream routers are joined to it.

Configuring hello options globally

Follow these steps to configure hello options globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the priority for DR election	**hello-option dr-priority** *priority*	Optional 1 by default
Configure IPv6 PIM neighbor timeout time	**hello-option holdtime** *interval*	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	**hello-option lan-delay** *interval*	Optional 500 milliseconds by default
Configure the prune override interval	**hello-option override-interval** *interval*	Optional 2,500 milliseconds by default
Disable join suppression	**hello-option neighbor-tracking**	Required Enabled by default

Configuring hello options on an interface

Follow these steps to configure hello options on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the priority for DR election	**pim ipv6 hello-option dr-priority** *priority*	Optional 1 by default
Configure IPv6 PIM neighbor timeout time	**pim ipv6 hello-option holdtime** *interval*	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	**pim ipv6 hello-option lan-delay** *interval*	Optional 500 milliseconds by default
Configure the prune override interval	**pim ipv6 hello-option override-interval** *interval*	Optional 2,500 milliseconds by default
Disable join suppression	**pim ipv6 hello-option neighbor-tracking**	Required Enabled by default

To do...	Use the command...	Remarks
Configure the interface to reject hello messages without a generation ID	**pim ipv6 require-genid**	Required By default, hello messages without Generation_ID are accepted.

Configuring IPv6 PIM Common Timers

IPv6 PIM routers discover IPv6 PIM neighbors and maintain IPv6 PIM neighboring relationships with other routers by periodically sending out hello messages.

Upon receiving a hello message, an IPv6 PIM router waits a random period, which is smaller than the maximum delay between hello messages, before sending out a hello message. This avoids collisions that occur when multiple IPv6 PIM routers send hello messages simultaneously.

An IPv6 PIM router periodically sends join/prune messages to its upstream for state update. A join/prune message contains the join/prune timeout time. The upstream router sets a join/prune timeout timer for each pruned downstream interface.

Any router that has lost assert election will prune its downstream interface and maintain the assert state for a period of time. When the assert state times out, the assert loser will resume IPv6 multicast forwarding.

When a router fails to receive subsequent IPv6 multicast data from the IPv6 multicast source S, the router does not immediately delete the corresponding (S, G) entry; instead, it maintains the (S, G) entry for a period of time, namely the IPv6 multicast source lifetime, before deleting the (S, G) entry.

Configuring IPv6 PIM common timers globally

Follow these steps to configure IPv6 PIM common timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the hello interval	**timer hello** *interval*	Optional 30 seconds by default
Configure the join/prune interval	**timer join-prune** *interval*	Optional 60 seconds by default
Configure the join/prune timeout time	**holdtime join-prune** *interval*	Optional 210 seconds by default
Configure assert timeout time	**holdtime assert** *interval*	Optional 180 seconds by default
Configure the IPv6 multicast source lifetime	**source-lifetime** *interval*	Optional 210 seconds by default

Configuring IPv6 PIM common timers on an interface

Follow these steps to configure IPv6 PIM common timers on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the hello interval	**pim ipv6 timer hello** *interval*	Optional 30 seconds by default
Configure the maximum delay between hello messages	**pim ipv6 triggered-hello-delay** *interval*	Optional 5 seconds by default
Configure the join/prune interval	**pim ipv6 timer join-prune** *interval*	Optional 60 seconds by default
Configure the join/prune timeout time	**pim ipv6 holdtime join-prune** *interval*	Optional 210 seconds by default
Configure assert timeout time	**pim ipv6 holdtime assert** *interval*	Optional 180 seconds by default

Note

If there are no special networking requirements, we recommend that you use the default settings.

Configuring Join/Prune Message Sizes

A larger join/prune message size will result in loss of a larger amount of information when a message is lost; with a reduced join/message size, the loss of a single message will bring relatively minor impact.

By controlling the maximum number of (S, G) entries in a join/prune message, you can effectively reduce the number of (S, G) entries sent per unit of time.

Follow these steps to configure join/prune message sizes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter IPv6 PIM view	**pim ipv6**	—
Configure the maximum size of a join/prune message	**jp-pkt-size** *packet-size*	Optional 8,100 bytes by default
Configure the maximum number of (S, G) entries in a join/prune message	**jp-queue-size** *queue-size*	Optional 1,020 by default

Displaying and Maintaining IPv6 PIM

To do...	Use the command...	Remarks
View the BSR information in the IPv6 PIM-SM domain and locally configured C-RP information in effect	**display pim ipv6 bsr-info**	Available in any view
View the information of IPv6 unicast routes used by IPv6 PIM	**display pim ipv6 claimed-route** [*ipv6-source-address*]	Available in any view
View the number of IPv6 PIM control messages	**display pim ipv6 control-message counters** [**message-type** { **probe** | **register** | **register-stop** } | [**interface** *interface-type interface-number* | **message-type** { **assert** | **bsr** | **crp** | **graft** | **graft-ack** | **hello** | **join-prune** | **state-refresh** }] *]	Available in any view
View the information about unacknowledged graft messages	**display pim ipv6 grafts**	Available in any view
View the IPv6 PIM information on an interface or all interfaces	**display pim ipv6 interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View the information of join/prune messages to send	**display pim ipv6 join-prune mode** { **sm** [**flags** *flag-value*] | **ssm** } [**interface** *interface-type interface-number* | **neighbor** *ipv6-neighbor-address*] * [**verbose**]	Available in any view
View IPv6 PIM neighboring information	**display pim ipv6 neighbor** [**interface** *interface-type interface-number* | *ipv6-neighbor-address* | **verbose**] *	Available in any view
View the content of the IPv6 PIM routing table	**display pim ipv6 routing-table** [*ipv6-group-address* [*prefix-length*] | *ipv6-source-address* [*prefix-length*] | **incoming-interface** [*interface-type interface-number* | **register**] | **outgoing-interface** { **include** | **exclude** | **match** } { *interface-type interface-number* | **register** } | **mode** *mode-type* | **flags** *flag-value* | **fsm**] *	Available in any view
View the RP information	**display pim ipv6 rp-info** [*ipv6-group-address*]	Available in any view
Reset IPv6 PIM control message counters	**reset pim ipv6 control-message counters** [**interface** *interface-type interface-number*]	Available in user view

IPv6 PIM Configuration Examples

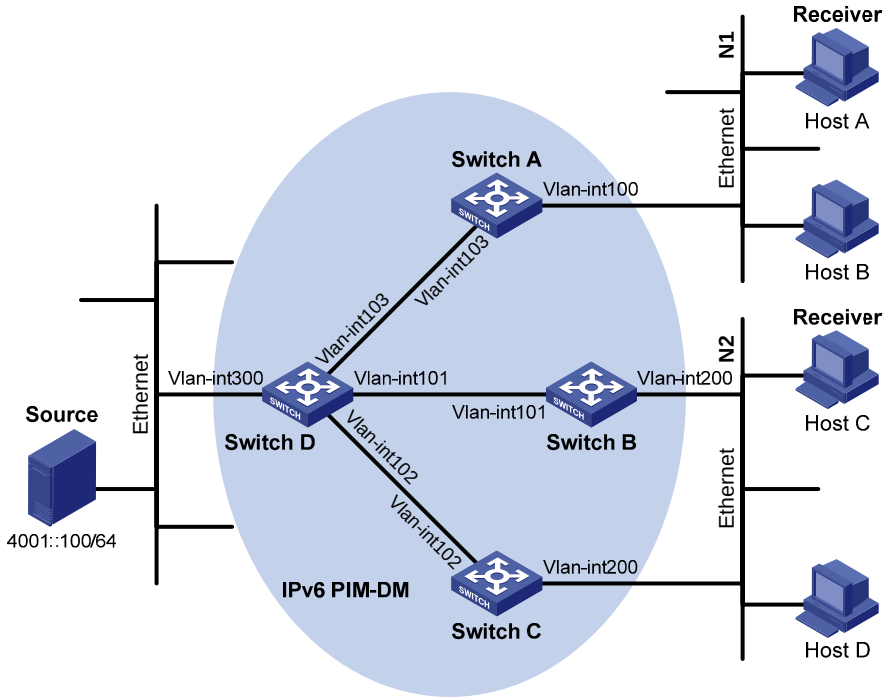
IPv6 PIM-DM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire IPv6 PIM domain operates in the dense mode.
- Host A and Host C are multicast receivers in two stub networks N1 and N2.

- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to N1 through VLAN-interface 100, and to Switch D through VLAN-interface 103.
- Switch B and Switch C connect to N2 through their respective VLAN-interface 200, and to Switch D through VLAN-interface 101 and VLAN-interface 102 respectively.
- MLDv1 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Figure 1-8 Network diagram for IPv6 PIM-DM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
	Vlan-int103	1002::1/64		Vlan-int103	1002::2/64
Switch B	Vlan-int200	2001::1/64		Vlan-int101	2002::2/64
	Vlan-int101	2002::1/64		Vlan-int102	3001::2/64
Switch C	Vlan-int200	2001::2/64			
	Vlan-int102	3001::1/64			

Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses and IPv6 unicast routing

Enable IPv6 forwarding on each switch and configure the IPv6 address and prefix length for each interface as per Figure 1-8. Detailed configuration steps are omitted here.

Configure OSPFv3 for interoperation among the switches in the PIM-DM domain. Ensure the network-layer interoperation in the PIM-DM domain and enable dynamic update of routing information among the switches through an IPv6 unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IPv6 multicast routing, and enable IPv6 PIM-DM and MLD

Enable IPv6 multicast routing on Switch A, enable IPv6 PIM-DM on each interface, and enable MLD on VLAN-interface 100, which connects Switch A to N1.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
```

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] pim ipv6 dm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim ipv6 dm
[SwitchA-Vlan-interface103] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A.

Enable IPv6 multicast routing on Switch D, and enable IPv6 PIM-DM on each interface.

```
<SwitchD> system-view
[SwitchD] multicast ipv6 routing-enable
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] pim ipv6 dm
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim ipv6 dm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 101
[SwitchD-Vlan-interface101] pim ipv6 dm
[SwitchD-Vlan-interface101] quit
[SwitchD] interface vlan-interface 102
[SwitchD-Vlan-interface102] pim ipv6 dm
[SwitchD-Vlan-interface102] quit
```

3) Verify the configuration

Use the **display pim ipv6 interface** command to view the IPv6 PIM configuration and running status on each interface. For example:

View the IPv6 PIM configuration information on Switch D.

```
[SwitchD] display pim ipv6 interface
 Interface          NbrCnt HelloInt   DR-Pri     DR-Address
 Vlan300            0      30         1          4001::1
                                                 (local)
 Vlan103            0      30         1          1002::2
                                                 (local)
 Vlan101            1      30         1          2002::2
                                                 (local)
 Vlan102            1      30         1          3001::2
                                                 (local)
```

Use the **display pim ipv6 neighbor** command to view the IPv6 PIM neighboring relationships among the switches. For example:

View the IPv6 PIM neighboring relationships on Switch D.

```
[SwitchD] display pim ipv6 neighbor
 Total Number of Neighbors = 3

 Neighbor           Interface           Uptime   Expires  Dr-Priority
 1002::1            Vlan103             00:04:00 00:01:29 1
 2002::1            Vlan101             00:04:16 00:01:29 3
```

```
 3001::1            Vlan102             00:03:54 00:01:17 5
```

Assume that Host A needs to receive the information addressed to IPv6 multicast group G (FF0E::101). After IPv6 multicast source S (4001::100/64) sends IPv6 multicast packets to the IPv6 multicast group G, an SPT is established through traffic flooding. Switches on the SPT path (Switch A and Switch D) have their (S, G) entries. Host A sends an MLD report to Switch A to join IPv6 multicast group G, and a (*, G) entry is generated on Switch A. You can use the **display pim IPv6 routing-table** command to view the IPv6 PIM routing table information on each switch. For example:

View the IPv6 PIM multicast routing table information on Switch A.

```
[SwitchA] display pim ipv6 routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::101)
     Protocol: pim-dm, Flag: WC
     UpTime: 00:01:24
     Upstream interface: NULL
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: mld, UpTime: 00:01:20, Expires: never

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:01:20
     Upstream interface: Vlan-interface103
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: pim-dm, UpTime: 00:01:20, Expires: never
```

The information on Switch B and Switch C is similar to that on Switch A.

View the IPv6 PIM multicast routing table information on Switch D.

```
[SwitchD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF0E::101)
     Protocol: pim-dm, Flag: LOC ACT
     UpTime: 00:02:19
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 3
         1: Vlan-interface103
```

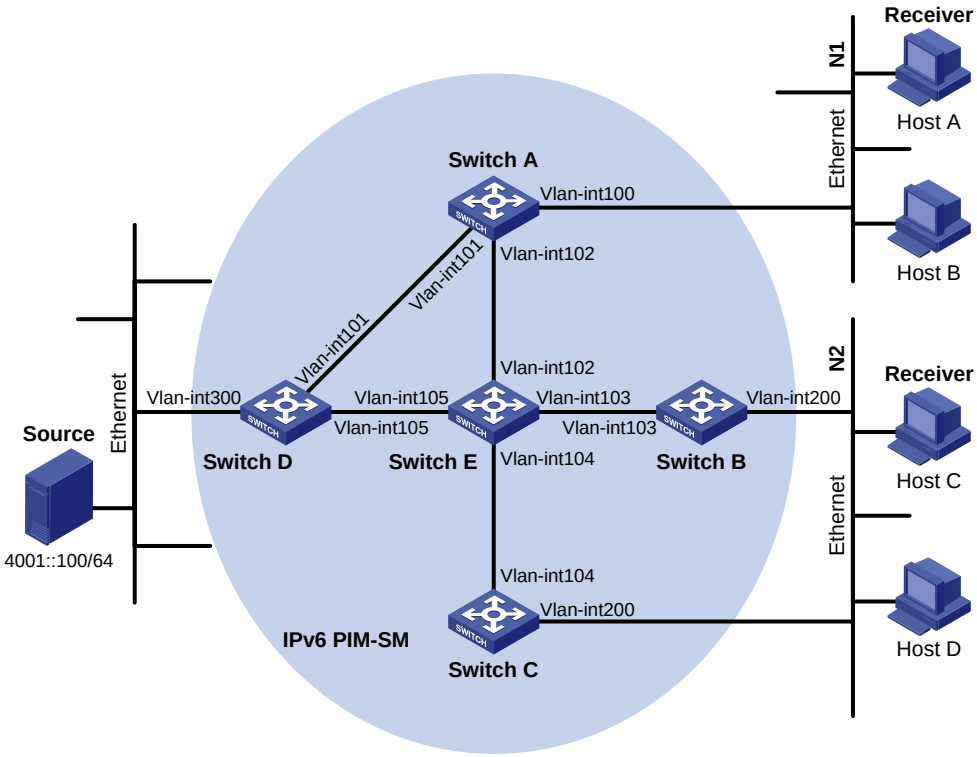
```
    Protocol: pim-dm, UpTime: 00:02:19, Expires: never
2: Vlan-interface101
    Protocol: pim-dm, UpTime: 00:02:19, Expires: never
3: Vlan-interface102
    Protocol: pim-dm, UpTime: 00:02:19, Expires: never
```

IPv6 PIM-SM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the sparse mode.
- Host A and Host C are IPv6 multicast receivers in two stub networks N1 and N2.
- Switch D connects to the network that comprises the IPv6 multicast source (Source) through VLAN-interface 300.
- Switch A connects to N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Vlan-interface 105 on Switch D and Vlan-interface 102 on Switch E act as C-BSRs and C-RPs; the C-BSR on Switch E has a higher priority; the IPv6 multicast group range served by the C-RP is FF0E::101/64; modify the hash mask length to map a certain number of consecutive IPv6 group addresses within the range to the two C-RPs.
- MLDv1 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Figure 1-9 Network diagram for IPv6 PIM-SM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
	Vlan-int101	1002::1/64		Vlan-int101	1002::2/64
	Vlan-int102	1003::1/64		Vlan-int105	4002::1/64
Switch B	Vlan-int200	2001::1/64	Switch E	Vlan-int104	3001::2/64
	Vlan-int103	2002::1/64		Vlan-int103	2002::2/64
Switch C	Vlan-int200	2001::2/64		Vlan-int102	1003::2/64
	Vlan-int104	3001::1/64		Vlan-int105	4002::2/64

Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses and IPv6 unicast routing

Enable IPv6 forwarding on each switch and configure the IPv6 address and prefix length for each interface as per Figure 1-9. Detailed configuration steps are omitted here.

Configure OSPFv3 for interoperation among the switches in the IPv6 PIM-SM domain. Ensure the network-layer interoperation in the IPv6 PIM-DM domain and enable dynamic update of routing information among the switches through an IPv6 unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IPv6 multicast routing, and enable IPv6 PIM-SM and MLD

Enable IPv6 multicast routing on Switch A, enable IPv6 PIM-SM on each interface, and enable MLD on VLAN-interface 100, which connects Switch A to N1.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] pim ipv6 sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
```

```
[SwitchA-Vlan-interface101] pim ipv6 sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim ipv6 sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable MLD on the corresponding interfaces on these two switches.

3) Configure a C-BSR and a C-RP

On Switch D, configure the service scope of RP advertisements, specify a C-BSR and a C-RP, and set the hash mask length to 128 and the priority of the C-BSR to 10.

```
<SwitchD> system-view
[SwitchD] acl ipv6 number 2005
[SwitchD-acl6-basic-2005] rule permit source ff0e::101 64
[SwitchD-acl6-basic-2005] quit
[SwitchD] pim ipv6
[SwitchD-pim6] c-bsr 4002::1 128 10
[SwitchD-pim6] c-rp 4002::1 group-policy 2005
[SwitchD-pim6] quit
```

On Switch E, configure the service scope of RP advertisements, specify a C-BSR and a C-RP, and set the hash mask length to 128 and the priority of the C-BSR to 20.

```
<SwitchE> system-view
[SwitchE] acl ipv6 number 2005
[SwitchE-acl6-basic-2005] rule permit source ff0e::101 64
[SwitchE-acl6-basic-2005] quit
[SwitchE] pim ipv6
[SwitchE-pim6] c-bsr 1003::2 128 20
[SwitchE-pim6] c-rp 1003::2 group-policy 2005
[SwitchE-pim6] quit
```

4) Verify the configuration

Use the **display pim ipv6 interface** command to view the IPv6 PIM configuration and running status on each interface. For example:

View the IPv6 PIM information on all interfaces of Switch A.

```
[SwitchA] display pim ipv6 interface
 Interface            NbrCnt HelloInt   DR-Pri    DR-Address
 Vlan100              0      30         1         1001::1
                                                  (local)
 Vlan101              1      30         1         1002::2
 Vlan102              1      30         1         1003::2
```

To view the BSR election information and the locally configured C-RP information in effect on a switch, use the **display pim ipv6 bsr-info** command. For example:

View the BSR information and the locally configured C-RP information in effect on Switch A.

```
[SwitchA] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
```

```
     Hash mask length: 128
     State: Accept Preferred
     Uptime: 00:04:22
     Expires: 00:01:46
```

View the BSR information and the locally configured C-RP information in effect on Switch D.

```
[SwitchD] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Elected
     Uptime: 00:05:26
     Expires: 00:01:45
 Candidate BSR Address: 4002::1
     Priority: 10
     Hash mask length: 128
     State: Candidate

 Candidate RP: 4002::1(Vlan-interface105)
     Priority: 0
     HoldTime: 130
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:48
```

View the BSR information and the locally configured C-RP information in effect on Switch E.

```
[SwitchE] display pim ipv6 bsr-info
 Elected BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Elected
     Uptime: 00:01:10
     Next BSR message scheduled at: 00:01:48
 Candidate BSR Address: 1003::2
     Priority: 20
     Hash mask length: 128
     State: Elected
Candidate RP: 1003::2(Vlan-interface102)
     Priority: 0
     HoldTime: 130
     Advertisement Interval: 60
     Next advertisement scheduled at: 00:00:48
```

To view the RP information discovered on a switch, use the **display pim ipv6 rp-info** command. For example:

View the RP information on Switch A.

```
[SwitchA] display pim ipv6 rp-info
 PIM-SM BSR RP information:
 prefix/prefix length: FF0E::101/64
     RP: 4002::1
```

```
    Priority: 0
    HoldTime: 130
    Uptime: 00:05:19
    Expires: 00:02:11

    RP: 1003::2
    Priority: 0
    HoldTime: 130
    Uptime: 00:05:19
    Expires: 00:02:11
```

Assume that Host A needs to receive information addressed to the IPv6 multicast group G (FF0E::100). The RP corresponding to the multicast group G is Switch E as a result of hash calculation, so an RPT will be built between Switch A and Switch E. When the IPv6 multicast source S (4001::100/64) registers with the RP, an SPT will be built between Switch D and Switch E. Upon receiving IPv6 multicast data, Switch A immediately switches from the RPT to the SPT. Switches on the RPT path (Switch A and Switch E) have a (*, G) entry, while switches on the SPT path (Switch A and Switch D) have an (S, G) entry. You can use the **display pim ipv6 routing-table** command to view the PIM routing table information on the switches. For example:

View the IPv6 PIM multicast routing table information on Switch A.

```
[SwitchA] display pim ipv6 routing-table
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::100)
     RP: 1003::2
     Protocol: pim-sm, Flag: WC
     UpTime: 00:03:45
     Upstream interface: Vlan-interface102
         Upstream neighbor: 1003::2
         RPF prime neighbor: 1003::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: mld, UpTime: 00:02:15, Expires: 00:03:06

 (4001::100, FF0E::100)
     RP: 1003::2
     Protocol: pim-sm, Flag: SPT ACT
     UpTime: 00:02:15
     Upstream interface: Vlan-interface101
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface100
             Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:06
```

The information on Switch B and Switch C is similar to that on Switch A.

View the IPv6 PIM multicast routing table information on Switch D.

```
[SwitchD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF0E::100)
     RP: 1003::2
     Protocol: pim-sm, Flag: SPT LOC ACT
     UpTime: 00:14:44
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface105
             Protocol: mld, UpTime: 00:14:44, Expires: 00:02:26
```

View the IPv6 PIM multicast routing table information on Switch E.

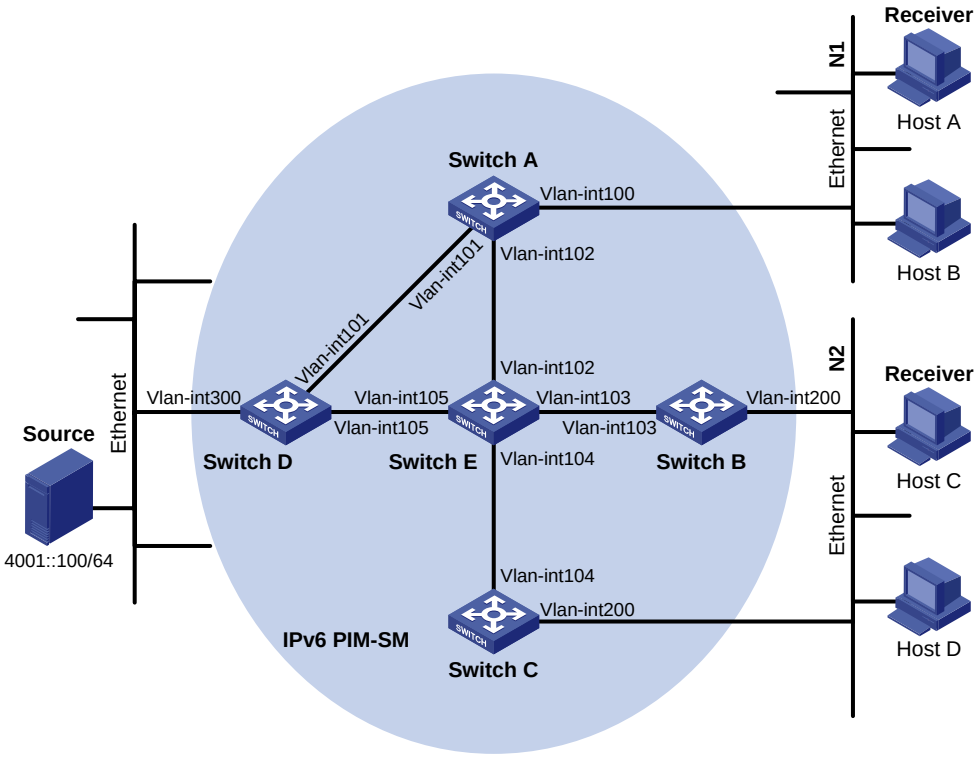
```
[SwitchE] display pim ipv6 routing-table
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, FF0E::100)
     RP: 1003::2 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:16:56
     Upstream interface: Register
         Upstream neighbor: 4002::1
         RPF prime neighbor: 4002::1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface102
             Protocol: pim-sm, UpTime: 00:16:56, Expires: 00:02:34
```

IPv6 PIM-SSM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the SSM mode.
- Host A and Host C are IPv6 multicast receivers in two stub networks N1 and N2.
- Switch D connects to the network that comprises the IPv6 multicast source (Source) through VLAN-interface 300.
- Switch A connects to N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Switch E connects to Switch A, Switch B, Switch C and Switch D.
- The SSM group range is FF3E::/64.
- MLDv2 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Figure 1-10 Network diagram for IPv6 PIM-SSM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	1001::1/64	Switch D	Vlan-int300	4001::1/64
	Vlan-int101	1002::1/64		Vlan-int101	1002::2/64
	Vlan-int102	1003::1/64		Vlan-int105	4002::1/64
Switch B	Vlan-int200	2001::1/64	Switch E	Vlan-int104	3001::2/64
	Vlan-int103	2002::1/64		Vlan-int103	2002::2/64
Switch C	Vlan-int200	2001::2/64		Vlan-int102	1003::2/64
	Vlan-int104	3001::1/64		Vlan-int105	4002::2/64

Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses and IPv6 unicast routing

Enable IPv6 forwarding on each switch and configure the IPv6 address and prefix length for each interface as per Figure 1-10. Detailed configuration steps are omitted here.

Configure OSPFv3 for interoperation among the switches in the IPv6 PIM-SM domain. Ensure the network-layer interoperation in the IPv6 PIM-SM domain and enable dynamic update of routing information among the switches through an IPv6 unicast routing protocol. Detailed configuration steps are omitted here.

2) Enable IPv6 multicast routing, and enable IPv6 PIM-SM and MLD

Enable IPv6 multicast routing on Switch A, enable IPv6 PIM-SM on each interface, and run MLDv2 on VLAN-interface 100, which connects Switch A to N1.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] mld version 2
[SwitchA-Vlan-interface100] pim ipv6 sm
[SwitchA-Vlan-interface100] quit
```

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim ipv6 sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable MLD on the corresponding interfaces on these two switches.

3) Configure the IPv6 SSM group range

Configure the IPv6 SSM group range to be FF3E::/64 on Switch A.

```
[SwitchA] acl ipv6 number 2000
[SwitchA-acl6-basic-2000] rule permit source ff3e:: 64
[SwitchA-acl6-basic-2000] quit
[SwitchA] pim ipv6
[SwitchA-pim6] ssm-policy 2000
[SwitchA-pim6] quit
```

The configuration on Switch B, Switch C, Switch D, and Switch E is similar to that on Switch A.

4) Verify the configuration

Use the **display pim ipv6 interface** command to view the IPv6 PIM configuration and running status on each interface. For example:

View the IPv6 PIM configuration information on Switch A.

```
[SwitchA] display pim ipv6 interface
 Interface             NbrCnt HelloInt   DR-Pri   DR-Address
 Vlan100               0      30         1        1001::1
                                                  (local)
 Vlan101               1      30         1        1002::2
 Vlan102               1      30         1        1003::2
```

Assume that Host A needs to receive the information a specific IPv6 multicast source S (4001::100/64) sends to IPv6 multicast group G (FF3E::101). Switch A builds an SPT toward the IPv6 multicast source. Switches on the SPT path (Switch A and Switch D) have generated an (S, G) entry, while Switch E, which is not on the SPT path, does not have IPv6 multicast routing entries. You can use the **display pim ipv6 routing-table** command to view the IPv6 PIM routing table information on each switch. For example:

View the IPv6 PIM multicast routing table information on Switch A.

```
[SwitchA] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF3E::101)
     Protocol: pim-ssm, Flag:
     UpTime: 00:00:11
     Upstream interface: Vlan-interface101
         Upstream neighbor: 1002::2
         RPF prime neighbor: 1002::2
     Downstream interface(s) information:
```

```
    Total number of downstreams: 1
        1: Vlan-interface100
            Protocol: mld, UpTime: 00:00:11, Expires: 00:03:25
```

The information on Switch B and Switch C is similar to that on Switch A.

View the IPv6 PIM multicast routing table information on Switch B.

```
[SwitchD] display pim ipv6 routing-table
 Total 0 (*, G) entry; 1 (S, G) entry

 (4001::100, FF3E::101)
     Protocol: pim-ssm, Flag: LOC
     UpTime: 00:08:02
     Upstream interface: Vlan-interface300
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface105
             Protocol: pim-ssm, UpTime: 00:08:02, Expires: 00:03:25
```

Troubleshooting IPv6 PIM Configuration

Failure of Building a Multicast Distribution Tree Correctly

Symptom

None of the routers in the network (including routers directly connected with IPv6 multicast sources and receivers) has IPv6 multicast forwarding entries. That is, a multicast distribution tree cannot be built correctly and clients cannot receive IPv6 multicast data.

Analysis

- An IPv6 PIM routing entry is created based on an IPv6 unicast route, whichever IPv6 PIM mode is running. Multicast works only when unicast does.
- IPv6 PIM must be enabled on the RPF interface. An RPF neighbor must be an IPv6 PIM neighbor as well. If IPv6 PIM is not enabled on the RPF interface or the RPF neighbor, the establishment of a multicast distribution tree will surely fail, resulting in abnormal multicast forwarding.
- IPv6 PIM requires that the same IPv6 PIM mode, namely DM or SM, must run on the entire network. Otherwise, the establishment of a multicast distribution tree will surely fail, resulting in abnormal multicast forwarding.

Solution

1) Check IPv6 unicast routes. Use the **display ipv6 routing-table** command to check whether a unicast route exists to the IPv6 multicast source or the RP.
2) Check that the RPF interface is IPv6 PIM enabled. Use the **display pim ipv6 interface** command to view the IPv6 PIM information on each interface. If IPv6 PIM is not enabled on the interface, use the **pim ipv6 dm** or **pim ipv6 sm** command to enable IPv6 PIM.
3) Check that the RPF neighbor is an IPv6 PIM neighbor. Use the **display pim ipv6 neighbor** command to view the PIM neighbor information.

4) Check that IPv6 PIM and MLD are enabled on the interfaces directly connecting to the IPv6 multicast source and to the receiver.
5) Check that the same IPv6 PIM mode is enabled on related interfaces. Use the **display pim ipv6 interface verbose** command to check whether the same PIM mode is enabled on the RPF interface and the corresponding interface of the RPF neighbor router.
6) Check that the same IPv6 PIM mode is enabled on all the routers in the entire network. Use the **display current-configuration** command to check the IPv6 PIM mode information on each interface. Make sure that the same IPv6 PIM mode is enabled on all the routers: IPv6 PIM-SM on all routers, or IPv6 PIM-DM on all routers.

IPv6 Multicast Data Abnormally Terminated on an Intermediate Router

Symptom

An intermediate router can receive IPv6 multicast data successfully, but the data cannot reach the last hop router. An interface on the intermediate router receives data but no corresponding (S, G) entry is created in the IPv6 PIM routing table.

Analysis

- When a router receives an IPv6 multicast packet, it decrements the hop limit value of the IPv6 multicast packet by 1 and recalculates the checksum value. The router then forwards the packet to all outgoing interfaces. If the **multicast ipv6 minimum-hoplimit** command is configured on the outgoing interfaces, the hop limit value of the packet must be larger than the configured minimum hop limit value; otherwise, the packet will be discarded.
- If an IPv6 multicast forwarding boundary has been configured through the **multicast ipv6 boundary** command, any IPv6 multicast packet will be kept from crossing the boundary, and therefore no routing entry can be created in the IPv6 PIM routing table.
- In addition, the **source-policy** command is used to filter received IPv6 multicast packets. If the IPv6 multicast data fails to pass the ACL rule defined in this command, IPv6 PIM cannot create the route entry, either.

Solution

1) Check the minimum hop limit value for IPv6 multicast forwarding. Use the **display current-configuration** command to check the minimum hop limit value for multicast forwarding. Increase the hop limit value or remove the **multicast ipv6 minimum-hoplimit** command configured on the interface.
2) Check the IPv6 multicast forwarding boundary configuration. Use the **display current-configuration** command to check the IPv6 multicast forwarding boundary settings. Use the **multicast ipv6 boundary** command to change the IPv6 multicast forwarding boundary settings.
3) Check the IPv6 multicast filter configuration. Use the **display current-configuration** command to check the IPv6 multicast filter configuration. Change the IPv6 ACL rule defined in the **source-policy** command so that the source/group address of the IPv6 multicast data can pass ACL filtering.

RPs Unable to Join SPT in IPv6 PIM-SM

Symptom

An RPT cannot be established correctly, or the RPs cannot join the SPT to the IPv6 multicast source.

Analysis

- As the core of an IPv6 PIM-SM domain, the RPs serves specific IPv6 multicast groups. Multiple RPs can coexist in a network. Make sure that the RP information on all routers is exactly the same, and a specific group is mapped to the same RP. Otherwise, IPv6 multicast will fail.
- In the case of the static RP mechanism, the same RP address must be configured on all the routers in the entire network, including static RPs, by means of the static RP command. Otherwise, IPv6 multicast will fail.

Solution

1) Check that a route is available to the RP. Carry out the **display ipv6 routing-table** command to check whether a route is available on each router to the RP.
2) Check the dynamic RP information. Use the **display pim ipv6 rp-info** command to check whether the RP information is consistent on all routers. In the case of inconsistent RP information, configure consistent RP address on all the routers.
3) Check the static RP configuration. Carry out the **display pim ipv6 rp-info** command to check whether the same RP address has been configured on all the routers throughout the network.

RPT Establishment Failure or Source Registration Failure in IPv6 PIM-SM

Symptom

C-RPs cannot unicast advertise messages to the BSR. The BSR does not advertise bootstrap messages containing C-RP information and has no unicast route to any C-RP. An RPT cannot be established correctly, or the DR cannot perform source register with the RP.

Analysis

- C-RPs periodically send advertisement messages to the BSR by unicast. If a C-RP does not have a route to the BSR, the BSR will be unable to receive the advertisements from the C-RP, and therefore the bootstrap messages of the BSR will not contain the information about that C-RP.
- The RP is the core of an IPv6 PIM-SM domain. Make sure that the RP information on all routers is exactly the same, a specific group is mapped to the same RP, and a unicast route is available to the RP.

Solution

1) Check whether routes to C-RPs, the RP and the BSR are available. Carry out the **display ipv6 routing-table** command to check whether routes are available on each router to the RP and the BSR, and whether a route is available between the RP and the BSR. Make sure that each C-RP has a unicast route to the BSR, the BSR has a unicast route to each C-RP, and all the routers in the entire network have a unicast route to the RP.
2) Check the RP and BSR information. IPv6 PIM-SM needs the support of the RP and BSR. Use the **display pim ipv6 bsr-info** command to check whether the BSR information is available on each router, and then use the **display pim ipv6 rp-info** command to check whether the RP information is correct.
3) View the IPv6 PIM neighboring relationships. Use the **display pim ipv6 neighbor** command to check whether the normal neighboring relationships have been established among the routers.

Table of Contents

1 IPv6 MBGP Configuration

When configuring IPv6 MBGP, go to these sections for information you are interested in:

Note

This chapter describes only configuration for IPv6 MBGP. For IPv6 BGP related information, refer to *IPv6 BGP Configuration* in the *IP Routing Volume*.

IPv6 MBGP Overview

BGP-4 is capable of carrying routing information for IPv4 only. IETF defined multi-protocol BGP extensions to carry routing information for multiple network layer protocols.

For an IPv6 network, the IPv6 multicast topology need be different from the IPv6 unicast topology. To meet the requirement, the multi-protocol BGP extensions enable IPv6 BGP to carry the IPv6 unicast Network Layer Reachability Information (NLRI) and IPv6 multicast NLRI separately, and the multicast NLRI is used to perform reverse path forwarding (RPF) exclusively. In this way, route selection for a destination through the IPv6 unicast routing table and through the IPv6 multicast routing table will have different results, ensuring the normal unicast and multicast operation across ASs.

Multi-protocol BGP is defined in RFC 2858 (Multiprotocol Extensions for BGP-4).

Multi-protocol BGP for IPv6 multicast is referred to as IPv6 multicast BGP (IPv6 MBGP).

Note

This document covers configuration tasks related to multi-protocol BGP for IPv6 multicast only. For BGP related information, refer to *BGP Configuration* in the *IP Routing Volume*.
For information about RPF, refer to *Multicast Routing and Forwarding* in the *IP Multicast Volume*.

IPv6 MBGP Configuration Task List

Complete the following tasks to configure IPv6 MBGP:

Task		Remarks
Configuring IPv6 MBGP Basic Functions	Configuring an IPv6 MBGP Peer	Required
	Configuring a Preferred Value for Routes from a Peer/Peer Group	Optional
Controlling Route Distribution and Reception	Injecting a Local IPv6 MBGP Route	Optional
	Configuring IPv6 MBGP Route Redistribution	Optional
	Configuring IPv6 MBGP Route Summarization	Optional
	Advertising a Default Route to a Peer or Peer Group	Optional
	Configure Outbound IPv6 MBGP Route Filtering	Optional
	Configuring Inbound IPv6 MBGP Route Filtering	Optional
	Configuring IPv6 MBGP Route Dampening	Optional
Configuring IPv6 MBGP Route Attributes	Configuring IPv6 MBGP Route Preferences	Optional
	Configuring the Default Local Preference	
	Configuring the MED Attribute	
	Configuring the NEXT_HOP Attribute	Optional
	Configuring the AS_PATH Attribute	Optional
Tuning and Optimizing IPv6 MBGP Networks	Configuring IPv6 MBGP Soft Reset	Optional
	Configuring the Maximum Number of Equal-Cost Routes for Load-Balancing	Optional
Configuring a Large Scale IPv6 MBGP Network	Configuring an IPv6 MBGP Peer Group	Optional
	Configuring IPv6 MBGP Community	Optional
	Configuring an IPv6 MBGP Route Reflector	Optional

Configuring IPv6 MBGP Basic Functions

Configuration Prerequisites

IPv6 MBGP is an application of multi-protocol BGP. Therefore, before configuring IPv6 MBGP, you need to

- Enable IPv6
- Configure network layer addresses for interfaces
- Complete BGP basic configuration

Configuring an IPv6 MBGP Peer

Follow these steps to configure an IPv6 MBGP peer

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable BGP and enter BGP view	**bgp** *as-number*	Required Not enabled by default

To do…	Use the command…	Remarks
Enter IPv6 address family view	**ipv6-family**	—
Specify a IPv6 BGP peer and its AS number	**peer** *ipv6-address* **as-number** *as-number*	Required Not configured by default
Quit to BGP view	**quit**	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Enable the IPv6 MBGP peer	**peer** *ipv6-address* **enable**	Required Not enabled by default.

Configuring a Preferred Value for Routes from a Peer/Peer Group

Follow these steps to configure a preferred value for routes from a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Specify a preferred value for routes received from the IPv6 MBGP peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value*	Optional The preferred value defaults to 0.

Caution

If you both reference a route policy and use the command **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* to set a preferred value for routes from a peer/peer group, the route policy sets a non-zero preferred value for routes matching it. Other routes not matching the route policy uses the value set with the command. If the preferred value in the route policy is zero, the routes matching it will also use the value set with the command. For information about using a route policy to set a preferred value, refer to the **peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* { **import** | **export** } command and the **apply preferred-value** *preferred-value* command in *Route Policy Commands* in the *IP Routing Volume*.

Controlling Route Distribution and Reception

Configuration Prerequisites

Before configuring this task, you need to:

- Enable IPv6.
- Configure the IPv6 MBGP basic functions.

Injecting a Local IPv6 MBGP Route

Follow these steps to inject a local IPv6 MBGP route:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Inject a network to the IPv6 MBGP routing table	**network** *ipv6-address prefix-length* [**route-policy** *route-policy-name* | **short-cut**]	Required Not injected by default

Configuring IPv6 MBGP Route Redistribution

Follow these steps to configure IPv6 MBGP route redistribution:

To do…	Use the command…	Description
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter the MBGP multicast address family view	**ipv6-family multicast**	—
Enable default route redistribution into the IPv6 MBGP routing table	**default-route imported**	Optional By default, default route redistribution is not allowed.
Enable route redistribution from another routing protocol	**import-route** *protocol* [*process-id*] [**med** *med-value* | **route-policy** *route-policy-name*] *	Required Not enabled by default

Note

If the **default-route imported** command is not configured, using the **import-route** command cannot redistribute any IGP default route.

Configuring IPv6 MBGP Route Summarization

To reduce the routing table size on medium and large BGP networks, you need to configure route summarization on IPv6 MBGP routers. BGP supports only manual summarization of IPv6 multicast routes.

Follow these steps to configure IPv6 MBGP route summarization:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure manual route summarization	**aggregate** *ipv6-address prefix-length* [**as-set** | **attribute-policy** *route-policy-name* | **detail-suppressed** | **origin-policy** *route-policy-name* | **suppress-policy** *route-policy-name*]*	Required Not configured by default.

Advertising a Default Route to a Peer or Peer Group

Follow these steps to advertise a default route to a peer or peer group

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Advertise a default route to an IPv6 MBGP peer or peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	Required Not advertised by default

Note

With the **peer default-route-advertise** command executed, the router sends a default route with the next hop being itself to the specified IPv6 MBGP peer/peer group, regardless of whether the default route is available in the routing table.

Configure Outbound IPv6 MBGP Route Filtering

Follow these steps to configure outbound IPv6 MBGP route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—

To do…	Use the command…	Remarks
Configure the filtering of outgoing routes	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **export** [*protocol process-id*]	Use any of the commands. No filtering is configured by default. You can configure filter policies as needed. If you configure multiple filter policies, they will be applied in the following order: • **filter-policy export** • **peer filter-policy export** • **peer as-path-acl export** • **peer ipv6-prefix export** • **peer route-policy export** A filter policy can be applied only after the previous one is passed; routing information can be advertised only after passing all the filter policies configured.
Specify an IPv6 ACL to filer routes advertised to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **export**	
Specify an AS path ACL to filter IPv6 MBGP routing information advertised to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **export**	
Specify an IPv6 prefix list to filer routes advertised to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **export**	
Apply a route policy to routes advertised to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	

Note

- Members of an IPv6 MBGP peer group must have the same outbound route filtering policy as the peer group.
- IPv6 BGP advertises redistributed routes passing the specified policy to the IPv6 MBGP peer.

Configuring Inbound IPv6 MBGP Route Filtering

Follow these steps to configure IPv6 MBGP inbound route filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—

To do…	Use the command…	Remarks
Configure inbound route filtering	**filter-policy** { *acl6-number* | **ipv6-prefix** *ipv6-prefix-name* } **import**	Use any of the commands By default, advertised routes are not filtered. You can configure a filtering policy as needed. If several filtering policies are configured, they are applied in the following sequence: • **filter-policy import** • **peer filter-policy import** • **peer as-path-acl import** • **peer ip-prefix import** • **peer route-policy import** A filter policy can be applied only after the previous one is passed; routing information can be received only after passing all the filter policies configured.
Apply a route policy to routes from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **import**	
Specify an IPv6 ACL to filter routes from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **import**	
Specify an AS path ACL to filter IPv6 BGP routing information from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **import**	
Specify an IPv6 prefix list to filer routes from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **import**	
Specify the upper limit of prefixes that can be imported from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-limit** *limit* [*percentage*]	Optional The number is unlimited by default.

Note

A peer can has an inbound route filtering policy different from that of the peer group it belongs to. That is, peer group members can have different inbound route filtering policies.

Configuring IPv6 MBGP Route Dampening

Follow these steps to configure IPv6 MBGP route dampening parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure IPv6 MBGP route dampening parameters	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*]*	Optional Not configured by default

Configuring IPv6 MBGP Route Attributes

This section describes how to use IPv6 MBGP route attributes to affect IPv6 MBGP route selection. IPv6 MBGP route attributes involve:

- IPv6 MBGP protocol preference
- Default LOCAL_PREF attribute
- MED attribute
- NEXT_HOP attribute
- AS_PATH attribute

Configuration Prerequisites

Before the configuration, accomplish the following tasks:

- Enable IPv6
- Configure the IPv6 MBGP basic functions

Configuring IPv6 MBGP Route Preferences

Follow these steps to configure IPv6 MBGP route preferences:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure preferences for external, internal, local IPv6 MBGP routes	**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	Optional The default preference values of external, internal and local routes are 255, 255, and 130, respectively.

Configuring the Default Local Preference

Follow these steps to configure the default local preference:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Set the default local preference	**default local-preference** *value*	Optional By default, the default local preference is 100.

Configuring the MED Attribute

Follow these steps to configure the MED attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure a default MED value	**default med** *med-value*	Optional By default, the default *med-value* is 0.
Enable the comparison of the MED for routes from different ASs	**compare-different-as-med**	Optional Not enabled by default
Enable the comparison of the MED for routes from each AS	**bestroute compare-med**	Optional Disabled by default
Enable the comparison of the MED for routes from confederation peers	**bestroute med-confederation**	Optional Disabled by default

Configuring the NEXT_HOP Attribute

You can use the **peer next-hop-local** command to specify the local router as the next hop of routes sent to an IPv6 multicast iBGP peer/peer group. If load balancing is configured, the router specifies itself as the next hop of routes sent to the IPv6 multicast iBGP peer/peer group regardless of whether the **peer next-hop-local** command is configured.

In a “third party next hop" network, that is, the local router has two IPv6 multicast eBGP peers in a broadcast network, the router does not specify itself as the next hop of routes sent to the EBGP peers by default.

Follow these steps to specify the router as the next hop of routes sent to a peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure the router as the next hop of routes sent to the peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **next-hop-local**	Optional By default, IPv6 MBGP specifies the local router as the next hop for routes sent to an eBGP peer/peer group, but not for routes sent to an iBGP peer/peer group.

Configuring the AS_PATH Attribute

Follow these steps to configure the AS_PATH attribute:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Allow the local AS number to appear in the AS-PATH of routes from a peer/peer group and specify the number of times that the local AS number can appear in the AS-PATH of routes from the peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **allow-as-loop** [*number*]	Optional Not allowed by default
Disable IPv6 MBGP from considering the AS_PATH during best route selection	**bestroute as-path-neglect**	Optional Enabled by default
Configure updates to a peer/peer group to carry only the public AS number	**peer** { *ipv6-group-name* | *ipv6-address* } **public-as-only**	Optional By default, outbound IPv6 MBGP updates can carry private AS numbers.

Tuning and Optimizing IPv6 MBGP Networks

Configuration Prerequisites

Before tuning and optimizing an OSPF network, perform the following tasks:

- Enable IPv6
- Configure the IPv6 MBGP basic functions

Configuring IPv6 MBGP Soft Reset

After modifying a route selection policy, you have to reset IPv6 MBGP connections to make it take effect, causing short time disconnections.

After the route-refresh capability is enabled on all IPv6 MBGP routers in a network, when a route selection policy is modified on a router, the local router can perform dynamic route updates without tearing down IPv6 MBGP connections.

If the peer does not support route-refresh, you can save all route updates from the peer. When the route selection policy changes, you can refresh the IPv6 MBGP routing table and apply the new policy without tearing down IPv6 MBGP connections.

Soft reset through route-refresh

If the peer is enabled with route-refresh, when the IPv6 MBGP route selection policy is modified on a router, the router advertises a route-refresh message to its IPv6 MBGP peers, which resend their routing information to the router after receiving the message. Therefore, the local router can perform dynamic route update and apply the new policy without tearing down IPv6 MBGP connections.

Follow these steps to configure IPv6 MBGP soft reset through route-refresh:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—

To do…	Use the command…	Remarks
Enter IPv6 address family view	**ipv6-family**	—
Enable IPv6 BGP route refresh for a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **capability-advertise route-refresh**	Optional Enabled by default

Perform a manual soft-reset

If the peer does not support route-refresh, you can use the **peer keep-all-routes** command to save all the route updates from the peer, and then use the **refresh bgp ipv6 multicast** command to soft-reset IPv6 MBGP connections to refresh the IPv6 MBGP routing table and apply the new policy without tearing down IPv6 MBGP connections.

Follow these steps to perform a manual soft-reset

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Keep all routes from a peer/peer group regardless of whether they pass the inbound filtering policy	**peer** { *ipv6-group-name* | *ipv6-address* } **keep-all-routes**	Required Not kept by default
Exit to user view	**return**	—
Soft-reset IPv6 MBGP connections manually	**refresh bgp ipv6 multicast** { **all** | *ipv6-address* | **group** *ipv6-group-name* | **external** | **internal** } { **export** | **import** }	Optional

Configuring the Maximum Number of Equal-Cost Routes for Load-Balancing

Follow these steps to configure the maximum number of equal-cost routes for load-balancing:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure the maximum number of equal-cost routes for load balancing	**balance** *number*	Required By default, load balancing is disabled.

Configuring a Large Scale IPv6 MBGP Network

Configuration Prerequisites

Before configuring the following tasks, you need to configure IPv6 MBGP basic functions.

Configuring an IPv6 MBGP Peer Group

For easy management and configuration, you can organize some IPv6 MBGP peers having the same route update policy into a group, known as a peer group. A policy configured for a peer group applies to all the members in the group.

Follow these steps to configure an IPv6 MBGP peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 address family view	**ipv6-family**	—
Create an IPv6 BGP peer group	**group** *ipv6-group-name* [**external** | **internal**]	Required
Add a peer to the peer group	**peer** *ipv6-address* **group** *ipv6-group-name* [**as-number** *as-number*]	Required By default, no peer is added.
Exit to BGP view	**quit**	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Enable the configured IPv6 unicast BGP peer group to create the IPv6 MBGP peer group	**peer** *ipv6-group-name* **enable**	Required
Add the IPv6 MBGP peer into the peer group	**peer** *ipv6-address* **group** *ipv6-group-name*	Required By default, no peer is added.

Caution

- To create an IPv6 MBGP peer group, you need to enable an existing IPv6 unicast peer group in IPv6 MBGP address family view.
- Before adding an IPv6 MBGP peer to the IPv6 MBGP peer group, you need to add the corresponding IPv6 BGP unicast peer to the corresponding IPv6 BGP unicast peer group.

Configuring IPv6 MBGP Community

A peer group allows a group of peers to share the same policy, while a community allows a group of IPv6 MBGP routers in multiple ASs to share the same policy. The community attribute is propagated among IPv6 MBGP peers and not restricted to AS boundaries.

You can reference a route policy to modify the community attribute for routes sent to a peer. In addition, you can define extended community attributes as needed.

Follow these steps to advertise the community attribute to an IPv6 MBGP peer/peer group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Advertise the community attribute to an IPv6 MBGP peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-community**	Required By default, no community attribute is advertised to any peer group/peer.
Advertise the extended community attribute to an IPv6 MBGP peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-ext-community**	Required By default, no extended community attribute is advertised to any peer/peer group.
Apply a route policy to routes sent to an IPv6 MBGP peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	Required Not configured by default

Note

- You need to configure a route policy to define the community attribute, and apply the policy to outgoing routes.
- For route policy configuration, refer to *Route Policy Configuration* in the *IP Routing Volume*.

Configuring an IPv6 MBGP Route Reflector

To guarantee connectivity between IPv6 multicast iBGP peers, you need to make them fully meshed, but it becomes unpractical when there are too many IPv6 multicast iBGP peers. Using route reflectors can solve the problem.

Follow these steps to configure an IPv6 BGP route reflector:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter BGP view	**bgp** *as-number*	—
Enter IPv6 MBGP address family view	**ipv6-family multicast**	—
Configure the router as a route reflector and specify an IPv6 MBGP peer/peer group as its client	**peer** { *ipv6-group-name* | *ipv6-address* } **reflect-client**	Required Not configured by default
Enable route reflection between clients	**reflect between-clients**	Optional Enabled by default

To do…	Use the command…	Remarks
Configure the cluster ID of the route reflector	**reflector cluster-id** *cluster-id*	Optional By default, a route reflector uses its router ID as the cluster ID.

Note

- The clients of a route reflector should not be fully meshed, and the route reflector reflects the routes of a client to the other clients. If the clients are fully meshed, you need to disable route reflection between clients to reduce routing costs.
- If a cluster has multiple route reflectors, you need to specify the same cluster ID for these route reflectors to avoid routing loops.

Displaying and Maintaining IPv6 MBGP

Displaying IPv6 MBGP

To do…	Use the command…	Remarks
Display the IPv6 MBGP peer group information	**display bgp ipv6 multicast group** [*ipv6-group-name*]	Available in any view
Display IPv6 MBGP routing information injected with the **network** command	**display bgp ipv6 multicast network**	Available in any view
Display the IPv6 MBGP AS path information of routes	**display bgp ipv6 multicast paths** [*as-regular-expression*]	Available in any view
Display IPv6 MBGP peer/peer group information	**display bgp ipv6 multicast peer** [[*ipv6-address*] **verbose**]	Available in any view
Display IPv6 MBGP routing table information	**display bgp ipv6 multicast routing-table** [*ipv6-address prefix-length*]	Available in any view
Display IPv6 MBGP routing information matching a AS path ACL	**display bgp ipv6 multicast routing-table as-path-acl** *as-path-acl-number*	Available in any view
Display IPv6 MBGP routing information with the specified community attribute	**display bgp ipv6 multicast routing-table community** [*aa:nn*<1-13>] [**no-advertise** | **no-export** | **no-export-subconfed**]* [**whole-match**]	Available in any view
Display routing information matching an IPv6 MBGP community list	**display bgp ipv6 multicast routing-table community-list** { *basic-community-list-number* [**whole-match**] | *adv-community-list-number* }&<1-16>	Available in any view
Display IPv6 MBGP dampened routing information	**display bgp ipv6 multicast routing-table dampened**	Available in any view
Display IPv6 MBGP dampening parameter information	**display bgp ipv6 multicast routing-table dampening parameter**	Available in any view

To do…	Use the command…	Remarks
Display IPv6 MBGP routing information originated from different ASs	**display bgp ipv6 multicast routing-table different-origin-as**	Available in any view
Display IPv6 MBGP routing flap statistics	**display bgp ipv6 multicast routing-table flap-info** [**regular-expression** *as-regular-expression* | **as-path-acl** *as-path-acl-number* | *network-address* [*prefix-length* [**longer-match**]]]	Available in any view
Display the IPv6 MBGP routes received from or advertised to the IPv6 MBGP peer or peer group	**display bgp ipv6 multicast routing-table peer** *ipv6-address* { **advertised-routes** | **received-routes** } [*network-address prefix-length* | **statistic**]	Available in any view
Display IPv6 multicast routing information matching an AS regular expression	**display bgp ipv6 multicast routing-table regular-expression** *as-regular-expression*	Available in any view
Display IPv6 MBGP routing statistics	**display bgp ipv6 multicast routing-table statistic**	Available in any view
Display the IPv6 MBGP routing table information	**display ipv6 multicast routing-table** [**verbose**]	Available in any view
Display the multicast routing information of the specified destination address	**display ipv6 multicast routing-table** *ipv6-address prefix-length* [**longer-match**] [**verbose**]	Available in any view

Resetting IPv6 MBGP Connections

When an IPv6 MBGP route policy is changed, you can make the new configuration effective by resetting the IPv6 MBGP connections.

To do…	Use the command…	Remarks
Reset specified IPv6 MBGP connections	**reset bgp ipv6 multicast** { *as-number* | *ipv6-address* | **all** | **group** *ipv6-group-name* | **external** | **internal** }	Available in user view

Clearing IPv6 MBGP Information

To do…	Use the command…	Remarks
Clear dampened IPv6 MBGP routing information and release suppressed routes	**reset bgp ipv6 multicast dampening** [*ipv6-address prefix-length*]	Available in user view
Clear IPv6 MBGP route flap statistics	**reset bgp ipv6 multicast flap-info** [*ipv6-address/prefix-length* | **regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number*]	Available in user view

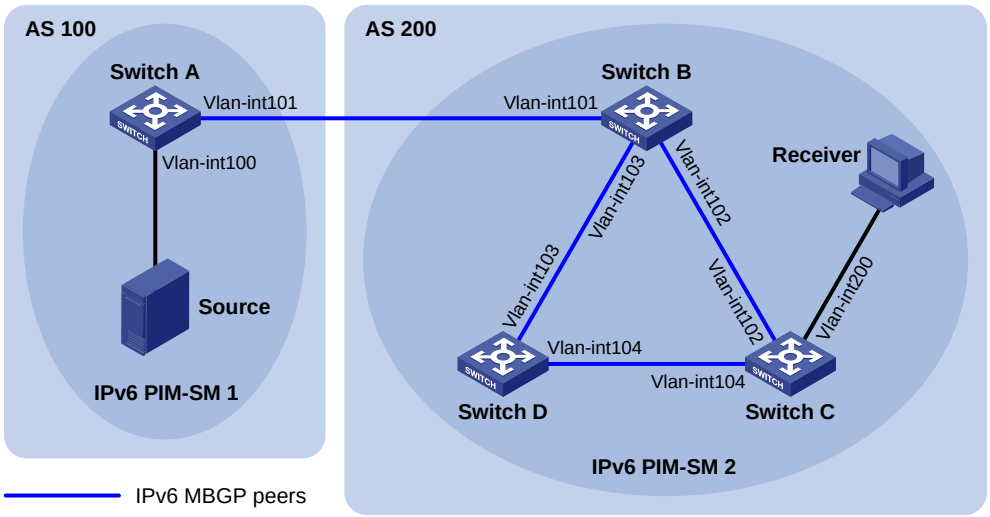
IPv6 MBGP Configuration Example

Network requirements

As shown in the following figure:

- IPv6 PIM-SM 1 is in AS 100 and IPv6 PIM-SM 2 is in AS 200. OSPFv3 is the IGP in the two ASs, and IPv6 MBGP runs between the two ASs to exchange IPv6 multicast route information.
- The IPv6 multicast source belongs to IPv6 PIM-SM 1 and the receiver belongs to IPv6 PIM-SM 2.
- It is required that the respective VLAN-interface 101 of Switch A and Switch B be configured as the C-BSR and C-RP of the respective IPv6 PIM-SM domains.

Figure 1-1 Network diagram for IPv6 MBGP configuration



Device	Interface	IP address	Device	Interface	IP address
Source	-	1002::100/64	Switch C	Vlan-int200	3002::1/64
Switch A	Vlan-int100	1002::1/64		Vlan-int102	2001::2/64
	Vlan-int101	1001::1/64		Vlan-int104	3001::1/64
Switch B	Vlan-int101	1001::2/64	Switch D	Vlan-int103	2002::2/64
	Vlan-int102	2001::1/64		Vlan-int104	3001::2/64
	Vlan-int103	2002::1/64			

Configuration procedure

1) Configure IPv6 addresses for interfaces as shown in the above figure (omitted).
2) Configure OSPFv3 (omitted).
3) Enable IPv6 multicast routing, IPv6 PIM-SM and MLD, and configure an IPv6 PIM-SM domain border.

Enable IPv6 multicast routing on Switch A, and enable IPv6 PIM-SM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim ipv6 sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 sm
[SwitchA-Vlan-interface101] quit
```

The configuration on Switch B and Switch D is similar to the configuration on Switch A.

Enable IPv6 multicast routing on Switch C, enable IPv6 PIM-SM on each interface, and enable MLD on the host-side interface VLAN-interface 200.

```
<SwitchC> system-view
[SwitchC] multicast ipv6 routing-enable
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim ipv6 sm
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 104
[SwitchC-Vlan-interface104] pim ipv6 sm
[SwitchC-Vlan-interface104] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim ipv6 sm
[SwitchC-Vlan-interface200] mld enable
[SwitchC-Vlan-interface200] quit
```

Configure an IPv6 PIM domain border on Switch A.

```
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim ipv6 bsr-boundary
[SwitchA-Vlan-interface101] quit
```

Configure an IPv6 PIM domain border on Switch B.

```
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim ipv6 bsr-boundary
[SwitchB-Vlan-interface101] quit
```

4) Configure the position of C-BSR and C-RP.

Configure the position of C-BSR and C-RP on Switch A.

```
[SwitchA] pim ipv6
[SwitchA-pim6] c-bsr 1001::1
[SwitchA-pim6] c-rp 1001::1
[SwitchA-pim6] quit
```

Configure the position of C-BSR and C-RP on Switch B.

```
[SwitchB] pim ipv6
[SwitchB-pim6] c-bsr 1001::2
[SwitchB-pim6] c-rp 1001::2
[SwitchB-pim6] quit
```

5) Configure BGP, specify the IPv6 MBGP peer and enable direct route redistribution.

On Switch A, configure the IPv6 MBGP peer and enable direct route redistribution.

```
[SwitchA] ipv6
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] ipv6-family
[SwitchA-bgp-af-ipv6] peer 1001::2 as-number 200
[SwitchA-bgp-af-ipv6] import-route direct
[SwitchA-bgp-af-ipv6] quit
[SwitchA-bgp] ipv6-family multicast
[SwitchA-bgp-af-ipv6-mul] peer 1001::2 enable
[SwitchA-bgp-af-ipv6-mul] import-route direct
```

```
[SwitchA-bgp-af-ipv6-mul] quit
[SwitchA-bgp] quit
```

On Switch B, configure the IPv6 MBGP peers and redistribute OSPF routes.

```
[SwitchB] ipv6
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 1001::1 as-number 100
[SwitchB-bgp-af-ipv6] import-route ospfv3 1
[SwitchB-bgp-af-ipv6] quit
[SwitchB-bgp] ipv6-family multicast
[SwitchB-bgp-af-ipv6-mul] peer 1001::1 enable
[SwitchB-bgp-af-ipv6-mul] import-route ospfv3 1
[SwitchB-bgp-af-ipv6-mul] quit
[SwitchB-bgp] quit
```

6) Verify the configuration

You can use the **display bgp ipv6 multicast peer** command to display IPv6 MBGP peers on a switch. For example, display IPv6 MBGP peers on Switch B.

```
[SwitchB] display bgp ipv6 multicast peer

 BGP local router ID : 2.2.2.2
 Local AS number : 200
 Total number of peers : 3                 Peers in established state : 3

  Peer        V    AS  MsgRcvd  MsgSent  OutQ PrefRcv Up/Down  State

  1001::1     4   100       56       56     0       0 00:40:54 Established
```

Table of Contents

1 MLD Snooping Configuration

When configuring MLD Snooping, go to these sections for information you are interested in:

MLD Snooping Overview

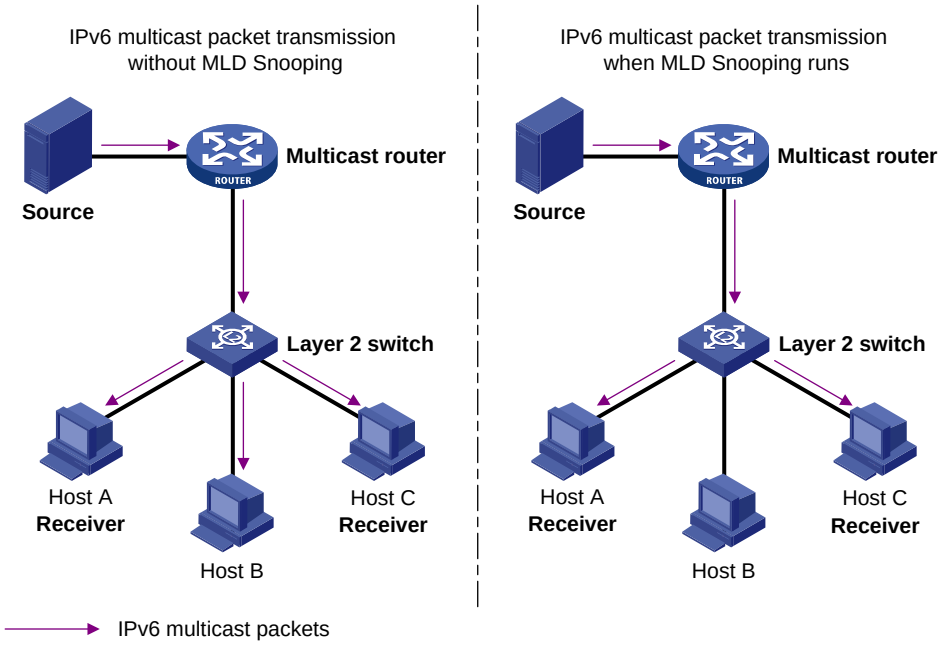
Multicast Listener Discovery Snooping (MLD Snooping) is an IPv6 multicast constraining mechanism that runs on Layer 2 devices to manage and control IPv6 multicast groups.

Introduction to MLD Snooping

By analyzing received MLD messages, a Layer 2 device running MLD Snooping establishes mappings between ports and multicast MAC addresses and forwards IPv6 multicast data based on these mappings.

As shown in Figure 1-1, when MLD Snooping is not running, IPv6 multicast packets are broadcast to all devices at Layer 2. When MLD Snooping runs, multicast packets for known IPv6 multicast groups are multicast to the receivers at Layer 2.

Figure 1-1 Before and after MLD Snooping is enabled on the Layer 2 device



MLD Snooping forwards multicast data to only the receivers requiring it at Layer 2. It brings the following advantages:

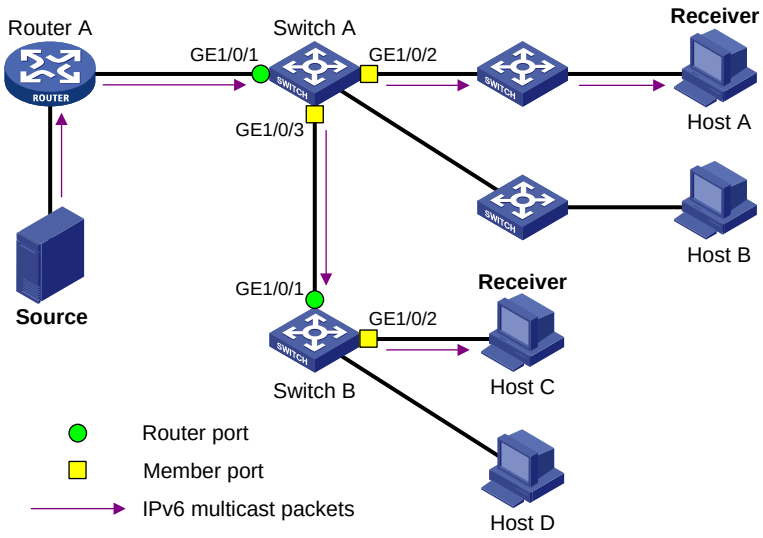
- Reducing Layer 2 broadcast packets, thus saving network bandwidth.
- Enhancing the security of multicast traffic.
- Facilitating the implementation of per-host accounting.

Basic Concepts in MLD Snooping

MLD Snooping related ports

As shown in Figure 1-2, Router A connects to the multicast source, MLD Snooping runs on Switch A and Switch B, Host A and Host C are receiver hosts (namely, IPv6 multicast group members).

Figure 1-2 MLD Snooping related ports



Ports involved in MLD Snooping, as shown in Figure 1-2, are described as follows:

- Router port: A router port is a port on the Ethernet switch that leads switch towards the Layer-3 multicast device (DR or MLD querier). In the figure, GigabitEthernet 1/0/1 of Switch A and GigabitEthernet 1/0/1 of Switch B are router ports. The switch registers all its local router ports in its router port list.
- Member port: A member port (also known as IPv6 multicast group member port) is a port on the Ethernet switch that leads towards multicast group members. In the figure, GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 of Switch A and GigabitEthernet 1/0/2 of Switch B are member ports. The switch registers all the member ports on the local device in its MLD Snooping forwarding table.

Note

- Whenever mentioned in this document, a router port is a router-connecting port on the switch, rather than a port on a router.
- Unless otherwise specified, router/member ports mentioned in this document include static and dynamic ports.
- On an MLD Snooping-enabled switch, the ports that received MLD general queries with the source address other than 0::0 or IPv6 PIM hello messages are dynamic router ports. For details about IPv6 PIM hello messages, see *IPv6 PIM Configuration* of the *IP Multicast Volume*.

Aging timers for dynamic ports in MLD Snooping

Table 1-1 Aging timers for dynamic ports in MLD Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Dynamic router port aging timer	For each dynamic router port, the switch sets a timer initialized to the dynamic router port aging time.	MLD general query of which the source address is not 0::0 or IPv6 PIM hello.	The switch removes this port from its router port list.
Dynamic member port aging timer	When a port dynamically joins an IPv6 multicast group, the switch sets a timer for the port, which is initialized to the dynamic member port aging time.	MLD report message.	The switch removes this port from the MLD Snooping forwarding table.

Note

The port aging mechanism of MLD Snooping works only for dynamic ports; a static port will never age out.

How MLD Snooping Works

A switch running MLD Snooping performs different actions when it receives different MLD messages, as follows:

Caution

The description about adding or deleting a port in this section is only for a dynamic port. Static ports can be added or deleted only through the corresponding configurations. For details, see Configuring Static Ports.

General queries

The MLD querier periodically sends MLD general queries to all hosts and routers (FF02::1) on the local subnet to find out whether IPv6 multicast group members exist on the subnet.

Upon receiving an MLD general query, the switch forwards it through all ports in the VLAN except the port on which it received the MLD query and performs the following:

- If the port on which it the switch received the MLD query is a dynamic router port in its router port list, the switch resets the aging timer for this dynamic router port.
- If the port is not included in its router port list, the switch adds it into its router port list as a dynamic router port and sets an aging timer for it.

Membership reports

A host sends an MLD report to the MLD querier in the following circumstances:

- Upon receiving an MLD query, an IPv6 multicast group member host responds with an MLD report.
- When intended to join an IPv6 multicast group, a host sends an MLD report to the MLD querier to announce that it is interested in the multicast information addressed to that IPv6 multicast group.

Upon receiving an MLD report, the switch forwards it through all the router ports in the VLAN, resolves the address of the reported IPv6 multicast group, and performs the following to the receiving port:

- If no forwarding table entry exists for the reported IPv6 multicast group, the switch creates an entry, adds the port as a dynamic member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported IPv6 multicast group, but the port is not included in the outgoing port list for that group, the switch adds the port as a dynamic member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported IPv6 multicast group and the port is included in the outgoing port list, which means that this port is already a dynamic member port, the switch resets the member port aging timer for that port.

Note

A switch does not forward an MLD report through a non-router port. The reason is as follows: Due to the MLD report suppression mechanism, if the switch forwards a report message through a member port, all the attached hosts listening to the reported IPv6 multicast address will suppress their own reports upon receiving this report, and this will prevent the switch from knowing whether the reported multicast group still has active members attached to that port.

For the description of MLD report suppression mechanism, refer to *MLD Configuration* in the *IP Multicast volume*.

Done messages

When a host leaves an IPv6 multicast group, the host sends an MLD done message to the multicast router.

When the switch receives an MLD done message on a dynamic member port, the switch first checks whether a forwarding table entry for the IPv6 multicast group address in the message exists, and, if one exists, whether the outgoing port list contains the port.

- If the forwarding table entry does not exist or if the outgoing port list does not contain the port, the switch discards the MLD done message instead of forwarding it to any port.
- If the forwarding table entry exists and the outgoing port list contains the port, the switch forwards the MLD done message to all router ports in the native VLAN. Because the switch does not know whether any other hosts attached to the port are still listening to that IPv6 multicast group address, the switch does not immediately remove the port from the outgoing port list of the forwarding table entry for that group; instead, it resets the aging timer for the port.

Upon receiving an MLD done message from a host, the MLD querier resolves the IPv6 multicast group address in the message and sends an MLD multicast-address-specific query to that IPv6 multicast group address through the port that received the MLD done message. Upon receiving the MLD multicast-address-specific query, the switch forwards it through all the router ports in the VLAN and all member ports for that IPv6 multicast group, and performs the following to the receiving port:

- If any MLD report in response to the MLD multicast-address-specific query is received on the port (suppose it is a dynamic member port) before its aging timer expires, this means that some host attached to the port is receiving or expecting to receive IPv6 multicast data for that IPv6 multicast group. The switch resets the aging timer for the port.
- If no MLD report in response to the MLD multicast-address-specific query is received on the port before its aging timer expires, this means that no hosts attached to the port are still listening to that IPv6 multicast group address. The switch removes the port from the outgoing port list of the forwarding table entry for that IPv6 multicast group when the aging timer expires.

Processing of IPv6 Multicast Protocol Messages

With Layer 3 multicast routing enabled, an MLD Snooping switch processes IPv6 multicast protocol messages differently under different conditions, specifically as follows:

1) If only MLD is enabled, or both MLD and IPv6 PIM are enabled on the switch, the switch handles IPv6 multicast protocol messages in the normal way.
2) In only IPv6 PIM is enabled on the switch:

- The switch broadcasts MLD messages as unknown messages in the VLAN.
- Upon receiving an IPv6 PIM hello message, the switch will maintain the corresponding dynamic router port.

3) When MLD is disabled on the switch:

- If IPv6 PIM is disabled, the switch deletes all its dynamic member ports and dynamic router ports.
- If IPv6 PIM is enabled, the switch deletes only its dynamic member ports without deleting its dynamic router ports.

Note

On a switch with Layer-3 IPv6 multicast routing enabled, use the **display mld group port-info** command to view Layer-2 port information.

For details about the **display mld group port-info** command, refer to *MLD Commands* in the *IP Multicast Volume*.

---

4) When IPv6 PIM is disabled on the switch:

- If MLD is disabled, the switch deletes all its dynamic router ports.

- If MLD is enabled, the switch maintains all its dynamic member ports and dynamic router ports.

Protocols and Standards

MLD Snooping is documented in:

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

MLD Snooping Configuration Task List

Complete these tasks to configure MLD Snooping:

Task		Remarks
Configuring Basic Functions of MLD Snooping	Enabling MLD Snooping	Required
	Configuring the Version of MLD Snooping	Optional
Configuring MLD Snooping Port Functions	Configuring Aging Timers for Dynamic Ports	Optional
	Configuring Static Ports	Optional
	Configuring Simulated Joining	Optional
	Configuring Fast Leave Processing	Optional
Configuring MLD Snooping Querier	Enabling MLD Snooping Querier	Optional
	Configuring MLD Queries and Responses	Optional
	Configuring Source IPv6 Addresses of MLD Queries	Optional
Configuring an MLD Snooping Policy	Configuring an IPv6 Multicast Group Filter	Optional
	Configuring IPv6 Multicast Source Port Filtering	Optional
	Configuring Dropping Unknown IPv6 Multicast Data	Optional
	Configuring MLD Report Suppression	Optional
	Configuring Maximum Multicast Groups that Can Be Joined on a Port	Optional
	Configuring IPv6 Multicast Group Replacement	Optional

Note

- Configurations made in MLD Snooping view are effective for all VLANs, while configurations made in VLAN view are effective only for ports belonging to the current VLAN. For a given VLAN, a configuration made in MLD Snooping view is effective only if the same configuration is not made in VLAN view.
- Configurations made in MLD Snooping view are effective for all ports; configurations made in Ethernet port view are effective only for the current port; configurations made in Layer 2 aggregate port view are effect only for the current port; configurations made in port group view are effective only for all the ports in the current port group. For a given port, a configuration made in MLD Snooping view is effective only if the same configuration is not made in Ethernet port view, Layer 2 aggregate port view or port group view.
- For MLD Snooping, configurations made on a Layer 2 aggregate port do not interfere with configurations made on its member ports, nor do they take part in aggregation calculations; configurations made on a member port of the aggregate group will not take effect until it leaves the aggregate group.

Configuring Basic Functions of MLD Snooping

Configuration Prerequisites

Before configuring the basic functions of MLD Snooping, complete the following tasks:

- Configure the corresponding VLANs

Before configuring the basic functions of MLD Snooping, prepare the following data:

- The version of MLD Snooping

Enabling MLD Snooping

Follow these steps to enable MLD Snooping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enable MLD Snooping globally and enter MLD Snooping view	**mld-snooping**	Required Disabled by default
Return to system view	**quit**	—
Enter VLAN view	**vlan** *vlan-id*	—
Enable MLD Snooping in the VLAN	**mld-snooping enable**	Required Disabled by default

Note

- MLD Snooping must be enabled globally before it can be enabled in a VLAN.
- After enabling MLD Snooping in a VLAN, you cannot enable MLD and/or IPv6 PIM on the corresponding VLAN interface, and vice versa.
- When you enable MLD Snooping in a specified VLAN, this function takes effect for ports in this VLAN only.

Configuring the Version of MLD Snooping

By configuring the MLD Snooping version, you actually configure the version of MLD messages that MLD Snooping can process.

- MLD Snooping version 1 can process MLDv1 messages, but cannot analyze and process MLDv2 messages, which will be flooded in the VLAN.
- MLD Snooping version 2 can process MLDv1 and MLDv2 messages.

Follow these steps to configure the version of MLD Snooping:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure the version of MLD Snooping	**mld-snooping version** *version-number*	Optional Version 1 by default

Caution

If you switch MLD Snooping from version 2 to version 1, the system will clear all MLD Snooping forwarding entries from dynamic joining, and will:

- Keep forwarding entries from version 2 static (*, G) joining;
- Clear forwarding entries from version 2 static (S, G) joining, which will be restored when MLD Snooping is switched back to version 2.

For details about static joining, refer to Configuring Static Ports.

Configuring MLD Snooping Port Functions

Configuration Prerequisites

Before configuring MLD Snooping port functions, complete the following tasks:

- Enable MLD Snooping in the VLAN or enable MLD on the desired VLAN interface
- Configure the corresponding port groups

Before configuring MLD Snooping port functions, prepare the following data:

- Aging time of dynamic router ports,
- Aging timer of dynamic member ports, and

- IPv6 multicast group and IPv6 multicast source addresses

Configuring Aging Timers for Dynamic Ports

If the switch receives no MLD general queries or IPv6 PIM hello messages on a dynamic router port, the switch removes the port from the router port list when the aging timer of the port expires.

If the switch receives no MLD reports for an IPv6 multicast group on a dynamic member port, the switch removes the port from the outgoing port list of the forwarding table entry for that IPv6 multicast group when the port aging timer expires.

If IPv6 multicast group memberships change frequently, you can set a relatively small value for the dynamic member port aging timer.

Configuring aging timers for dynamic ports globally

Follow these steps to configure aging timers for dynamic ports globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Configure dynamic router port aging time	**router-aging-time** *interval*	Optional 260 seconds by default
Configure dynamic member port aging time	**host-aging-time** *interval*	Optional 260 seconds by default

Configuring aging timers for dynamic ports in a VLAN

Follow these steps to configure aging timers for dynamic ports in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure dynamic router port aging time	**mld-snooping** **router-aging-time** *interval*	Optional 260 seconds by default
Configure dynamic member port aging time	**mld-snooping** **host-aging-time** *interval*	Optional 260 seconds by default

Configuring Static Ports

If all the hosts attached to a port is interested in the IPv6 multicast data addressed to a particular IPv6 multicast group, you can configure that port as a static member port for that IPv6 multicast group.

You can configure a port of a switch to be a static router port, through which the switch can forward all IPv6 multicast data it received.

Follow these steps to configure static ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure the port(s) as static member port(s)	**mld-snooping static-group** *ipv6-group-address* [**source-ip** *ipv6-source-address*] **vlan** *vlan-id*	Required No static member ports by default
Configure the port(s) as static router port(s)	**mld-snooping static-router-port** vlan *vlan-id*	Required No static router ports by default

Note

- An IPv6 static (S, G) join takes effect only if a valid IPv6 multicast source address is specified and MLD Snooping version 2 is currently running.
- A static member port does not respond to queries from the MLD querier; when static (*, G) or (S, G) joining is enabled or disabled on a port, the port does not send an unsolicited MLD report or an MLD done message.
- If MLD is enabled on the virtual interface of a VLAN on a switch that supports both MLD Snooping and MLD and you want a port in that VLAN to be a static member port for an IPv6 multicast group or an IPv6 multicast source and group, in addition to configuring the port as a static member port, you need to use the **mld static-group** command to configure the VLAN interface to be a static member of the IPv6 multicast group or source and group. For details of the **mld static-group** command, refer to *MLD Commands* in the *IP Multicast Volume*.
- Static member ports and static router ports never age out. To remove such a port, you need to use the corresponding **undo** command.

Configuring Simulated Joining

Generally, a host running MLD responds to MLD queries from the MLD querier. If a host fails to respond due to some reasons, the multicast router will deem that no member of this IPv6 multicast group exists on the network segment, and therefore will remove the corresponding forwarding path.

To avoid this situation from happening, you can enable simulated joining on a port of the switch, namely configure the port as a simulated member host for an IPv6 multicast group. When an MLD query is received, simulated host gives a response. Thus, the switch can continue receiving IPv6 multicast data.

A simulated host acts like a real host, as follows:

- When a port is configured as a simulated member host, the switch sends an unsolicited MLD report through that port.
- After a port is configured as a simulated member host, the switch responds to MLD general queries by sending MLD reports through that port.

- When the simulated joining function is disabled on a port, the switch sends an MLD done message through that port.

Follow these steps to configure simulated joining:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure simulated joining	**mld-snooping host-join** *ipv6-group-address* [**source-ip** *ipv6-source-address*] **vlan** *vlan-id*	Required Disabled by default

Note

- Each simulated host is equivalent to an independent host. For example, when receiving an MLD query, the simulated host corresponding to each configuration responds respectively.
- Unlike a static member port, a port configured as a simulated member host will age out like a dynamic member port.

Configuring Fast Leave Processing

The fast leave processing feature allows the switch to process MLD done messages in a fast way. With the fast leave processing feature enabled, when receiving an MLD done message on a port, the switch immediately removes that port from the outgoing port list of the forwarding table entry for the indicated IPv6 multicast group. Then, when receiving MLD done multicast-address-specific queries for that IPv6 multicast group, the switch will not forward them to that port.

In VLANs where only one host is attached to each port, fast leave processing helps improve bandwidth and resource usage.

Configuring fast leave processing globally

Follow these steps to configure fast leave processing globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Enable fast leave processing	**fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Configuring fast leave processing on a port or a group of ports

Follow these steps to configure fast leave processing on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable fast leave processing	**mld-snooping fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Caution

If fast leave processing is enabled on a port to which more than one host is connected, when one host leaves an IPv6 multicast group, the other hosts connected to port and interested in the same IPv6 multicast group will fail to receive IPv6 multicast data addressed to that group.

Configuring MLD Snooping Querier

Configuration Prerequisites

Before configuring MLD Snooping querier, complete the following task:

- Enable MLD Snooping in the VLAN.

Before configuring MLD Snooping querier, prepare the following data:

- MLD general query interval,
- MLD last-member query interval,
- Maximum response time for MLD general queries,
- Source IPv6 address of MLD general queries, and
- Source IPv6 address of MLD multicast-address-specific queries.

Enabling MLD Snooping Querier

In an IPv6 multicast network running MLD, a multicast router or Layer 3 multicast switch is responsible for sending periodic MLD general queries, so that all Layer 3 multicast devices can establish and maintain multicast forwarding entries, thus to forward multicast traffic correctly at the network layer. This router or Layer 3 switch is called MLD querier.

However, a Layer 2 multicast switch does not support MLD, and therefore cannot send MLD general queries by default. By enabling MLD Snooping querier on a Layer 2 switch in a VLAN where multicast traffic needs to be Layer-2 switched only and no Layer 3 multicast devices are present, the Layer 2 switch will act as the MLD querier to send periodic MLD queries, thus allowing multicast forwarding entries to be established and maintained at the data link layer.

Follow these steps to enable the MLD Snooping querier:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—

To do...	Use the command...	Remarks
Enable the MLD Snooping querier	**mld-snooping querier**	Required Disabled by default

Caution

It is meaningless to configure an MLD Snooping querier in an IPv6 multicast network running MLD. Although an MLD Snooping querier does not take part in MLD querier elections, it may affect MLD querier elections because it sends MLD general queries with a low source IPv6 address.
For details about MLD querier, see *MLD Configuration* of the *IP Multicast Volume*.

Configuring MLD Queries and Responses

You can tune the MLD general query interval based on actual condition of the network.

Upon receiving an MLD query (general query or multicast-address-specific query), a host starts a timer for each IPv6 multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time (the host obtains the value of the maximum response time from the Max Response Time field in the MLD query it received). When the timer value comes down to 0, the host sends an MLD report to the corresponding IPv6 multicast group.

An appropriate setting of the maximum response time for MLD queries allows hosts to respond to queries quickly and avoids bursts of MLD traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expire simultaneously.

- For MLD general queries, you can configure the maximum response time to fill their Max Response time field.
- For MLD multicast-address-specific queries, you can configure the MLD last-member query interval to fill their Max Response time field. Namely, for MLD multicast-address-specific queries, the maximum response time equals to the MLD last-member query interval.

Configuring MLD queries and responses globally

Follow these steps to configure MLD queries and responses globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Configure the maximum response time for MLD general queries	**max-response-time** *interval*	Optional 10 seconds by default
Configure the MLD last-member query interval	**last-listener-query-interval** *interval*	Optional 1 second by default

Configuring MLD queries and responses in a VLAN

Follow these steps to configure MLD queries and responses in a VLAN

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure MLD query interval	**mld-snooping query-interval** *interval*	Optional 125 seconds by default
Configure the maximum response time for MLD general queries	**mld-snooping max-response-time** *interval*	Optional 10 seconds by default
Configure the MLD last-member query interval	**mld-snooping last-listener-query-interval** *interval*	Optional 1 second by default

Caution

Make sure that the MLD query interval is greater than the maximum response time for MLD general queries; otherwise undesired deletion of IPv6 multicast members may occur.

Configuring Source IPv6 Addresses of MLD Queries

This configuration allows you to change the source IPv6 address of MLD queries.

Follow these steps to configure source IPv6 addresses of MLD queries:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Configure the source IPv6 address of MLD general queries	**mld-snooping general-query source-ip** { **current-interface** | *ipv6-address* }	Optional FE80::02FF:FFFF:FE00:0001 by default
Configure the source IPv6 address of MLD multicast-address-specific queries	**mld-snooping special-query source-ip** { **current-interface** | *ipv6-address* }	Optional FE80::02FF:FFFF:FE00:0001 by default

Caution

The source IPv6 address of MLD query messages may affect MLD querier election within the segment.

Configuring an MLD Snooping Policy

Configuration Prerequisites

Before configuring an MLD Snooping policy, complete the following tasks:

- Enable MLD Snooping in the VLAN or enable MLD on the desired VLAN interface

Before configuring an MLD Snooping policy, prepare the following data:

- IPv6 ACL rule for IPv6 multicast group filtering
- The maximum number of IPv6 multicast groups that can pass the ports

Configuring an IPv6 Multicast Group Filter

On a MLD Snooping–enabled switch, the configuration of an IPv6 multicast group filter allows the service provider to define limits of multicast programs available to different users.

In an actual application, when a user requests a multicast program, the user's host initiates an MLD report. Upon receiving this report message, the switch checks the report against the configured ACL rule. If the port on which the report was received can join this IPv6 multicast group, the switch adds an entry for this port in the MLD Snooping forwarding table; otherwise the switch drops this report message. Any IPv6 multicast data that fails the ACL check will not be sent to this port. In this way, the service provider can control the VOD programs provided for multicast users.

Configuring an IPv6 multicast group filter globally

Follow these steps to configure an IPv6 multicast group globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Configure an IPv6 multicast group filter	**group-policy** *acl6-number* [**vlan** *vlan-list*]	Required No IPv6 filter configured by default.

Configuring an IPv6 multicast group filter on a port or a group of ports

Follow these steps to configure an IPv6 multicast group filer on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure an IPv6 multicast group filter	**mld-snooping group-policy** *acl6-number* [**vlan** *vlan-list*]	Required No IPv6 filter configured by default.

Configuring IPv6 Multicast Source Port Filtering

With the IPv6 multicast source port filtering feature enabled on a port, the port can be connected with IPv6 multicast receivers only rather than with multicast sources, because the port will block all IPv6 multicast data packets while it permits multicast protocol packets to pass.

If this feature is disabled on a port, the port can be connected with both multicast sources and IPv6 multicast receivers.

Configuring IPv6 multicast source port filtering globally

Follow these steps to configure IPv6 multicast source port filtering:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Enable IPv6 multicast source port filtering	**source-deny port** *interface-list*	Required Disabled by default

Configuring IPv6 multicast source port filtering on a port or a group of ports

Follow these steps to configure IPv6 multicast source port filtering on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable IPv6 multicast source port filtering	**mld-snooping source-deny**	Required Disabled by default

Note

Some models of devices, when enabled to filter IPv6 multicast data based on the source ports, are automatically enabled to filter IPv4 multicast data based on the source ports.

Configuring Dropping Unknown IPv6 Multicast Data

Unknown IPv6 multicast data refers to IPv6 multicast data for which no forwarding entries exist in the MLD Snooping forwarding table: When the switch receives such IPv6 multicast traffic:

- With the function of dropping unknown IPv6 multicast data enabled, the switch drops all unknown IPv6 multicast data received.
- With the function of dropping unknown IPv6 multicast data disabled, the switch floods unknown IPv6 multicast data in the VLAN to which the unknown IPv6 multicast data belongs.

Follow these steps to enable dropping unknown IPv6 multicast data in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter VLAN view	**vlan** *vlan-id*	—
Enable dropping unknown IPv6 multicast data	**mld-snooping drop-unknown**	Required Disabled by default

Configuring MLD Report Suppression

When a Layer 2 device receives an MLD report from an IPv6 multicast group member, the Layer 2 device forwards the message to the Layer 3 device directly connected with it. Thus, when multiple members belonging to an IPv6 multicast group exist on the Layer 2 device, the Layer 3 device directly connected with it will receive duplicate MLD reports from these members.

With the MLD report suppression function enabled, within a query interval, the Layer 2 device forwards only the first MLD report of an IPv6 group to the Layer 3 device and will not forward the subsequent MLD reports from the same multicast group to the Layer 3 device. This helps reduce the number of packets being transmitted over the network.

Follow these steps to configure MLD report suppression:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Enable MLD report suppression	**report-aggregation**	Optional Enabled by default

Configuring Maximum Multicast Groups that Can Be Joined on a Port

By configuring the maximum number of IPv6 multicast groups that can be joined on a port or a group of ports, you can limit the number of multicast programs available to VOD users, thus to control the traffic on the port.

Follow these steps configure the maximum number of IPv6 multicast groups that can be joined on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Configure the maximum number of IPv6 multicast groups that can be joined on a port	**mld-snooping group-limit** *limit* [**vlan** *vlan-list*]	Optional 512 by default.

Note

- When the number of IPv6 multicast groups that can be joined on a port reaches the maximum number configured, the system deletes all the forwarding entries persistent to that port from the MLD Snooping forwarding table, and the hosts on this port need to join IPv6 multicast groups again.
- If you have configured static or simulated joining on a port, however, when the number of IPv6 multicast groups on the port exceeds the configured threshold, the system deletes all the forwarding entries persistent to that port from the MLD Snooping forwarding table and applies the static or simulated joining again, until the number of IPv6 multicast groups joined by the port comes back within the configured threshold.

Configuring IPv6 Multicast Group Replacement

For some special reasons, the number of IPv6 multicast groups passing through a switch or port may exceed the number configured for the switch or the port. In addition, in some specific applications, an IPv6 multicast group newly joined on the switch needs to replace an existing IPv6 multicast group automatically. A typical example is “channel switching”, namely, by joining the new multicast group, a user automatically switches from the current IPv6 multicast group to the new one.

To address this situation, you can enable the IPv6 multicast group replacement function on the switch or certain ports. When the number of IPv6 multicast groups a switch or a port has joined exceeds the limit.

- If the IPv6 multicast group replacement is enabled, the newly joined IPv6 multicast group automatically replaces an existing IPv6 multicast group with the lowest IPv6 address.
- If the IPv6 multicast group replacement is not enabled, new MLD reports will be automatically discarded.

Configuring IPv6 multicast group replacement globally

Follow these steps to configure IPv6 multicast group replacement globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter MLD Snooping view	**mld-snooping**	—
Enable IPv6 multicast group replacement	**overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Configuring IPv6 multicast group replacement on a port or a group of ports

Follow these steps to configure IPv6 multicast group replacement on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter Ethernet port/Layer 2 aggregate port view or port group view	**interface** *interface-type interface-number*	Required Use either approach
	port-group manual *port-group-name*	
Enable IPv6 multicast group replacement	**mld-snooping overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Caution

Be sure to configure the maximum number of IPv6 multicast groups allowed on a port (refer to Configuring Maximum Multicast Groups that Can Be Joined on a Port) before enabling IPv6 multicast group replacement. Otherwise, the IPv6 multicast group replacement functionality will not take effect.

Displaying and Maintaining MLD Snooping

To do…	Use the command...	Remarks
View MLD Snooping multicast group information	**display mld-snooping group** [**vlan** *vlan-id*] [**slot** *slot-number*] [**verbose**]	Available in any view
View the statistics information of MLD messages learned by MLD Snooping	**display mld-snooping statistics**	Available in any view
Clear MLD Snooping multicast group information	**reset mld-snooping group** { *ipv6-group-address* | **all** } [**vlan** *vlan-id*]	Available in user view
Clear the statistics information of all kinds of MLD messages learned by MLD Snooping	**reset mld-snooping statistics**	Available in user view

Note

- The **reset mld-snooping group** command works only on an MLD Snooping–enabled VLAN, but not on a VLAN with MLD enabled on its VLAN interface.
- The **reset mld-snooping group** command cannot clear the MLD Snooping multicast group information for static joining.

MLD Snooping Configuration Examples

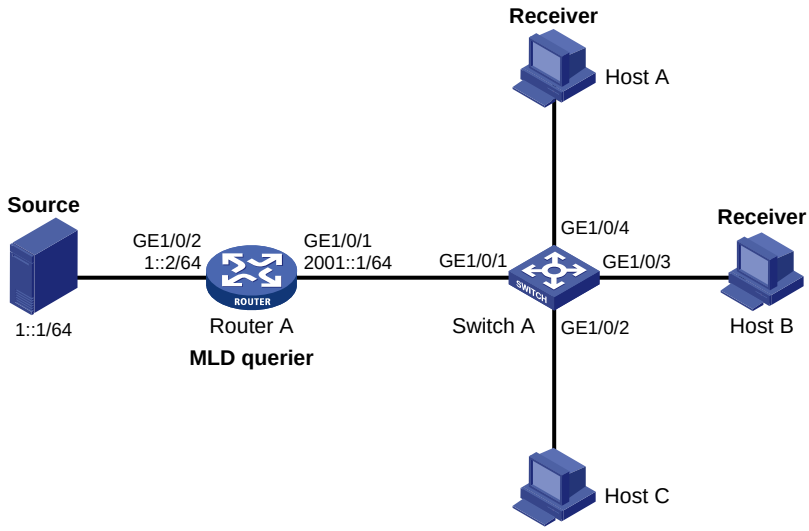
Configuring IPv6 Group Policy and Simulated Joining

Network requirements

- As shown in Figure 1-3, Router A connects to the IPv6 multicast source through GigabitEthernet 1/0/2 and to Switch A through GigabitEthernet 1/0/1. Router A is the MLD querier on the subnet.
- MLDv1 is required on Router A, MLD Snooping version 1 is required on Switch A, and Router A will act as the MLD querier on the subnet.
- It is required that the receivers, Host A and Host B, attached to Switch A can receive IPv6 multicast traffic addressed to IPv6 multicast group FF1E::101 only.
- It is required that IPv6 multicast data for group FF1E::101 can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A even if Host A and Host B accidentally, temporarily stop receiving IPv6 multicast data.

Network diagram

Figure 1-3 Network diagram for IPv6 group policy simulated joining configuration



Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses

Enable IPv6 forwarding and configure an IPv6 address and prefix length for each interface as per Figure 1-3. The detailed configuration steps are omitted.

2) Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLDv1 on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] mld enable
[RouterA-GigabitEthernet1/0/1] pim ipv6 dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
```

```
[RouterA-GigabitEthernet1/0/2] pim ipv6 dm
[RouterA-GigabitEthernet1/0/2] quit
```

3) Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to this VLAN, and enable MLD Snooping and the function of dropping IPv6 unknown multicast traffic in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] mld-snooping drop-unknown
[SwitchA-vlan100] quit
```

Configure an IPv6 multicast group filter so that the hosts in VLAN 100 can join only the IPv6 multicast group FF1E::101.

```
[SwitchA] acl ipv6 number 2001
[SwitchA-acl6-basic-2001] rule permit source ff1e::101 128
[SwitchA-acl6-basic-2001] quit
[SwitchA] mld-snooping
[SwitchA–mld-snooping] group-policy 2001 vlan 100
[SwitchA–mld-snooping] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 as simulated hosts for IPv6 multicast group FF1E::101.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

4) Verify the configuration

View the detailed MLD Snooping multicast group information in VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 1 port.
            GE1/0/1               (D) ( 00:01:30 )
```

```
IP group(s):the following ip group(s) match to one mac group.
  IP group address:FF1E::101
    (::, FF1E::101):
      Attribute:    Host Port
      Host port(s):total 2 port.
        GE1/0/3                 (D) ( 00:03:23 )
        GE1/0/4                 (D) ( 00:04:10 )
MAC group(s):
  MAC group address:3333-0000-1001
      Host port(s):total 2 port.
        GE1/0/3
        GE1/0/4
```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A have joined IPv6 multicast group FF1E::101.

Static Port Configuration

Network requirements

- As shown in Figure 1-4, Router A connects to an IPv6 multicast source (Source) through GigabitEthernet 1/0/2, and to Switch A through GigabitEthernet 1/0/1.
- MLDv1 is to run on Router A, and MLDv1 Snooping is to run on Switch A, Switch B and Switch C, with Router A acting as the MLD querier.
- Host A and host C are permanent receivers of IPv6 multicast group FF1E::101. GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C are required to be configured as static member ports for multicast group 224.1.1.1 to enhance the reliability of multicast traffic transmission.
- Suppose STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked under normal conditions, and IPv6 multicast traffic flows to the receivers attached to Switch C only along the path of Switch A—Switch B—Switch C.
- It is required to configure GigabitEthernet 1/0/3 that connects Switch A to Switch C as a static router port, so that IPv6 multicast traffic can flow to the receivers nearly uninterruptedly along the path of Switch A—Switch C in the case that the path of Switch A—Switch B—Switch C gets blocked.

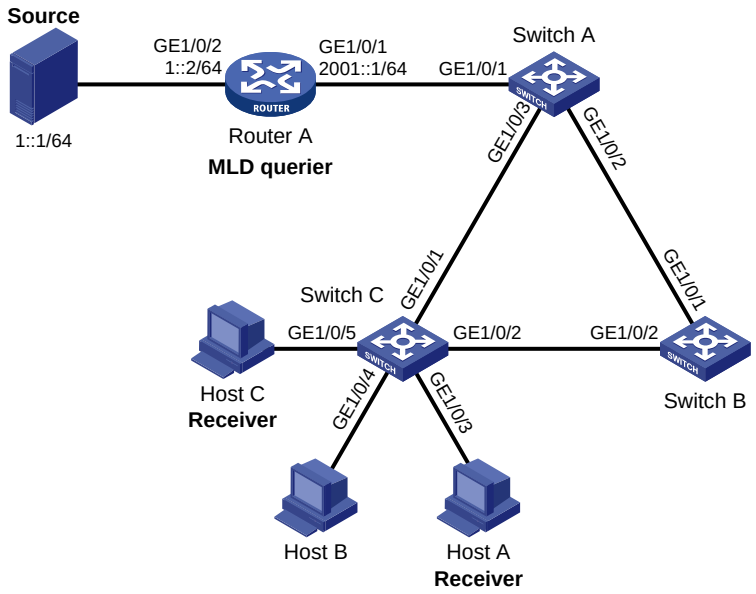
Note

If no static router port is configured, when the path of Switch A—Switch B—Switch C gets blocked, at least one MLD query-response cycle must be completed before the IPv6 multicast data can flow to the receivers along the new path of Switch A—Switch C, namely IPv6 multicast delivery will be interrupted during this process.

For details about the Spanning Tree Protocol (STP), refer to *MSTP Configuration* in the *Access Volume*.

Network diagram

Figure 1-4 Network diagram for static port configuration



Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses

Enable IPv6 forwarding and configure an IPv6 address and prefix length for each interface as per Figure 1-4.

2) Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLD on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] mld enable
[RouterA-GigabitEthernet1/0/1] pim ipv6 dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim ipv6 dm
[RouterA-GigabitEthernet1/0/2] quit
```

3) Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
[SwitchA-vlan100] mld-snooping enable
```

```
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 to be a static router port.

```
[SwitchA] interface gigabitethernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] mld-snooping static-router-port vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
```

4) Configure Switch B

Enable MLD Snooping globally.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 gigabitethernet 1/0/2
[SwitchB-vlan100] mld-snooping enable
[SwitchB-vlan100] quit
```

5) Configure Switch C

Enable MLD Snooping globally.

```
<SwitchC> system-view
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/5
[SwitchC-vlan100] mld-snooping enable
[SwitchC-vlan100] quit
```

Configure GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 as static member ports for IPv6 multicast group FF1E::101.

```
[SwitchC] interface GigabitEthernet 1/0/3
[SwitchC-GigabitEthernet1/0/3] mld-snooping static-group ff1e::101 vlan 100
[SwitchC-GigabitEthernet1/0/3] quit
[SwitchC] interface GigabitEthernet 1/0/5
[SwitchC-GigabitEthernet1/0/5] mld-snooping static-group ff1e::101 vlan 100
[SwitchC-GigabitEthernet1/0/5] quit
```

6) Verify the configuration

View the detailed MLD Snooping multicast group information in VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
```

```
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 2 port.
            GE1/0/1               (D) ( 00:01:30 )
            GE1/0/3               (S)
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:FF1E::101
        (::, FF1E::101):
          Attribute:    Host Port
          Host port(s):total 1 port.
            GE1/0/2               (D) ( 00:03:23 )
    MAC group(s):
      MAC group address:3333-0000-0101
          Host port(s):total 1 port.
            GE1/0/2
```

As shown above, GigabitEthernet 1/0/3 of Switch A has become a static router port.

View the detailed MLD Snooping multicast group information in VLAN 100 on Switch C.

```
[SwitchC] display mld-snooping group vlan 100 verbose
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).

  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):100.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 1 port.
            GE1/0/2               (D) ( 00:01:23 )
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:FF1E::101
        (::, FF1E::101):
          Attribute:    Host Port
          Host port(s):total 2 port.
            GE1/0/3               (S)
            GE1/0/5               (S)
    MAC group(s):
      MAC group address:3333-0000-0101
          Host port(s):total 2 port.
            GE1/0/3
            GE1/0/5
```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/5 on Switch C have become static member ports for IPv6 multicast group FF1E::101.

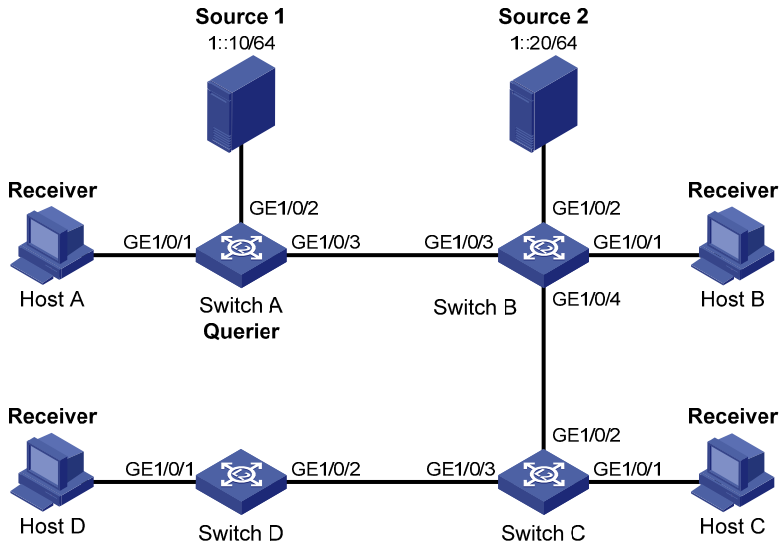
MLD Snooping Querier Configuration

Network requirements

- As shown in Figure 1-5, in a Layer-2-only network environment, two multicast sources Source 1 and Source 2 send IPv6 multicast data to multicast groups FF1E::101 and FF1E::102 respectively, Host A and Host C are receivers of multicast group FF1E::101, while Host B and Host D are receivers of multicast group FF1E::102.
- MLDv1 is enabled on all the receivers and MLDv1 Snooping is enabled on all the switches. Switch A, which is close to the multicast sources, is chosen as the MLD Snooping querier.
- To prevent flooding of unknown multicast traffic within the VLAN, it is required to configure all the switches to drop unknown multicast data packets.

Network diagram

Figure 1-5 Network diagram for MLD Snooping querier configuration



Configuration procedure

1) Configure Switch A

Enable IPv6 forwarding and enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100 and assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/3
```

Enable MLD Snooping and the function of dropping unknown IPv6 multicast data packets in VLAN 100.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] mld-snooping drop-unknown
```

Configure MLD Snooping querier feature in VLAN 100.

```
[SwitchA-vlan100] mld-snooping querier
[SwitchA-vlan100] quit
```

2) Configure Switch B

Enable IPv6 forwarding and enable MLD Snooping globally.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 100, add GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 into VLAN 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port gigabitethernet 1/0/1 to gigabitethernet 1/0/4
```

Enable the MLD Snooping feature and the function of dropping unknown IPv6 multicast data packets in VLAN 100.

```
[SwitchB-vlan100] mld-snooping enable
[SwitchB-vlan100] mld-snooping drop-unknown
[SwitchB-vlan100] quit
```

Configurations of Switch C and Switch D are similar to the configuration of Switch B.

3) Verify the configuration

When the MLD Snooping querier starts to work, all the switches but the querier receive MLD general queries. Use the **display mld-snooping statistics** command to view the statistics information of these MLD messages received.

View the MLD message statistics on Switch B.

```
[SwitchB-vlan100] display mld-snooping statistics
  Received MLD general queries:3.
  Received MLDv1 specific queries:0.
  Received MLDv1 reports:12.
  Received MLD dones:0.
  Sent     MLDv1 specific queries:0.
  Received MLDv2 reports:0.
  Received MLDv2 reports with right and wrong records:0.
  Received MLDv2 specific queries:0.
  Received MLDv2 specific sg queries:0.
  Sent     MLDv2 specific queries:0.
  Sent     MLDv2 specific sg queries:0.
  Received error MLD messages:0.
```

Troubleshooting MLD Snooping

Switch Fails in Layer 2 Multicast Forwarding

Symptom

A switch fails to implement Layer 2 multicast forwarding.

Analysis

MLD Snooping is not enabled.

Solution

1) Enter the **display current-configuration** command to view the running status of MLD Snooping.
2) If MLD Snooping is not enabled, use the **mld-snooping** command to enable MLD Snooping globally, and then use **mld-snooping enable** command to enable MLD Snooping in VLAN view.
3) If MLD Snooping is disabled only for the corresponding VLAN, just use the **mld-snooping enable** command in VLAN view to enable MLD Snooping in the corresponding VLAN.

Configured IPv6 Multicast Group Policy Fails to Take Effect

Symptom

Although an IPv6 multicast group policy has been configured to allow hosts to join specific IPv6 multicast groups, the hosts can still receive IPv6 multicast data addressed to other groups.

Analysis

- The IPv6 ACL rule is incorrectly configured.
- The IPv6 multicast group policy is not correctly applied.
- The function of dropping unknown IPv6 multicast data is not enabled, so unknown IPv6 multicast data is flooded.

Solution

1) Use the **display acl ipv6** command to check the configured IPv6 ACL rule. Make sure that the IPv6 ACL rule conforms to the IPv6 multicast group policy to be implemented.
2) Use the **display this** command in MLD Snooping view or the corresponding port view to check whether the correct IPv6 multicast group policy has been applied. If not, use the **group-policy** or **mld-snooping group-policy** command to apply the correct IPv6 multicast group policy.
3) Use the **display current-configuration** command to check whether the function of dropping unknown IPv6 multicast data is enabled. If not, use the **drop-unknown** or **mld-snooping drop-unknown** command to enable the function of dropping unknown IPv6 multicast data.

Table of Contents

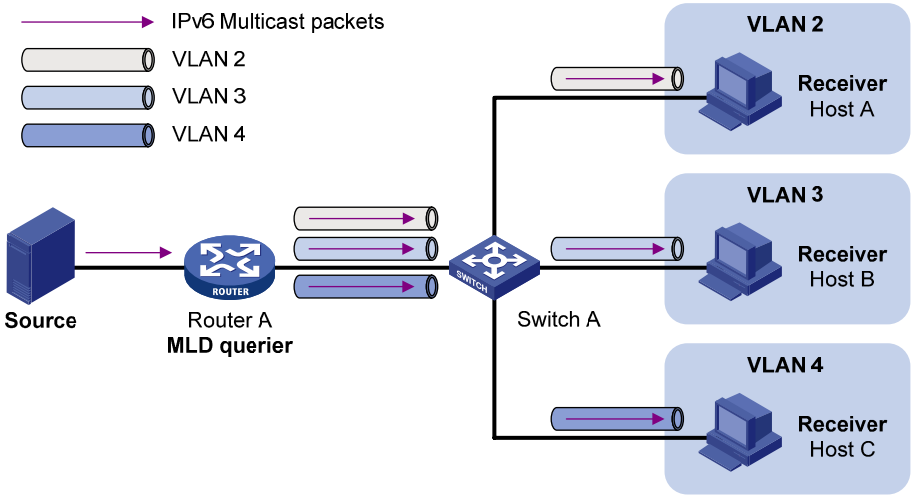
1 IPv6 Multicast VLAN Configuration

When configuring IPv6 multicast VLAN, go to these sections for information you are interested in:

Introduction to IPv6 Multicast VLAN

As shown in Figure 1-1, in the traditional IPv6 multicast programs-on-demand mode, when hosts, Host A, Host B and Host C, belonging to different VLANs require IPv6 multicast programs on demand service, the Layer 3 device, Router A, needs to forward a separate copy of the multicast traffic in each user VLAN to the Layer 2 device, Switch A. This results in not only waste of network bandwidth but also extra burden on the Layer 3 device.

Figure 1-1 Multicast transmission without IPv6 multicast VLAN



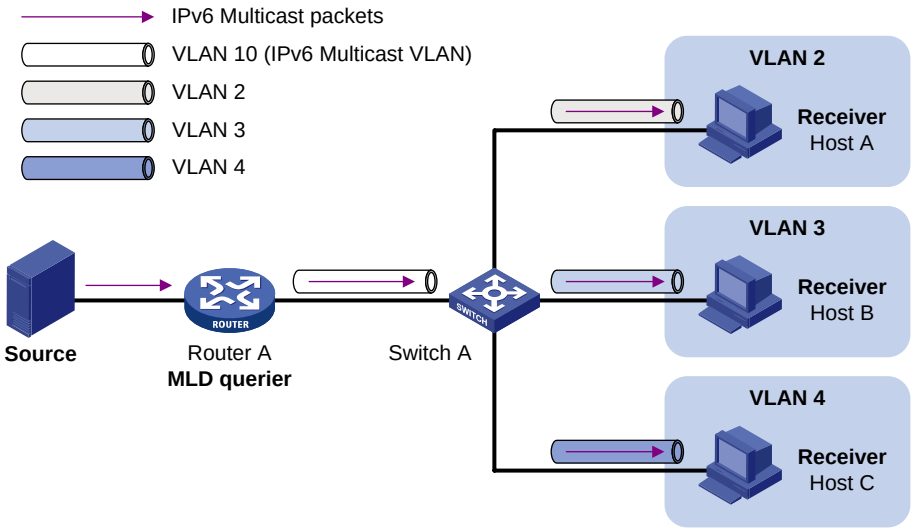
The IPv6 multicast VLAN feature configured on the Layer 2 device is the solution to this issue. With the IPv6 multicast VLAN feature, the Layer 3 device needs to replicate the multicast traffic only in the IPv6 multicast VLAN instead of making a separate copy of the multicast traffic in each user VLAN. This saves the network bandwidth and lessens the burden of the Layer 3 device.

The IPv6 multicast VLAN feature can be implemented in two approaches, as described below:

Sub-VLAN-based IPv6 multicast VLAN

As shown in Figure 1-2, Host A, Host B and Host C are in three different user VLANs. On Switch A, configure VLAN 10 as an IPv6 multicast VLAN, configure all the user VLANs as sub-VLANs of this IPv6 multicast VLAN, and enable MLD snooping in the IPv6 multicast VLAN.

Figure 1-2 Sub-VLAN-based IPv6 multicast VLAN

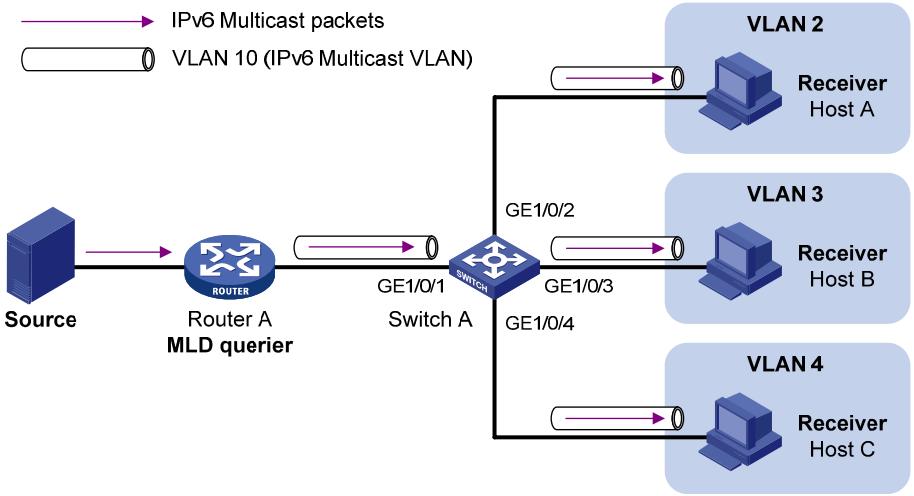


After the configuration, MLD snooping manages router ports in the IPv6 multicast VLAN and member ports in the sub-VLANs. When forwarding multicast data to Switch A, Router A needs to send only one copy of multicast traffic to Switch A in the IPv6 multicast VLAN, and Switch A distributes the traffic to the IPv6 multicast VLAN's sub-VLANs that contain receivers.

Port-based IPv6 multicast VLAN

As shown in Figure 1-3, Host A, Host B and Host C are in three different user VLANs. All the user ports are hybrid ports. On Switch A, configure VLAN 10 as an IPv6 multicast VLAN, assign all the user ports to this IPv6 multicast VLAN, and enable MLD Snooping in the IPv6 multicast VLAN and all the user VLANs.

Figure 1-3 Port-based IPv6 multicast VLAN



After the configuration, upon receiving an MLD message on a user port, Switch A tags the message with the IPv6 multicast VLAN ID and relays it to the MLD querier, so that MLD Snooping can uniformly manage the router ports and member ports in the IPv6 multicast VLAN. When forwarding multicast data to Switch A, Router A needs to send only one copy of multicast traffic to Switch A in the IPv6 multicast VLAN, and Switch A distributes the traffic to all the member ports in the IPv6 multicast VLAN.

Note

- For information about MLD Snooping, router ports, and member ports, refer to *MLD Snooping Configuration* in the *IP Multicast Volume*.
- For information about VLAN tags, refer to *VLAN Configuration* in the *Access Volume*.

IPv6 Multicast VLAN Configuration Task List

Complete the following tasks to configure IPv6 multicast VLAN:

Configuration task		Remarks
Configuring IPv6 Sub-VLAN-Based IPv6 Multicast VLAN		Required Use either approach.
Configuring Port-Based IPv6 Multicast VLAN	Configuring User Port Attributes	
	Configuring IPv6 Multicast VLAN Ports	

Note

If you have configured both sub-VLAN-based IPv6 multicast VLAN and port-based IPv6 multicast VLAN on a device, the port-based IPv6 multicast VLAN configuration is given preference.

Configuring IPv6 Sub-VLAN-Based IPv6 Multicast VLAN

Configuration Prerequisites

Before configuring sub-VLAN-based IPv6 multicast VLAN, complete the following tasks:

- Create VLANs as required
- Enable MLD Snooping in the VLAN to be configured as an IPv6 multicast VLAN

Configuring Sub-VLAN-Based IPv6 Multicast VLAN

In this approach, you configure a VLAN as an IPv6 multicast VLAN, and configure user VLANs as sub-VLANs of the IPv6 multicast VLAN.

Follow these steps to configure sub-VLAN-based IPv6 multicast VLAN:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view	**multicast-vlan ipv6** *vlan-id*	Required No IPv6 multicast VLAN configured by default

To do…	Use the command…	Remarks
Configure the specified VLAN(s) as sub-VLAN(s) of the IPv6 multicast VLAN	**subvlan** *vlan-list*	Required By default, an IPv6 multicast VLAN has no sub-VLANs.

Note

- You cannot configure IPv6 multicast VLAN on a device with IP multicast routing enabled.
- The VLAN to be configured as an IPv6 multicast VLAN must exist.
- The VLANs to be configured as the sub-VLANs of the IPv6 multicast VLAN must exist and must not be sub-VLANs of another IPv6 multicast VLAN.
- The total number of sub-VLANs of an IPv6 multicast VLAN must not exceed 127.

Configuring Port-Based IPv6 Multicast VLAN

When configuring port-based IPv6 multicast VLAN, you need to configure the attributes of each user port and then assign the ports to the IPv6 multicast VLAN.

Note

- A user port can be configured as a multicast VLAN port only if it is of the Ethernet or Layer 2 aggregate port type.
- Configurations made in Ethernet port view are effective only for the current port; configurations made in Layer 2 aggregate port view are effective only for the current port; configurations made in port group view are effective for all the ports in the current port group.

Configuration Prerequisites

Before configuring port-based IPv6 multicast VLAN, complete the following tasks:

- Create VLANs as required
- Enable MLD Snooping in the VLAN to be configured as an IPv6 multicast VLAN
- Enable MLD Snooping in all the user VLANs

Configuring User Port Attributes

Configure the user ports as hybrid ports to permit packets of the specified user VLAN to pass and configure the user VLAN to which the user ports belong as the default VLAN.

Configure the user ports to permit packets of the IPv6 multicast VLAN to pass and untag the packets. Thus, upon receiving multicast packets tagged with the IPv6 multicast VLAN ID from the upstream device, the Layer 2 device untags the multicast packets and forwards them to its downstream device.

Follow these steps to configure user port attributes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Enter port view or port group view	**interface** *interface-type interface-number*	Required Use either approach.
	port-group manual *port-group-name*	
Configue the user port link type as hybrid	**port link-type hybrid**	Required Access by default
Specify the user VLAN that comprises the current user port(s) as the default VLAN	**port hybrid pvid vlan** *vlan-id*	Required VLAN 1 by default
Configure the current user ports to permit packets of the specified IPv6 multicast VLAN to pass and untag the packets	**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required By default, a hybrid port permits only packets of VLAN 1 to pass.

Note

For details about the **port link-type**, **port hybrid pvid vlan**, and **port hybrid vlan** commands, refer to VLAN Commands in the Access Volume.

Configuring IPv6 Multicast VLAN Ports

In this approach, you need to configure a VLAN as an IPv6 multicast VLAN and then assign user ports to this IPv6 multicast VLAN by either adding the user ports in the IPv6 multicast VLAN or specifying the IPv6 multicast VLAN on the user ports. These two methods give the same result.

Configure IPv6 multicast VLAN ports in IPv6 multicast VLAN view

Follow these steps to configure IPv6 multicast VLAN ports in IPv6 multicast VLAN view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view	**multicast-vlan ipv6** *vlan-id*	Required No IPv6 multicast VLAN configured by default
Assign port(s) to the IPv6 multicast VLAN	**port** *interface-list*	Required By default, an IPv6 multicast VLAN has no ports.

Configure IPv6 multicast VLAN ports in terface view or port group view

Follow these steps to configure IPv6 multicast VLAN ports in port view or port group view:

To do…	Use this command…	Remarks
Enter system view	**system-view**	—
Configure the specified VLAN as an IPv6 multicast VLAN and enter IPv6 multicast VLAN view	**multicast-vlan ipv6** *vlan-id*	Required Not an IPv6 multicast VLAN by default.
Return to system view	**quit**	—
Enter port view or port group view	**interface** *interface-type interface-number*	Required Use either command.
	port-group manual *port-group-name*	
Configure the port(s) as port(s) of the IPv6 muticast VLAN	**port multicast-vlan ipv6** *vlan-id*	Required By default, a user port does not belong to any IPv6 multicast VLAN.

Note

- You cannot configure IPv6 multicast VLAN on a device with multicast routing enabled.
- The VLAN to be configured as an IPv6 multicast VLAN must exist.
- A port can belong to only one IPv6 multicast VLAN.

Displaying and Maintaining IPv6 Multicast VLAN

To do…	Use the command…	Remarks
Display information about an IPv6 multicast VLAN	**display multicast-vlan ipv6** [*vlan-id*]	Available in any view

IPv6 Multicast VLAN Configuration Examples

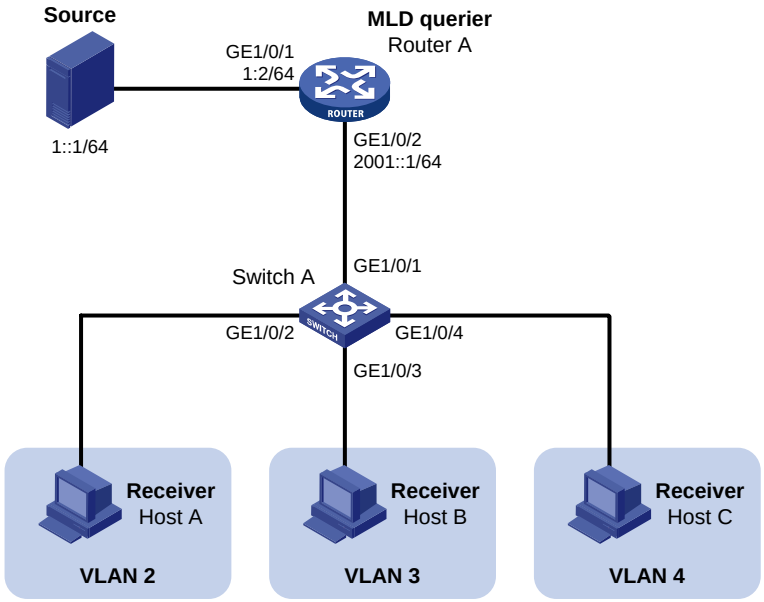
Sub-VLAN-Based Multicast VLAN Configuration Example

Network requirements

- As shown in Figure 1-4, Router A connects to an IPv6 multicast source through GigabitEthernet 1/0/1 and to Switch A, through GigabitEthernet 1/0/2.
- MLDv1 is required on Router A, and MLD Snooping is required on Switch A. Router A is the MLD querier.
- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 10, GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 belong to VLAN 2 through VLAN 4 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 of Switch A.

- The IPv6 multicast source sends IPv6 multicast data to the IPv6 multicast group FF1E::101. Host A, Host B, and Host C are receivers of the IPv6 multicast group.
- Configure the sub-VLAN-based IPv6 multicast VLAN feature so that Router A just sends IPv6 multicast data to Switch A through the IPv6 multicast VLAN and Switch A forwards the traffic to the receivers that belong to different user VLANs.

Figure 1-4 Network diagram for sub-VLAN-based IPv6 multicast VLAN configuration



Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses

Enable IPv6 forwarding on each device and configure an IPv6 address and address prefix for each interface as per Figure 1-4. The detailed configuration steps are omitted here.

2) Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface and enable MLD on the host-side interface GigabitEthernet 1/0/2.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] pim ipv6 dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim ipv6 dm
[RouterA-GigabitEthernet1/0/2] mld enable
```

3) Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 2 and assign GigabitEthernet 1/0/2 to this VLAN.

```
[SwitchA] vlan 2
```

```
[SwitchA-vlan2] port gigabitethernet 1/0/2
[SwitchA-vlan2] quit
```

The configuration for VLAN 3 and VLAN 4 is similar to the configuration for VLAN 2.

Create VLAN 10, assign GigabitEthernet 1/0/1 to this VLAN and enable MLD Snooping in the VLAN.

```
[SwitchA] vlan 10
[SwitchA-vlan10] port gigabitethernet 1/0/1
[SwitchA-vlan10] mld-snooping enable
[SwitchA-vlan10] quit
```

Configure VLAN 10 as an IPv6 multicast VLAN and configure VLAN 2 through VLAN 4 as its sub-VLANs.

```
[SwitchA] multicast-vlan ipv6 10
[SwitchA-ipv6-mvlan-10] subvlan 2 to 4
[SwitchA-ipv6-mvlan-10] quit
```

4) Verify the configuration

Display information about the IPv6 multicast VLAN.

```
[SwitchA] display multicast-vlan ipv6
 Total 1 IPv6 multicast-vlan(s)
 IPv6 Multicast vlan 10
   subvlan list:
    vlan 2-4
   port list:
    no port
```

View the MLD Snooping IPv6 multicast group information on Switch A.

```
[SwitchA] display mld-snooping group
  Total 4 IP Group(s).
  Total 4 IP Source(s).
  Total 4 MAC Group(s).
  Port flags: D-Dynamic port, S-Static port, C-Copy port
  Subvlan flags: R-Real VLAN, C-Copy VLAN
  Vlan(id):2.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
    Router port(s):total 0 port.
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:FF1E::101
        (::, FF1E::101):
          Host port(s):total 1 port.
            GE1/0/2               (D)
    MAC group(s):
      MAC group address:3333-0000-0101
          Host port(s):total 1 port.
            GE1/0/2
  Vlan(id):3.
    Total 1 IP Group(s).
    Total 1 IP Source(s).
```

```
  Total 1 MAC Group(s).
  Router port(s):total 0 port.
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:FF1E::101
      (::, FF1E::101):
        Host port(s):total 1 port.
          GE1/0/3               (D)
  MAC group(s):
    MAC group address:3333-0000-0101
        Host port(s):total 1 port.
          GE1/0/3
Vlan(id):4.
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).
  Router port(s):total 0 port.
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:FF1E::101
      (::, FF1E::101):
        Host port(s):total 1 port.
          GE1/0/4               (D)
  MAC group(s):
    MAC group address:3333-0000-0101
        Host port(s):total 1 port.
          GE1/0/4
Vlan(id):10.
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).
  Router port(s):total 1 port.
          GE1/0/1               (D)
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:FF1E::101
      (::, FF1E::101):
        Host port(s):total 0 port.
  MAC group(s):
    MAC group address:3333-0000-0101
        Host port(s):total 0 port.
```

As shown above, MLD Snooping is maintaining the router port in the IPv6 multicast VLAN (VLAN 10) and the member ports in the sub-VLANs (VLAN 2 through VLAN 4).

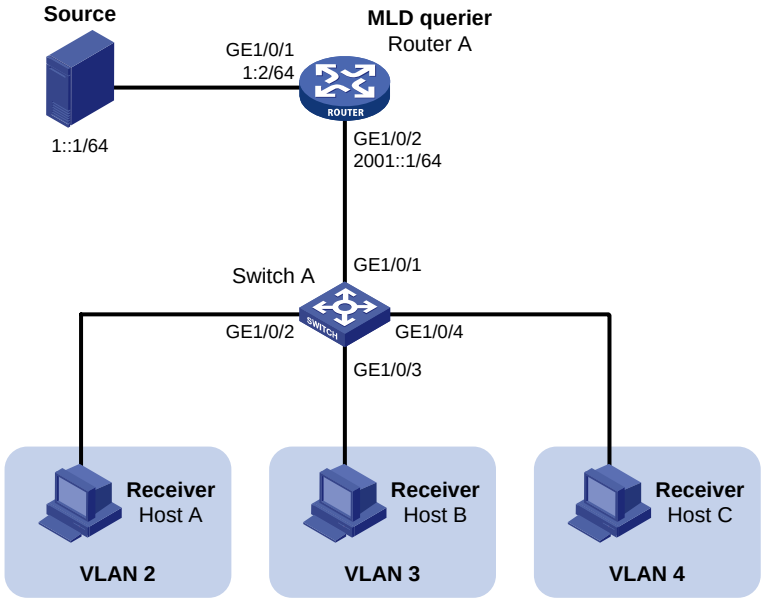
Port-Based Multicast VLAN Configuration Example

Network requirements

- As shown in Figure 1-5, Router A connects to an IPv6 multicast source (Source) through GigabitEthernet 1/0/1, and to Switch A through GigabitEthernet 1/0/2.
- MLDv1 is required on Router A. MLDv1 Snooping is required on Switch A. Router A acts as the MLD querier.

- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 10, GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 belong to VLAN 2 through VLAN 4 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 of Switch A.
- The IPv6 multicast source sends IPv6 multicast data to IPv6 multicast group FF1E::101. Host A, Host B, and Host C are receivers of the IPv6 multicast group.
- Configure the port-based IPv6 multicast VLAN feature so that Router A just sends IPv6 multicast data to Switch A through the IPv6 multicast VLAN and Switch A forward the IPv6 multicast data to the receivers that belong to different user VLANs.

Figure 1-5 Network diagram for port-based IPv6 multicast VLAN configuration



Configuration procedure

1) Enable IPv6 forwarding and configure IPv6 addresses

Enable IPv6 forwarding on each device and configure the IPv6 address and address prefix for each interface as per Figure 1-5. The detailed configuration steps are omitted here.

2) Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLD on the host-side interface GigabitEthernet 1/0/2.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface gigabitethernet 1/0/1
[RouterA-GigabitEthernet1/0/1] ipv6 pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface gigabitethernet 1/0/2
[RouterA-GigabitEthernet1/0/2] ipv6 pim dm
[RouterA-GigabitEthernet1/0/2] mld enable
```

3) Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 10, assign GigabitEthernet 1/0/1 to VLAN 10, and enable MLD Snooping in this VLAN.

```
[SwitchA] vlan 10
[SwitchA-vlan10] port gigabitethernet 1/0/1
[SwitchA-vlan10] mld-snooping enable
[SwitchA-vlan10] quit
```

Create VLAN 2 and enable MLD Snooping in the VLAN.

```
[SwitchA] vlan 2
[SwitchA-vlan2] mld-snooping enable
[SwitchA-vlan2] quit
```

The configuration for VLAN 3 and VLAN 4 is similar. The detailed configuration steps are omitted.

Configure GigabitEthernet 1/0/2 as a hybrid port. Configure VLAN 2 as the default VLAN. Configue GigabitEthernet 1/0/2 to permit packets of VLAN 2 to pass and untag the packets when forwarding them.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-type hybrid
[SwitchA-GigabitEthernet1/0/2] port hybrid pvid vlan 2
[SwitchA-GigabitEthernet1/0/2] port hybrid vlan 2 untagged
[SwitchA-GigabitEthernet1/0/2] port hybrid vlan 10 untagged
[SwitchA-GigabitEthernet1/0/2] quit
```

The configuration for GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 is similar. The detailed configuration steps are omitted.

Configure VLAN 10 as an IPv6 multicast VLAN.

```
[SwitchA] multicast-vlan ipv6 10
```

Assign GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 to IPv6 multicast VLAN 10.

```
[SwitchA-ipv6-mvlan-10] port gigabitethernet 1/0/2 to gigabitethernet 1/0/3
[SwitchA-ipv6-mvlan-10] quit
```

Assign GigabitEthernet 1/0/4 to IPv6 multicast VLAN 10.

```
[SwitchA] interface gigabitethernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] port multicast-vlan ipv6 10
[SwitchA-GigabitEthernet1/0/4] quit
```

4) Verify the configuration

View the IPv6 multicast VLAN information on Switch A.

```
[SwitchA] display multicast-vlan ipv6
 Total 1 IPv6 multicast-vlan(s)
 IPv6 Multicast vlan 10
   subvlan list:
    no subvlan
   port list:
    GE1/0/2                GE1/0/3                GE1/0/4
```

View the MLD Snooping multicast group information on Switch A.

```
[SwitchA] display mld-snooping group
  Total 1 IP Group(s).
  Total 1 IP Source(s).
```

```
Total 1 MAC Group(s).

Port flags: D-Dynamic port, S-Static port, C-Copy port
Subvlan flags: R-Real VLAN, C-Copy VLAN
Vlan(id):10.
  Total 1 IP Group(s).
  Total 1 IP Source(s).
  Total 1 MAC Group(s).
  Router port(s):total 1 port.
          GE1/0/1                (D)
  IP group(s):the following ip group(s) match to one mac group.
    IP group address:FF1E::101
      (::, FF1E::101):
        Host port(s):total 3 port.
          GE1/0/2                (D)
          GE1/0/3                (D)
          GE1/0/4                (D)
  MAC group(s):
    MAC group address:3333-0000-0101
        Host port(s):total 3 port.
          GE1/0/2
          GE1/0/3
          GE1/0/4
```

As shown above, MLD Snooping is maintaining router ports and member ports in VLAN 10.

QoS Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The QoS Volume is organized as follows:

Features	Description
QoS	This document describes: • QoS overview • Traffic classification configuration • Traffic policing Configuration • Traffic shaping Configuration • Line rate configuration • QoS policy configuration • Congestion management • Congestion avoidance configuration • Priority mapping configuration • Traffic mirroring configuration
User Profile	User profile provides a configuration template to save predefined configurations. This document describes: • Creating a User Profile • Configuring a User Profile • Enabling a User Profile

Table of Contents

1 QoS Overview

This chapter covers these topics:

Introduction

Quality of Service (QoS) is a concept concerning service demand and supply. It reflects the ability to meet customer needs. Generally, QoS focuses on improving services under certain conditions rather than grading services precisely.

In an internet, QoS evaluates the ability of the network to forward packets using different services. The evaluation can be based on different criteria because the network may provide various services. Generally, QoS refers to the ability to provide improved service by solving the core issues such as delay, jitter, and packet loss ratio in the packet forwarding process.

Traditional Packet Forwarding Service

On traditional IP networks, devices treat all packets equally and handle them using the first in first out (FIFO) policy. All packets share the resources of the network and devices. How many resources the packets can obtain completely depends on the time they arrive. This service is called best-effort. It delivers packets to their destinations as possibly as it can, without any guarantee for delay, jitter, packet loss ratio, reliability and so on.

This service policy is only suitable for applications insensitive to bandwidth and delay, such as WWW, file transfer and e-mail.

New Requirements from Emerging Services

The Internet has been growing along with the fast development of networking technologies. More and more people use the Internet to transmit data, share video and do a lot of other things.

Besides traditional applications such as WWW, e-mail and FTP, network users are enjoying new services such as tele-education, telemedicine, video telephone, videoconference and Video-on-Demand (VoD). Enterprise users expect to connect their regional branches together with VPN technologies to carry out operational applications, for instance, to access the database of the company or to monitor remote devices through Telnet.

These new applications have one thing in common, that is, they all have special requirements for bandwidth, delay, and jitter. For example, videoconference and VoD require high bandwidth, low delay and jitter. As for mission-critical applications, such as transactions and Telnet, they may not require high bandwidth but do require low delay and preferential service during congestion.

The emerging applications demand higher service performance of IP networks. Better network services during packets forwarding are required, such as providing dedicated bandwidth, reducing packet loss ratio, managing and avoiding congestion, regulating network traffic, and setting the precedence of packets. To meet these requirements, a network must provide more improved services.

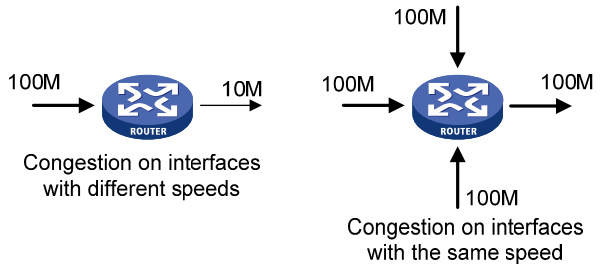
Congestion: Causes, Impacts, and Countermeasures

Network congestion is a major factor degrading the service quality of a traditional network. Congestion is a situation where the forwarding rate decreases due to insufficient resources, resulting in extra delay.

Causes

Congestion easily occurs in complex packet switching circumstances in the Internet. The following figure shows two common cases:

Figure 1-1 Traffic congestion causes



- The traffic enters a device from a high speed link and is forwarded over a low speed link.
- The packet flows enter a device from several interfaces at the same rate and are forwarded out an interface at the same rate as well.

When traffic arrives at the line speed, a bottleneck will be created at the outgoing interface causing congestion.

Besides bandwidth bottlenecks, congestion can be caused by resource shortage in various forms such as insufficient processor time, buffer, and memory, and by network resource exhaustion resulting from excessive arriving traffic in certain periods.

Impacts

Congestion may bring these negative results:

- Increased delay and jitter during packet transmission
- Decreased network throughput and resource use efficiency
- Network resource (memory in particular) exhaustion and even system breakdown

It is obvious that congestion hinders resource assignment for traffic and thus degrades service performance. The chance of congestion is high in switched networks and multi-user application environments. To improve the service performance of your network, you must address the congestion issues.

Countermeasures

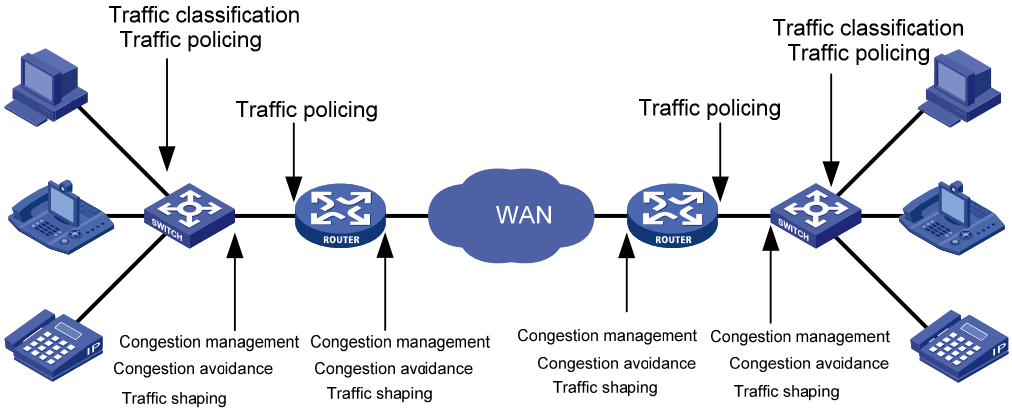
A simple solution for congestion is to increase network bandwidth. However, it cannot address all the problems of congestion.

A more effective solution is to provide differentiated services for different applications through traffic control and resource allocation. In this way, resources can be used more properly. During resources allocation and traffic control, the direct or indirect factors that might cause network congestion should be controlled to reduce the probability of congestion. Once congestion occurs, resource allocation should be performed according to the characteristics and demands of applications to minimize the effects of congestion on QoS.

Major Traffic Management Techniques

End-to-end QoS model

Figure 1-2 End-to-end QoS model



As shown in Figure 1-2, traffic classification, traffic policing, traffic shaping, congestion management, and congestion avoidance are the foundations for a network to provide differentiated services. Mainly they implement the following functions:

- Traffic classification uses certain match criteria to organize packets with different characteristics into different classes, and is the prerequisite for providing differentiated services. Traffic classification is usually applied in the inbound direction of a port.
- Traffic policing polices particular flows entering a device according to configured specifications and is usually applied in the inbound direction of a port. When a flow exceeds the specification, some restriction or punishment measures can be taken to prevent overconsumption of network resources and protect the commercial benefits of the carrier.
- Traffic shaping proactively adjusts the output rate of traffic to adapt traffic to the network resources of the downstream device and avoid unnecessary packet drop and congestion. Traffic shaping is usually applied in the outbound direction of a port.
- Congestion management provides measures for handling resource competition during network congestion and is usually applied in the outbound direction of a port. Generally, it stores packets in queues, and then uses a scheduling algorithm to arrange the forwarding sequence of the packets.
- Congestion avoidance monitors the usage status of network resources and is usually applied in the outbound direction of a port. As congestion becomes worse, it actively reduces the amount of traffic by dropping packets.

Among these traffic management technologies, traffic classification is the basis for providing differentiated services by classifying packets with certain match criteria. Traffic policing, traffic shaping, congestion management, and congestion avoidance manage network traffic and resources in different ways to realize differentiated services.

This section is focused on traffic classification, and the subsequent sections will introduce the other technologies in details.

Traffic Classification

Traffic classification organizes packets with different characteristics into different classes using match criteria. It is the basis for providing differentiated services.

You can define match criteria based on the IP precedence bits in the type of service (ToS) field of the IP packet header, or based on other header information such as IP addresses, MAC addresses, IP protocol field, and port numbers. Contents other than the header information in packets are rarely used for traffic classification. You can define a class for packets with a common quintuple (source address, source port number, protocol number, destination address and destination port number), or for all packets to a certain network segment.

When packets are classified at network boundaries, the precedence bits in the ToS field of the IP packet header are generally re-set. In this way, IP precedence can be adopted as a classification criterion for the packets in the network. IP precedence can also be used in queuing to prioritize traffic. The downstream network can either inherit the classification results from its upstream network or re-classify the packets according to its own criteria.

To provide differentiated services, traffic classes must be associated with certain traffic control actions or resource allocation actions. What traffic control actions should be adopted depends on the current phase and the resources of the network. For example, CIR is adopted to police packets when they enter the network; generic traffic shaping (GTS) is performed on packets when they flow out of the node; queue scheduling is performed when congestion happens; congestion avoidance measures are taken when the congestion deteriorates.

2 QoS Policy Configuration

When configuring QoS policy, go to these sections for information that you are interested in:

Overview

QoS policy includes the following three elements: class, traffic behavior and policy. You can bind the specified class to the specified traffic behavior through QoS policies to facilitate the QoS configuration.

Class

Class is used for identifying traffic.

The elements of a class include the class name and classification rules.

You can use commands to define a series of rules to classify packets. Additionally, you can use commands to define the relationship among classification rules: “**and”** and “**or”**.

- **and**: The devices considers a packet to be of a specific class when the packet matches all the specified classification rules.
- **or**: The device considers a packet be of a specific class when the packet matches one of the specified classification rules.

Traffic behavior

Traffic behavior is used to define all the QoS actions performed on packets.

The elements of a QoS behavior include traffic behavior name and actions defined in traffic behavior.

You can use commands to define multiple actions in a traffic behavior.

Policy

Policy is used to bind the specified class to the specified traffic behavior.

The elements of a policy include the policy name and the name of the classification-to-behavior binding.

Configuring a QoS Policy

The procedure for configuring QoS policy is as follows:

1) Define a class and define a group of traffic classification rules in class view.
2) Define a traffic behavior and define a group of QoS actions in traffic behavior view.
3) Define a policy and specify a traffic behavior corresponding to the class in policy view.

Defining a Class

To define a class, you need to create a class and then define rules in the corresponding class view.

Configuration procedure

Follow these steps to define a class:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a class and enter the corresponding class view	**traffic classifier** *classifier-name* [**operator** { **and** | **or** }]	Required By default, the **and** keyword is specified. That is, the relation between the rules in the class view is logic AND. This operation leads you to class view.
Define a rule used to match packets	**if-match** *match-criteria*	Required

match-criteria: Matching rules to be defined for a class. Table 2-1 describes the available forms of this argument.

Table 2-1 The form of the *match-criteria* argument

Form	Description
acl *access-list-number*	Specifies an ACL to match packets. The access-list-number argument is in the range 2000 to 4999. In a class configured with the operator **and**, the logical relationship between rules defined in the referenced IPv4 ACL is **or**.
acl ipv6 *access-list-number*	Specifies an IPv6 ACL to match IPv6 packets. The access-list-number argument is in the range 2000 to 3999. In a class configured with the operator **and**, the logical relationship between rules defined in the referenced IPv6 ACL is **or**.
any	Specifies to match all packets.
customer-dot1p *8021p-list*	Specifies to match packets by 802.1p precedence of the customer network. The *8021p-list* argument is a list of CoS values, in the range of 0 to 7. **Note** *Even though you can provide up to eight space-separated CoS values for this argument, the Switch 4800G supports only one CoS value in a rule. If you configure multiple CoS values in a rule, the rule cannot be issued.*
customer-vlan-id *vlan-id-list*	Specifies to match the packets of specified VLANs of user networks. The *vlan-id-list* argument specifies a list of VLAN IDs, in the form of *vlan-id* **to** *vlan-id* or multiple discontinuous VLAN IDs (separated by space). You can specify up to eight VLAN IDs for this argument at a time. VLAN ID is in the range 1 to 4094. In a class configured with the operator **and**, the logical relationship between the customer VLAN IDs specified for the **customer-vlan-id** keyword is **or**.
destination-mac *mac-address*	Specifies to match the packets with a specified destination MAC address.

Form	Description
dscp *dscp-list*	Specifies to match packets by DSCP precedence. The *dscp-list* argument is a list of DSCP values in the range of 0 to 63. **Note** *Even though you can provide up to eight space-separated DSCP values for this argument, the Switch 4800G supports only one DSCP value in a rule. If you configure multiple DSCP values in a rule, the rule cannot be issued.*
ip-precedence *ip-precedence-list*	Specifies to match packets by IP precedence. The *ip-precedence-list* argument is a list of IP precedence values in the range of 0 to 7. **Note** *Even though you can provide up to eight space-separated IP precedence values for this argument, the Switch 4800G supports only one IP precedence value in a rule. If you configure multiple IP precedence values in a rule, the rule cannot be issued.*
protocol *protocol-name*	Specifies to match the packets of a specified protocol. The *protocol-name* argument can be IP or IPv6.
service-dot1p *8021p-list*	Specifies to match packets by 802.1p precedence of the service provider network. The *8021p-list* argument is a list of CoS values in the range of 0 to 7. **Note** Even though you can provide up to eight space-separated CoS values for this argument, the *Switch 4800G* supports only one CoS value in a rule. If you configure multiple CoS values in a rule, the rule cannot be issued.
service-vlan-id *vlan-id-list*	Specifies to match the packets of the VLANs of the operator's network. The *vlan-id-list* argument is a list of VLAN IDs, in the form of *vlan-id* **to** *vlan-id* or multiple discontinuous VLAN IDs (separated by space). You can specify up to eight VLAN IDs for this argument at a time. VLAN ID is in the range of 1 to 4094. In a class configured with the operator **and**, the logical relationship between the service VLAN IDs specified for the **service-vlan-id** keyword is **or**.
source-mac *mac-address*	Specifies to match the packets with a specified source MAC address.

Note

Suppose the logical relationship between classification rules is **and**. Note the following when using the **if-match** command to define matching rules.

- If multiple matching rules with the **acl** or **acl ipv6** keyword specified are defined in a class, the actual logical relationship between these rules is **or** when the policy is applied.
- If multiple matching rules with the **customer-vlan-id** or **service-vlan-id** keyword specified are defined in a class, the actual logical relationship between these rules is **or**.

Configuration example

1) Network requirements

Configure a class named test to match the packets with their IP precedence being 6.

2) Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create the class. (This operation leads you to class view.)

```
[Sysname] traffic classifier test
```

Define the classification rule.

```
[Sysname-classifier-test] if-match ip-precedence 6
```

Defining a Traffic Behavior

To define a traffic behavior, you need to create a traffic behavior and then configure attributes for it in traffic behavior view.

Configuration procedure

Follow these steps to define a traffic behavior:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a traffic behavior and enter the corresponding traffic behavior view	**traffic behavior** *behavior-name*	Required *behavior-name*: Behavior name. This operation leads you to traffic behavior view

To do…	Use the command…	Remarks
Configure accounting action	**accounting**	Required You can configure the traffic behavior as required.
Configure traffic policing action	**car cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**pir** *peak-information-rate*] [**green** *action*] [**red** *action*] [**yellow** *action*]	
Configure traffic filtering behavior	**filter** { **deny** | **permit** }	
Configure traffic mirroring action	**mirror-to** { **cpu** | **interface** *interface-type interface-number* }	
Configure nested VLAN tag action	**nest top-most vlan-id** *vlan-id*	
Configure traffic redirecting action	**redirect** { **cpu** | **interface** *interface-type interface-number* | **next-hop** { *ipv4-add* [*ipv4-add*] | *ipv6-add* [*interface-type interface-number*] [*ipv6-add* [*interface-type interface-number*]] } }	
Remark the customer network VLAN ID for packets	**remark customer-vlan-id** *vlan-id-value*	
Remark DSCP value for packets	**remark dscp** *dscp-value*	
Remark 802.1p priority for packets	**remark dot1p** *8021p*	
Remark drop precedence for packets	**remark drop-precedence** *drop-precedence-value*	
Remark IP precedence for packets	**remark ip-precedence** *ip-precedence-value*	
Remark local precedence for packets	**remark local-precedence** *local-precedence*	
Remark the service provider network VLAN ID for packets	**remark service-vlan-id** *vlan-id-value*	

Configuration example

1) Network requirements

Create a traffic behavior named test, configuring traffic policing action for it, with the CAR being 640 kbps.

2) Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create the traffic behavior (This operation leads you to traffic behavior view).

```
[Sysname] traffic behavior test
```

Configure traffic policing action for the traffic behavior.

```
[Sysname-behavior-test] car cir 640
```

Defining a Policy

A policy associates a class with a traffic behavior. Each traffic behavior is comprised of a group of QoS actions. A device executes these QoS actions in the order they are defined.

Follow these steps to associate a traffic behavior with a class:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a policy (This operation leads you to policy view)	**qos policy** *policy-name*	—
Specify the traffic behavior for a class	**classifier** *classifier-name* **behavior** *behavior-name* [**mode do1q-tag-manipulation**]	Required

Note

In a QoS policy with multiple class-to-traffic-behavior associations, if the action of creating an outer VLAN tag, the action of setting customer network VLAN ID, or the action of setting service provider network VLAN ID is configured in a traffic behavior, we recommend you not to configure any other action in this traffic behavior. Otherwise, the QoS policy may not function as expected after it is applied.

Applying a Policy

You can apply a QoS policy in different views as follows:

- In port or port group view, the policy applies to the inbound or outbound direction of an interface or a group of interfaces;
- In user profile view, the policy applies to the traffic sent or received by the online users;
- In VLAN view, the policy applies to the inbound or outbound direction of a VLAN;
- In system view, the policy applies to the inbound or outbound direction of all ports globally.

Note

- You cannot modify the classification rules, traffic behaviors, and classifier-behavior associations in a QoS policy already applied. To check whether a QoS policy has been applied successfully, use the **display qos policy global** command and the **display qos policy interface** command.
- The switch may save the applications of some QoS policies that have failed to be applied due to insufficient hardware resources in the configuration file. After the switch reboots, these policies may preempt other user configurations for resources, resulting in loss of configurations. Suppose that the **user-bind** command is configured on GigabitEthernet 1/0/2, and the application of a QoS policy to GigabitEthernet 1/0/1 is saved in the configuration file even though the application has failed due to insufficient resources. After the switch reboots, it may assign resources to have the QoS policy take effect preferentially, while the **user-bind** configuration may be lost due to insufficient resources.

Applying a QoS policy to a port/port group

A policy can be applied to multiple ports. Only one policy can be applied in one direction (inbound or outbound) of a port/port group.

Follow these steps to apply the QoS policy to a port/port group:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Apply an associated policy		**qos apply policy** *policy-name* { **inbound** | **outbound** }	Required

Note

If a QoS policy is applied in the outbound direction of an interface, the QoS policy cannot influence local packets (local packets refer to the important protocol packets that maintain the normal operation of the device. QoS must not process such packets to avoid packet drop. Commonly used local packets are: link maintenance packets, IS-IS packets, OSPF packets, RIP packets, BGP packets, LDP packets, RSVP packets, and SSH packets and so on.)

Applying a QoS policy to online users

You can apply a QoS policy to traffic of multiple online users. You can apply only one policy in one direction (inbound or outbound) of the traffic of online users. To modify a QoS policy already applied, remove the QoS policy application first.

Follow these steps to apply a QoS policy to traffic of online users:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter user profile view	**user-profile** *profile-name* [**dot1x** | **portal**]	Required The configuration made in user profile view takes effect when the user-profile is active and the corresponding users are online.
Apply the QoS policy	**qos apply policy** *policy-name* { **inbound** | **outbound** }	Required
Return to system view	**quit**	—
Activate the user profile	**user-profile** *profile-name* **enable**	Required Inactive by default.

Note

- When a user profile is active, you cannot configure or remove the QoS policy applied to it.
- The QoS policies applied in user profile view support only the **remark**, **car**, and **filter** actions.
- Do not apply an empty QoS policy in user profile view, because even if you can do that, the user profile cannot be activated.
- Refer to *User Profile Configuration* in the *System Volume* for more information about user profiles.

Applying a QoS policy to a VLAN

Follow these steps to apply the QoS policy to a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Apply the QoS policy to the specified VLAN(s)	**qos vlan-policy** *policy-name* **vlan** *vlan-id-list* { **inbound** | **outbound** }	Required

Note

- QoS policies cannot be applied to dynamic VLANs, for example, VLANs created by GVRP.
- Do not apply a QoS policy to a VLAN and the ports in the VLAN at the same time.
- A policy configured with the **nest** action, the **remark customer-vlan-id** action, or the **remark service-vlan-id** action cannot be applied to a VLAN.

Applying a QoS policy globally

A QoS policy applied globally takes effect on all ports on the device. Only one policy can be applied globally in one direction (inbound or outbound).

Follow these steps to apply a QoS policy globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Apply the QoS policy globally	**qos apply policy** *policy-name* **global** { **inbound** | **outbound** }	Required

Note

A policy configured with the **nest** action, the **remark customer-vlan-id** action, or the **remark service-vlan-id** action cannot be applied globally.

Support for QoS actions in different directions

Before creating and applying a QoS policy, you must be aware that some QoS actions are supported only in a particular traffic direction, as shown in Table 2-2:

Table 2-2 Support for QoS actions in different traffic directions

Direction (right) Action (below)	Inbound	Outbound
Traffic accounting	Supported	Supported
Traffic policing	Supported	Supported
Traffic filtering	Supported	Supported
Traffic mirroring	Supported	Supported
Tagging packets with an outer VLAN tag	Supported	Not supported
Traffic redirecting	Supported	Not supported
Marking customer VLAN IDs	Not supported	Supported
Marking 802.1p priority	Supported	Supported
Marking drop precedence	Supported	Not supported
Marking DSCP precedence	Supported	Supported

Direction (right) Action (below)	Inbound	Outbound
Marking IP precedence	Supported	Supported
Marking local precedence	Supported	Not supported
Marking service VLAN IDs	Supported	Supported

Caution

Follow these rules when configuring a behavior. Otherwise the corresponding QoS policy cannot be applied successfully.

- The action of creating an outer VLAN tag cannot be configured simultaneously with any other action except the traffic filtering action or the action of setting 802.1p precedence in the same traffic behavior. The action of creating an outer VLAN tag must be applied to basic QinQ-enabled ports or port groups.
- When the action of setting the service provider network VLAN ID is applied in the inbound direction, any other action except the traffic filtering action or the action of setting 802.1p precedence cannot be configured in the same traffic behavior.
- When the action of mirroring traffic is applied in the outbound direction, any other action cannot be configured in the same traffic behavior.

Configuration example

1) Configuration example 1

Configure a QoS policy **test_policy**. Associate the traffic behavior **test_behavior** with the traffic class **test_class** in the policy, and apply the policy to:

- the inbound direction of GigabitEthernet 1/0/1.
- the inbound direction of VLAN 200, VLAN 300, VLAN 400, VLAN 500, VLAN 600, VLAN 700, VLAN 800, and VLAN 900.
- the inbound direction globally.

Configuration procedure:

Enter system view.

```
<Sysname> system-view
```

Create a policy (This operation leads you to policy view).

```
[Sysname] qos policy test_policy
[Sysname-qospolicy-test_policy]
```

Associate the traffic behavior **test_behavior** with the class **test_class**.

```
[Sysname-qospolicy-test_policy] classifier test_class behavior test_behavior
[Sysname-qospolicy-test_policy] quit
```

Apply the QoS policy to the inbound direction of GigabitEthernet 1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos apply policy test_policy inbound
[Sysname-GigabitEthernet1/0/1] quit
```

Apply the QoS policy to the inbound direction of the specified VLANs.

```
[Sysname] qos vlan-policy test_policy vlan 200 300 400 500 600 700 800 900 inbound
```

Apply the QoS policy globally in the inbound direction.

```
[Sysname] qos apply policy test_policy global inbound
```

2) Configuration example 2

Apply the QoS policy **test_policy** in the inbound direction of the online users of the 802.1x user profile **user**.

```
<Sysname> system-view
[Sysname] user-profile user dot1x
[Sysname-user-profile-DOT1X-user] qos apply policy test_policy inbound
[Sysname-user-profile-DOT1X-user] quit
[Sysname] user-profile user enable
```

Displaying and Maintaining QoS Policies

To do…	Use the command…	Remarks
Display information about a class and the corresponding actions associated by a policy	**display qos policy user-defined** [*policy-name* [**classifier** *classifier-name*]]	Available in any view
Display information about the policies applied on a port	**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**]	Available in any view
Display information about a traffic behavior	**display traffic behavior user-defined** [*behavior-name*]	Available in any view
Display information about a class	**display traffic classifier user-defined** [*classifier-name*]	Available in any view
Display information about a global QoS policy	**display qos policy global** [**slot** *slot-number*] [**inbound** | **outbound**]	Available in any view
Display information about QoS policies applied to VLANs	**display qos vlan-policy** { **name** *policy-name* | **vlan** [*vlan-id*] } [**slot** *slot-number*] [**inbound** | **outbound**]	Available in any view
Clear the statistics of a global QoS policy	**reset qos policy global** [**inbound** | **outbound**]	Available in user view
Clear the statistics of QoS policies applied to VLANs	**reset qos vlan-policy** [**vlan** *vlan-id*] [**inbound** | **outbound**]	Available in user view

3 Priority Mapping

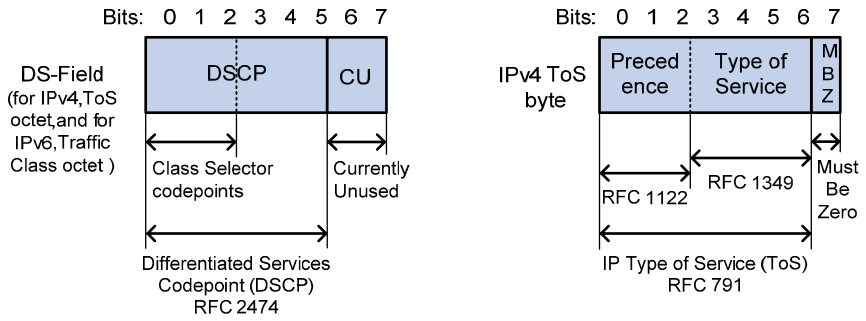
When configuring priority mapping, go to these sections for information you are interested in:

Priority Overview

The following describes several types of precedence:

1) IP precedence, ToS precedence, and DSCP precedence

Figure 3-1 DS field and ToS field



The ToS field in an IP header contains eight bits, which are described as follows:

- The first three bits indicate IP precedence in the range of 0 to 7.
- Bit 3 to bit 6 indicate ToS precedence in the range of 0 to 15.
- RFC2474 re-defines the ToS field in the IP packet header, which is called the DS field. The first six (bit 0 to bit 5) bits of the DS field indicate DSCP precedence in the range of 0 to 63. The last two bits (bit 6 and bit 7) are reserved bits.

Table 3-1 Description on IP Precedence

IP Precedence (decimal)	IP Precedence (binary)	Description
0	000	Routine
1	001	priority
2	010	immediate
3	011	flash
4	100	flash-override
5	101	critical

IP Precedence (decimal)	IP Precedence (binary)	Description
6	110	internet
7	111	network

In a network providing differentiated services, traffics are grouped into the following four classes, and packets are processed according to their DSCP values.

- Expedited Forwarding (EF) class: In this class, packets can be forwarded regardless of link share of other traffic. The class is suitable for preferential services with low delay, low packet loss ratio, low jitter, and assured bandwidth (such as virtual leased line);
- Assured forwarding (AF) class: This class is further divided into four subclasses (AF1/2/3/4) and a subclass is further divided into three drop priorities, so the AF service level can be segmented. The QoS rank of the AF class is lower than that of the EF class;
- Class selector (CS) class: This class comes from the IP ToS field and includes eight subclasses;
- Best Effort (BE) class: This class is a special class without any assurance in the CS class. The AF class can be degraded to the BE class if it exceeds the limit. Current IP network traffic belongs to this class by default.

Table 3-2 Description on DSCP precedence values

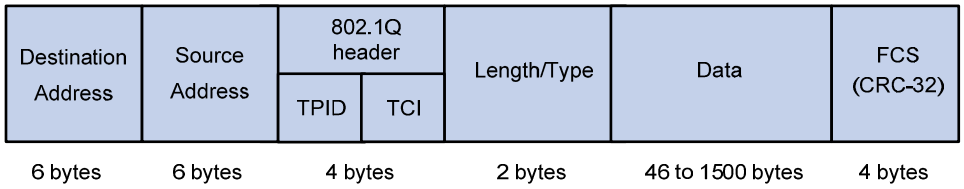
DSCP value (decimal)	DSCP value (binary)	Description
46	101110	ef
10	001010	af11
12	001100	af12
14	001110	af13
18	010010	af21
20	010100	af22
22	010110	af23
26	011010	af31
28	011100	af32
30	011110	af33
34	100010	af41
36	100100	af42
38	100110	af43
8	001000	cs1
16	010000	cs2
24	011000	cs3
32	100000	cs4
40	101000	cs5
48	110000	cs6

DSCP value (decimal)	DSCP value (binary)	Description
56	111000	cs7
0	000000	be (default)

2) 802.1p priority

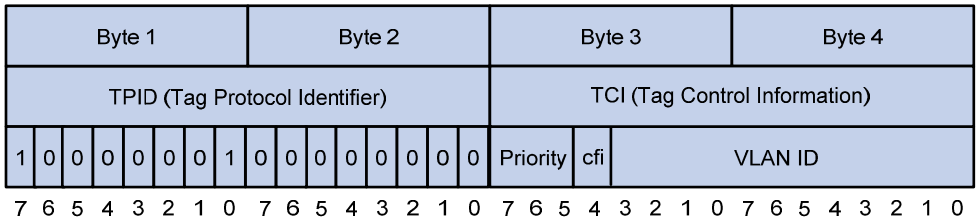
802.1p priority lies in Layer 2 packet headers and is applicable to occasions where the Layer 3 packet header does not need analysis but QoS must be assured at Layer 2.

Figure 3-2 An Ethernet frame with an 802.1Q tag header



As shown in Figure 3-2, the 4-byte 802.1Q tag header contains a 2-byte Tag Protocol Identifier (TPID) whose value is 8100 and a 2-byte Tag Control Information (TCI). TPID is a new class defined by IEEE to indicate a packet with an 802.1Q tag. Figure 3-3 describes the detailed contents of an 802.1Q tag header.

Figure 3-3 802.1Q tag header



In the figure above, the 3-bit priority field in TCI is 802.1p priority in the range of 0 to 7. In the figure above, the priority field (three bits in length) in TCI is 802.1p priority (also known as CoS precedence), which ranges from 0 to 7.

Table 3-3 Description on 802.1p priority

802.1p priority (decimal)	802.1p priority (binary)	Description
0	000	best-effort
1	001	background
2	010	spare
3	011	excellent-effort
4	100	controlled-load
5	101	video
6	110	voice
7	111	network-management

The precedence is called 802.1p priority because the related applications of this precedence are defined in detail in the 802.1p specifications.

Priority Mapping Overview

When a packet reaches a switch, the switch assigns the packet parameters according to it configuration, such as 802.1p precedence, DSCP precedence, IP precedence, local precedence, and drop precedence.

The local precedence and drop precedence are described as follows.

- Local precedence is the precedence that the switch assigns to a packet and it is corresponding to the number of an outbound queue on the port. Local precedence takes effect only on the local switch.
- Drop precedence is a parameter that is referred to when dropping packets. The higher the drop precedence, the more likely a packet is dropped.

Depending on whether a received packet is 802.1q-tagged, the switch marks it with priority as follows:

1) For an 802.1q-untagged packet

When a packet carrying no 802.1q tag reaches a port, the switch uses the port priority as the 802.1p precedence value of the received packet, searches for the local precedence value corresponding to the port priority of the receiving port in the 802.1p-precedence-to-local-precedence mapping table, assigns the local precedence value to the packet, and enqueues the packet according to the local precedence value.

2) For an 802.1q-tagged packet

When an 802.1q tagged packet reaches the port of a switch, you can specify a priority trust mode for the port, trusting port priority or trusting packet priority.

- Trusting packet priority

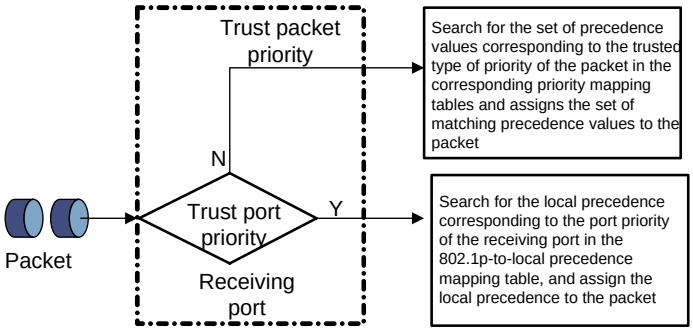
In this mode, the switch searches for the set of precedence values corresponding to the trusted type (802.1p precedence or DSCP precedence) of priority of the packet in the corresponding priority mapping tables and assigns the set of matching precedence values to the packet.

- Trusting port priority

In this mode, the switch replaces the 802.1p priority of the received packet with the port priority, searches for the local precedence corresponding to the port priority of the receiving port in the 802.1p-to-local precedence mapping table, assigns the local precedence to the packet, and enqueues the packet according to the local precedence value.

You can configure the priority trust mode of a port as required. The priority mapping process on a switch is as shown in Figure 3-4.

Figure 3-4 Priority mapping process in the case of supporting trusting port priority



When trusting packet priority, Switch 4800G provide the following two priority trust modes:can trust one of the following two priority types:

- Trusting the DSCP precedence of received packets. In this mode, the switch searches the **dscp-dot1p/dp/dscp** mapping table based on the DSCP precedence of the received packet for the 802.1p precedence/drop precedence/DSCP precedence to be used to mark the packet. Then the switch searches the **dot1p-lp** mapping table based on the marked 802.1p precedence for the corresponding local precedence and marks the received packet with the local precedence.
- Trusting the 802.1p precedence of received packets. In this mode, if a packet is received without an 802.1q tag, the switch takes the priority of the receiving port as the 802.1p precedence of the packet and then based on the priority searches the **dot1p-dp/lp** mapping table for the local/drop precedence for the packet. If packet is received with an 802.1q tag, the switch searches the **dot1p-dp/lp** mapping table based on the 802.1p precedence in the tag for local/drop precedence for the packet.

The default **dot1p-lp/dp** mapping and **dscp-dot1p/dp/dscp** mapping provided by Switch 4800G series Ethernet switches are shown in the following two tables.

- **dot1p-dp**: 802.1p-priority-to-drop-precedence mapping table
- **dot1p-lp**: 802.1p-priority-to-local-precedence mapping table
- **dscp-dot1p**: DSCP-precedence-to-802.1p-priority mapping table
- **dscp-dp**: DSCP-precedence-to-drop-precedence mapping table, applicable to only IP packets
- **dscp-dscp**: DSCP-precedence-to-DSCP-precedence mapping table, applicable to only IP packets

Table 3-4 The default values of **dot1p-lp** mapping and **dot1p-dp** mapping

Imported priority value	dot1p-lp mapping	dot1p-dp mapping
802.1p precedence (**dot1p**)	Local precedence (**lp**)	Drop precedence (**dp**)
0	2	0
1	0	0
2	1	0
3	3	0
4	4	0
5	5	0
6	6	0
7	7	0

Table 3-5 The default values of **dscp-dp** mapping, **dscp-dot1p** mapping, and **dscp-dscp** mapping

Imported priority value	dscp-dp mapping	dscp-dot1p mapping	dscp-dscp mapping
DSCP precedence (**dscp**)	Drop precedence (**dp**)	802.1p precedence (**dot1p**)	DSCP precedence (**dscp**)
0 to 7	0	0	0 to 7
8 to 15	0	1	8 to 15
16 to 23	0	2	16 to 23
24 to 31	0	3	24 to 31
32 to 39	0	4	32 to 39
40 to 47	0	5	40 to 47
48 to 55	0	6	48 to 55
56 to 63	0	7	56 to 63

Configuring a Priority Mapping Table

You can modify the priority mapping tables in a switch as required.

Follow the two steps to configure priority mapping tables:

- Enter priority mapping table view;
- Configure priority mapping parameters.

Configuration Prerequisites

The new priority mapping table is determined.

Configuration Procedure

Follow these steps to configure a priority mapping table:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter priority mapping table view	**qos map-table** { **dot1p-dp** | **dot1p-lp** | **dscp-dot1p** | **dscp-dp** | **dscp-dscp** }	Required To configure a priority mapping table, you need to enter the corresponding priority mapping table view.
Configure priority mapping parameters	**import** *import-value-list* **export** *export-value*	Required The newly configured mapping entries overwrite the corresponding previous entries.

Note

You cannot configure to map any DSCP value to drop precedence 1.

Configuration Example

Network requirements

Modify the **dot1p-lp** mapping table as those listed in Table 3-6.

Table 3-6 The specified **dot1p-lp** mapping

802.1p precedence	Local precedence
0	0
1	0
2	1
3	1
4	2
5	2
6	3
7	3

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter **dot1p-lp** priority mapping table view.

```
[Sysname] qos map-table dot1p-lp
```

Modify **dot1p-lp** priority mapping parameters.

```
[Sysname-maptbl-dot1p-lp] import 0 1 export 0
[Sysname-maptbl-dot1p-lp] import 2 3 export 1
[Sysname-maptbl-dot1p-lp] import 4 5 export 2
[Sysname-maptbl-dot1p-lp] import 6 7 export 3
```

Configuring the Port Priority

By default, if a port receives packets without 802.1q tags, the switch takes the priority of the receiving port as the 802.1p precedence of the received packets, searches the **dot1p-lp/dp** mapping table for the corresponding local precedence and drop precedence according to the 802.1p precedence of the received packets, and then marks the received packets with the corresponding local precedence and drop precedence.

Port priority is in the range 0 to 7. You can set the port priority as required.

Configuration Prerequisites

The port priority of the port is determined.

Configuration Procedure

Follow these steps to configure port priority:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure port priority		**qos priority** *priority-value*	Required By default, the port priority is 0.

Configuration Example

Network requirements

Configure the port priority to 7.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure port priority of GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos priority 7
```

Configuring Port Priority Trust Mode

You can configure the switch to trust the DSCP precedence of the received packets. In this case, the switch searches the **dscp-dot1p/dp/dscp** mapping table for the corresponding precedence according to the DSCP precedence of the packets and marks the received packets with the precedence.

Configuration Prerequisites

It is determined to trust the DSCP precedence or the 802.1p precedence of received packets.

Configuration Procedure

Follow these steps to configure the port priority trust mode:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure to trust the DSCP precedence or the 802.1p precedence of the received packets		**qos trust** { **dscp** | **dot1p** }	Perform either of the two operations. By default, the port priority is trusted.
Configure to trust the port priority		**undo qod trust**	

Configuration Example

Network requirements

Configure to trust the DSCP precedence of the received packets.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter port view.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1]
```

Configure to trust the DSCP precedence of the received packets.

```
[Sysname-GigabitEthernet1/0/1] qos trust dscp
```

Displaying and Maintaining Priority Mapping

To do…	Use the command…	Remarks
Display the information about a specified priority mapping table	**display qos map-table** [**dot1p-dp** | **dot1p-lp** | **dscp-dot1p** | **dscp-dp** | **dscp-dscp**]	Available in any view
Display the priority trust mode configured for a port	**display qos trust interface** [*interface-type interface-number*]	

4 Traffic Policing, Traffic Shaping, and Line Rate Configuration

When configuring traffic classification, traffic policing, traffic shaping, and line rate, go to these section for information you are interested in:

Traffic Policing, Traffic Shaping, and Line Rate Overview

If the traffic from users is not limited, a large amount of continuous burst packets will result in worse network congestion. The traffic of users must be limited in order to make better use of the limited network resources and provide better service for more users. For example, if a traffic flow obtains only the resources committed to it within a certain period of time, network congestion due to excessive burst traffic can be avoided.

- Traffic policing and traffic shaping are traffic control policies for limiting traffic and resource usage by supervising the traffic. The prerequisite for traffic policing and traffic shaping is to determine whether or not the traffic exceeds the set threshold. Traffic control policies are adopted only when the traffic exceeds the set threshold. Generally, token bucket is used for evaluating traffic.
- The line rate of a physical interface specifies the maximum rate for forwarding packets. Line rate also uses token buckets for traffic control.

Traffic Evaluation and the Token Bucket

The token bucket can be considered as a container with a certain capacity to hold tokens. The system puts tokens into the bucket at the set rate. When the token bucket is full, the extra tokens will overflow and the number of tokens in the bucket stops increasing.

Evaluating traffic with the token bucket

The evaluation for the traffic specification is based on whether the number of tokens in the bucket can meet the need of packet forwarding. If the number of tokens in the bucket is enough to forward the packets, the traffic is conforming to the specification; otherwise, the traffic is nonconforming or excess.

When the token bucket evaluates the traffic, its parameter configurations include:

- Average rate: The rate at which tokens are put into the bucket, namely, the permitted average rate of the traffic. It is generally set to committed information rate (CIR).
- Burst size: The capacity of the token bucket, namely, the maximum traffic size that is permitted in each burst. It is generally set to committed burst size (CBS). The set burst size must be greater than the maximum packet length.

An evaluation is performed on the arrival of each packet. In each evaluation, if the bucket has enough tokens for use, the traffic is controlled within the specification and a number of tokens equivalent to the

packet forwarding authority must be taken out; otherwise, this means too many tokens have been used — the traffic is in excess of the specification.

Complicated Evaluation

You can set two token buckets (referred to as the C bucket and E bucket respectively) in order to evaluate more complicated conditions and implement more flexible regulation policies. For example, traffic policing uses four parameters:

- CIR: Rate at which tokens are put into the C bucket, that is, the average packet transmission or forwarding rate allowed by the C bucket.
- CBS: Size of the C bucket, that is, transient burst of traffic that the C bucket can forward.
- Peak information rate (PIR): Rate at which tokens are put into the E bucket, that is, the average packet transmission or forwarding rate allowed by the E bucket.
- Excess burst size (EBS): Size of the E bucket, that is, transient burst of traffic that the E bucket can forward.

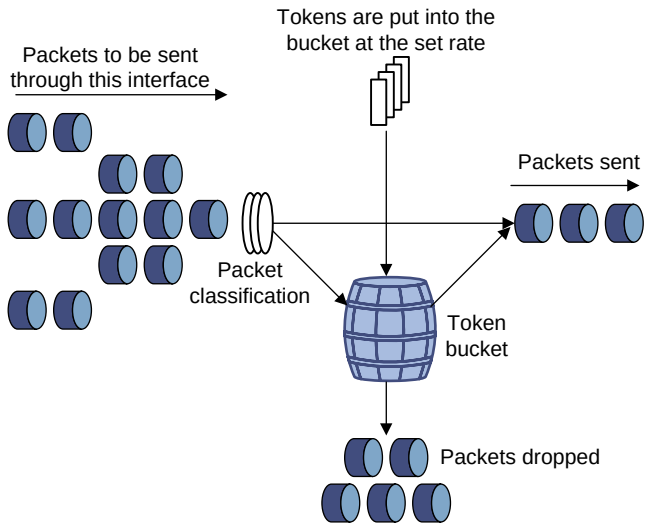
In each evaluation, packets are measured against the buckets:

- If the C bucket has enough tokens, packets are colored green.
- If the C bucket does not have enough tokens but the E bucket has enough tokens, packets are colored yellow.
- If neither the C bucket nor the E bucket has sufficient tokens, packets are colored red.

Traffic Policing

The typical application of traffic policing is to supervise the specification of certain traffic into the network and limit it within a reasonable range, or to "discipline" the extra traffic. In this way, the network resources and the interests of the operators are protected. For example, you can limit HTTP packets to be within 50% of the network bandwidth. If the traffic of a certain connection is excess, traffic policing can choose to drop the packets or to reset the priority of the packets.

Figure 4-1 Diagram for TP



Traffic policing is widely used in policing the traffic into the network of internet service providers (ISPs). Traffic policing can classify the policed traffic and perform pre-defined policing actions based on different evaluation results. These actions include:

- Forwarding conforming packets or non-conforming packets.

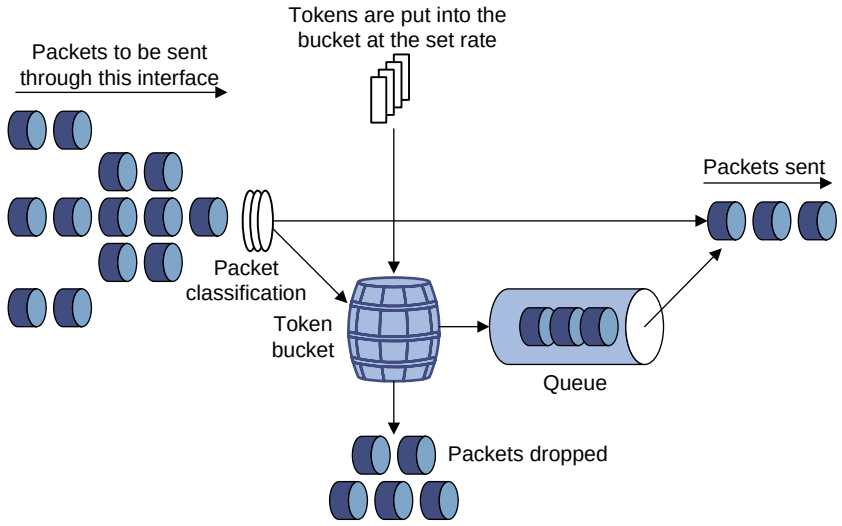
- Dropping conforming or non-conforming packets.
- Marking a conforming packet with a new 802.1p precedence value and forwarding the packet.
- Marking a conforming packet with a new IP precedence value and forwarding the packet.
- Marking a conforming packet or a non-conforming packet with a new DSCP precedence value and forwarding the packet.

Traffic Shaping

Traffic shaping provides measures to adjust the rate of outbound traffic actively. A typical traffic shaping application is to limit the local traffic output rate according to the downstream traffic policing parameters.

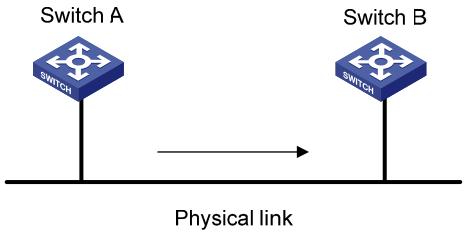
The difference between traffic policing and GTS is that packets to be dropped in traffic policing are cached in a buffer or queue in GTS, as shown in Figure 4-2. When there are enough tokens in the token bucket, these cached packets are sent at an even rate. Traffic shaping may result in an additional delay while traffic policing does not.

Figure 4-2 Diagram for GTS



For example, in Figure 4-3, Switch A sends packets to Switch B. Switch B performs traffic policing on packets from Switch A and drops packets exceeding the limit.

Figure 4-3 GTS application



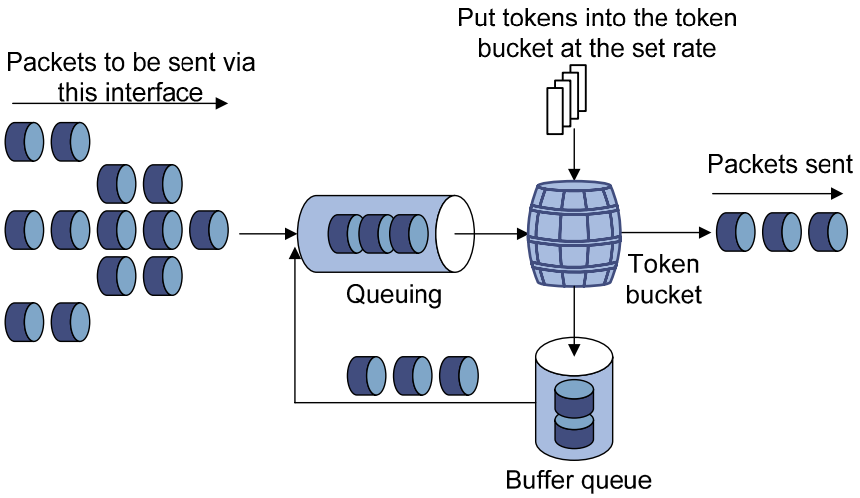
You can perform traffic shaping for the packets on the outgoing interface of Switch A to avoid unnecessary packet loss. Packets exceeding the limit are cached in Switch A. Once resources are released, traffic shaping takes out the cached packets and sends them out. In this way, all the traffic sent to Switch B conforms to the traffic specification defined in Switch B.

Line Rate

The line rate of a physical interface specifies the maximum rate for forwarding packets (including critical packets).

Line rate also uses token buckets for traffic control. With line rate configured on an interface, all packets to be sent through the interface are first handled by the token bucket at line rate. If there are enough tokens in the token bucket, packets can be forwarded; otherwise, packets are put into QoS queues for congestion management. In this way, the traffic passing the physical interface is controlled.

Figure 4-4 Line rate implementation



In the token bucket approach to traffic control, burst traffic can be transmitted so long as enough tokens are available in the token bucket; if tokens are inadequate, packets cannot be transmitted until the required number of tokens are generated in the token bucket. Thus, traffic rate is restricted to the rate for generating tokens, thus limiting traffic rate and allowing bursty traffic.

Compared with traffic policing, line rate can only limit traffic rate on a physical interface. Since traffic policing operates at the IP layer, it can limit the rate of different flows on a port. However, traffic policing ignores packets not processed by the IP layer. To limit the rate of all the packets on interfaces, using line rate is easier.

GTS/Line Rate Configuration

Configuring GTS

Configuration procedure

Follow these steps to configure GTS:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter interface view	**interface** *interface-type interface-number*	Use either command. Settings in interface view take effect on the current interface; settings in port group view take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	

To do…	Use the command…	Remarks
Configure GTS for a queue	**qos gts queue** *queue-number* **cir** *committed-information-rate* [**cbs** *committed-burst-size*]	Required

GTS configuration example

Configure GTS on GigabitEthernet 1/0/1, shaping packets when the sending rate exceeds 640 kbps in queue 2.

Enter system view.

```
<Sysname> system-view
```

Enter interface view.

```
[Sysname] interface gigabitethernet 1/0/1
```

Configure GTS parameters.

```
[Sysname-GigabitEthernet1/0/1] qos gts queue 2 cir 640
```

Line Rate Configuration Procedure

Configuration procedure

Follow these steps to configure line rate:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter interface view or port group view	Enter port view	**interface** *interface-type interface-number*	Enter either view. Settings in interface view take effect on the current interface; settings in port group view take effect on all ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure line rate		**qos lr outbound cir** *committed-information-rate* [**cbs** *committed-burst-size*]	Required

Line rate configuration examples

Limit the outbound rate of GigabitEthernet 1/0/1 to 640 kbps.

Enter system view.

```
<Sysname> system-view
```

Enter interface view.

```
[Sysname] interface GigabitEthernet 1/0/1
```

Configure line rate parameter and limit the outbound rate to 640 kbps.

```
[Sysname-GigabitEthernet1/0/1] qos lr outbound cir 640
```

Displaying and Maintaining Line Rate/GTS

To do…	Use the command…	Remarks
Display the GTS configuration of interfaces	**display qos gts interface** [*interface-type interface-number*]	Available in any view
Display the line rate configuration of interfaces	**display qos lr interface** [*interface-type interface-number*]	Available in any view

5 Congestion Management

When configuring congestion management, go to these section for information that you are interested in:

Overview

When the rate at which the packets arrive is higher than the rate at which the packets are transmitted on an interface, congestion occurs on this interface. If there is not enough storage space to store these packets, parts of them will be lost. Packet loss may cause the transmitting device to retransmit the packets because the lost packets time out, which causes a malicious cycle.

The core of congestion management is how to schedule the resources and determine the sequence of forwarding packets when congestion occurs. Congestion management processing includes queue creating, traffic classification, packet enqueuing, and queue scheduling.

Congestion Management Policy

Queuing technology is generally adopted to solve the congestion problem. The queuing technology is to classify the traffic according to a specified queue-scheduling algorithm and then use the specified priority algorithm to forward the traffic. Each queuing algorithm is used to solve specific network traffic problems and affects the parameters such as bandwidth allocation, delay and delay jitter.

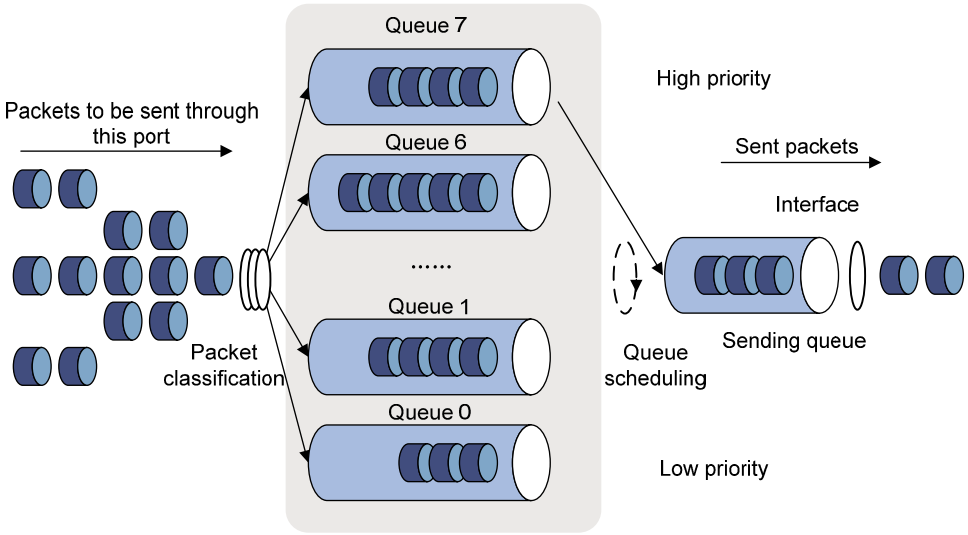
The Switch 4800G series support the following four queue scheduling methods:

- Scheduling all queues with the strict priority (SP) algorithm.
- Scheduling all queues with the weighted round robin (WRR) algorithm.
- Scheduling all queues with the weighted fair queuing (WFQ) algorithm
- Scheduling some queues with the SP algorithm and some with the WRR algorithm.

This section describe how SP, WRR, WFQ, and SP+WRR work in details.

1) SP queue-scheduling algorithm

Figure 5-1 Diagram for SP queuing



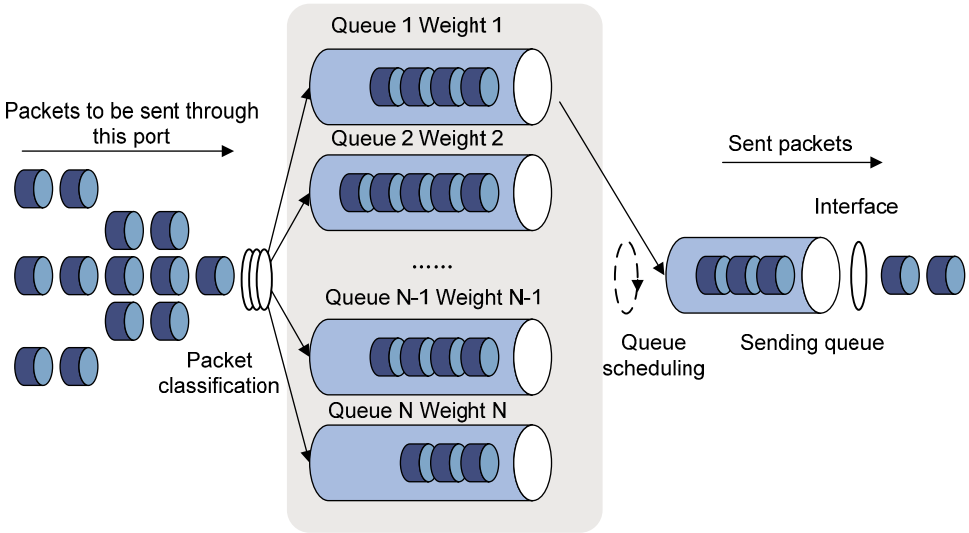
SP queue-scheduling algorithm is specially designed for critical service applications. An important feature of critical services is that they demand preferential service in congestion in order to reduce the response delay. Assume that there are eight output queues on the port and the preferential queue classifies the eight output queues on the port into eight classes, which are queue7, queue6, queue5, queue4, queue3, queue2, queue1, and queue0. Their priorities decrease in order.

In queue scheduling, SP sends packets in the queue with higher priority strictly following the priority order from high to low. When the queue with higher priority is empty, packets in the queue with lower priority are sent. You can put critical service packets into the queues with higher priority and put non-critical service (such as e-mail) packets into the queues with lower priority. In this case, critical service packets are sent preferentially and non-critical service packets are sent when critical service groups are not sent.

The disadvantage of SP queue is that: if there are packets in the queues with higher priority for a long time in congestion, the packets in the queues with lower priority will be “starved” because they are not served.

2) WRR queue-scheduling algorithm

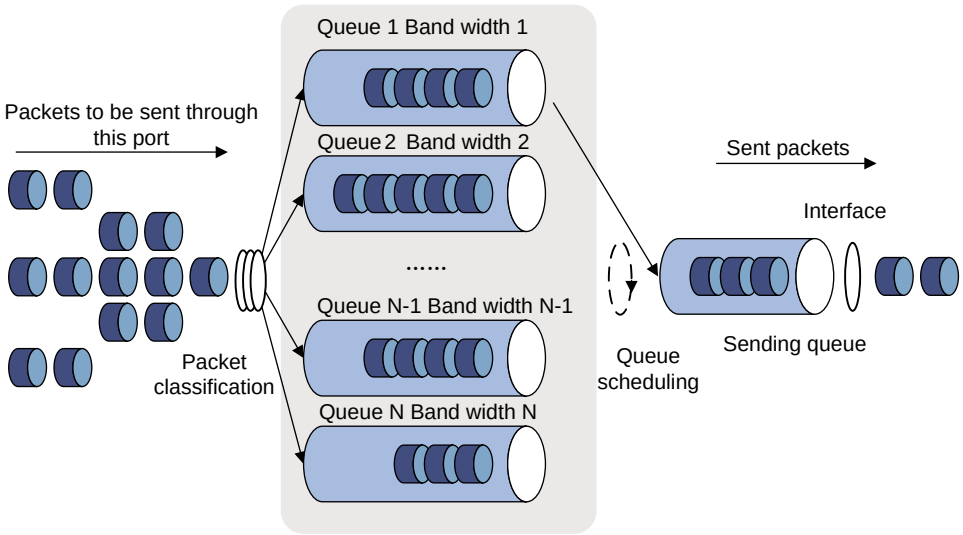
Figure 5-2 Diagram for WRR queuing



A port of the switch supports eight outbound queues. The WRR queue-scheduling algorithm schedules all the queues in turn to ensure that every queue can be assigned a certain service time. Assume there are eight output queues on the port. The eight weight values (namely, w 7, w 6, w 5, w 4, w 3, w 2, w 1, and w 0) indicating the proportion of assigned resources are assigned to the eight queues respectively. On a 100M port, you can configure the weight values of WRR queue-scheduling algorithm to 50, 30, 10, 10, 50, 30, 10, and 10 (corresponding to w7, w6, w5, w4, w3, w2, w1, and w0 respectively). In this way, the queue with the lowest priority can be assured of 5 Mbps of bandwidth at least, thus avoiding the disadvantage of SP queue-scheduling algorithm that packets in low-priority queues are possibly not to be served for a long time. Another advantage of WRR queue-scheduling algorithm is that though the queues are scheduled in turn, the service time for each queue is not fixed, that is to say, if a queue is empty, the next queue will be scheduled immediately. In this way, the bandwidth resources are fully utilized.

3) WFQ queue-scheduling algorithm

Figure 5-3 Diagram for WFQ queuing



Before WFQ is introduced, you need to understand fair queuing (FQ). FQ is designed for fairly sharing network resources, reducing the delay and jitter of all traffic. FQ takes all the aspects into consideration:

- Different queues have fair dispatching opportunities for delay balancing among streams.
- Short packets and long packets are fairly scheduled: if there are long packets and short packets in queues, statistically the short packets should be scheduled preferentially to reduce the jitter between packets on the whole.

Compared with FQ, WFQ takes weights into account when determining the queue scheduling order. Statistically, WFQ gives high priority traffic more scheduling opportunities than low priority traffic. WFQ can automatically classify traffic according to the “session” information of traffic (protocol type, TCP or UDP source/destination port numbers, source/destination IP addresses, IP precedence bits in the ToS field, etc), and try to provide as many queues as possible so that each traffic flow can be put into these queues to balance the delay of every traffic flow on a whole. When dequeuing packets, WFQ assigns the outgoing interface bandwidth to each traffic flow by the precedence. The higher precedence value a traffic flow has, the more bandwidth it gets.

The Switch 4800G series switches introduce the minimum guaranteed bandwidth mechanism, and use it in conjunction with WFQ as follows:

- The minimum guaranteed bandwidth configuration guarantees a certain amount of bandwidth for each WFQ queue.

- The allocable bandwidth (allocable bandwidth = the total bandwidth – the sum of the minimum guaranteed bandwidth for each queue) is divided and allocated to each queue based on queue precedence.

For example, assume that the total bandwidth of an interface is 10 Mbps and there are five flows on the interface, with the precedence being 0, 1, 2, 3, and 4 respectively and the minimum guaranteed bandwidth being 128 kbps, 128 kbps, 128 kbps, 64 kbps, and 64 kbps respectively. Then,

- The allocable bandwidth = 10 Mbps – (128 + 128 + 128 + 64 + 64) kbps = 9.5 Mbps
- The total allocable bandwidth quota is the sum of all the (precedence value + 1)s, that is, 1 + 2 + 3 + 4 + 5 = 15.
- The bandwidth percentage assigned to each flow is (precedence value of the flow + 1)/total allocable bandwidth quota. The bandwidth percentages for flows are 1/15, 2/15, 3/15, 4/15, and 5/15 respectively.
- The bandwidth allocated to a queue = Minimum guaranteed bandwidth + bandwidth allocated to the queue from the allocable bandwidth

Because WFQ can balance the delay and jitter of every flow when congestion occurs, it is effectively applied in some special occasions. For example, WFQ is adopted in the assured forwarding (AF) services of the Resource Reservation Protocol (RSVP). In Generic Traffic Shaping (GTS), WFQ is used to schedule buffered packets.

4) SP+WRR queue scheduling algorithm

You can implement SP+WRR queue scheduling on a port by assigning some queues on the port to the SP scheduling group and the others to the WRR scheduling group (that is, group 1). Packets in the SP scheduling group are scheduled preferentially. When the SP scheduling group is empty, packets in the WRR scheduling group are scheduled. Queues in the SP scheduling group are scheduled by SP. Queues in the WRR scheduling group are scheduled by WRR.

Configuring an SP Queue

By default, WRR queue scheduling algorithm is adopted on all the ports. You can adopt SP queue scheduling algorithm instead as required.

Configuration Procedure

Follow these steps to configure SP queues:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure SP queue scheduling algorithm		**qos sp**	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.

Configuration Example

Network requirements

Configure GigabitEthernet1/0/1 to adopt SP queue scheduling algorithm.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure an SP queue for GigabitEthernet1/0/1 port.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos sp
```

Configuring a WRR Queue

Configuration Procedure

Follow these steps to configure WRR queues:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group
	Enter port group view	**port-group manual** *port-group-name*	
Enable WRR on the port		**qos wrr**	Optional Enabled by default.
Configure WRR queue scheduling		**qos wrr** *queue-id* **group** *group-id* **weight** *schedule-value*	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.

Configuration Example

Network requirements

Configure WRR queue scheduling algorithm on GigabitEthernet1/0/1, and assign weight 1, 2, 4, 6, 8, 10, 12, and 14 to queue 0 through queue 7.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure the WRR queues on GigabitEthernet1/0/1 port.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos wrr
[Sysname-GigabitEthernet1/0/1] qos wrr 0 group 1 weight 1
```

```
[Sysname-GigabitEthernet1/0/1] qos wrr 1 group 1 weight 2
[Sysname-GigabitEthernet1/0/1] qos wrr 2 group 1 weight 4
[Sysname-GigabitEthernet1/0/1] qos wrr 3 group 1 weight 6
[Sysname-GigabitEthernet1/0/1] qos wrr 4 group 1 weight 8
[Sysname-GigabitEthernet1/0/1] qos wrr 5 group 1 weight 10
[Sysname-GigabitEthernet1/0/1] qos wrr 6 group 1 weight 12
[Sysname-GigabitEthernet1/0/1] qos wrr 7 group 1 weight 14
```

Configuring a WFQ Queue

By default, all ports adopt the WRR queue algorithm. You can configure a port to use the WFQ algorithm instead as required.

Configuration Procedure

Follow these steps to configure WFQ queues:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group
	Enter port group view	**port-group manual** *port-group-name*	
Adopt the WFQ queue scheduling on the port		**qos wfq**	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.
Configure the minimum guaranteed bandwidth for a WFQ queue		**qos bandwidth queue** *queue-id* **min** *bandwidth-value*	Optional By default, the minimum guaranteed bandwidth of a queue is 64 kbps.
Configure a scheduling weight for the specified queue		**qos wfq** *queue-id* **weight** *schedule-value*	Optional By default, the scheduling weight of a WFQ queue is 1.

Configuration Example

Network requirements

Enable WFQ on GigabitEthernet 1/0/1 and assign weight values 1, 2, 4, 6, 8, 10, 12, and 14 to queues 0 through 7 respectively.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure the WFQ queues on GigabitEthernet1/0/1 port.

```
[Sysname] interface GigabitEthernet 1/0/1
```

```
[Sysname-GigabitEthernet1/0/1] qos wfq
[Sysname-GigabitEthernet1/0/1] qos wfq 0 weight 1
[Sysname-GigabitEthernet1/0/1] qos wfq 1 weight 2
[Sysname-GigabitEthernet1/0/1] qos wfq 2 weight 4
[Sysname-GigabitEthernet1/0/1] qos wfq 3 weight 6
[Sysname-GigabitEthernet1/0/1] qos wfq 4 weight 8
[Sysname-GigabitEthernet1/0/1] qos wfq 5 weight 10
[Sysname-GigabitEthernet1/0/1] qos wfq 6 weight 12
[Sysname-GigabitEthernet1/0/1] qos wfq 7 weight 14
```

Configuring SP+WRR Queues

By default, all ports adopt the WRR queue algorithm. You can configure a port to use the SP+WRR queue scheduling algorithm as required.

Configuration Procedure

Follow these steps to configure SP + WRR queues:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Perform either of the two operations. The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Enable the WRR queue scheduling on the port		**qos wrr**	Required
Configure SP queue scheduling		**qos wrr** *queue-id* **group sp**	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.
Configure WRR queue scheduling		**qos wrr** *queue-id* **group** *group-id* **weight** *schedule-value*	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.

Configuration Example

Network requirements

- Configure to adopt SP+WRR queue scheduling algorithm on GigabitEthernet1/0/1.
- Configure queue 0, queue 1, queue 2 and queue 3 on GigabitEthernet1/0/1 to be in SP queue scheduling group.
- Configure queue 4, queue 5, queue 6 and queue 7 on GigabitEthernet1/0/1 to be in WRR queue scheduling group, with the weight being 2, 4, 6 and 8 respectively.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enable the SP+WRR queue scheduling algorithm on GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos wrr
[Sysname-GigabitEthernet1/0/1] qos wrr 0 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 1 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 2 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 3 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 4 group 1 weight 2
[Sysname-GigabitEthernet1/0/1] qos wrr 5 group 1 weight 4
[Sysname-GigabitEthernet1/0/1] qos wrr 6 group 1 weight 6
[Sysname-GigabitEthernet1/0/1] qos wrr 7 group 1 weight 8
```

Displaying and Maintaining Congestion Management

To do…	Use the command…	Remarks
Display WRR queue configuration information	**display qos wrr interface** [*interface-type interface-number*]	Available in any view
Display SP queue configuration information	**display qos sp interface** [*interface-type interface-number*]	
Display WFQ queue configuration information	**display qos wfq interface** [*interface-type interface-number*]	

6 Congestion Avoidance

When configuring congestion avoidance, go to these sections for information you are interested in:

- Congestion Avoidance Overview
- Configuring WRED
- Displaying and Maintaining WRED

Congestion Avoidance Overview

Serious congestion causes great damages to the network resources, and therefore some measures must be taken to avoid such congestion. As a flow control mechanism, congestion avoidance can actively drop packets when congestion deteriorates through monitoring the utilization of network resources (such as queues or memory buffers) to prevent network overload.

Compared to point-to-point flow control, this flow control mechanism is of broader sense because it can control the load of more flows in a device. When dropping packets from a source end, it can still cooperate well with the flow control mechanism (such as TCP flow control) at the source end to better adjust the network traffic to a reasonable load status. The combination of the packet drop policy of the local device and the flow control mechanism at the source end can maximize throughput and utilization rate of the network and minimize packet loss and delay.

Traditional packet drop policy

The traditional packet drop policy is tail drop. When the length of a queue reaches the maximum threshold, all the subsequent packets are dropped.

Such a policy results in global TCP synchronization. That is, if packets from multiple TCP connections are dropped, these TCP connections go into the state of congestion avoidance and slow start to reduce traffic, but traffic peak occurs later. Consequently, the network traffic jitters all the time.

RED and WRED

You can use random early detection (RED) or weighted random early detection (WRED) to avoid global TCP synchronization.

The RED or WRED algorithm sets an upper threshold and lower threshold for each queue, and processes the packets in a queue as follows:

- When the queue size is shorter than the lower threshold, no packet is dropped;
- When the queue size reaches the upper threshold, all subsequent packets are dropped;
- When the queue size is between the lower threshold and the upper threshold, the received packets are dropped at random. The longer a queue is, the higher the drop probability is. However, a maximum drop probability exists.

Different from RED, WRED determines differentiated drop policies for packets with different IP precedence values. Packets with a lower IP precedence are more likely to be dropped.

Both RED and WRED avoid global TCP synchronization by randomly dropping packets. When the sending rate of a TCP session slows down after its packets are dropped, the other TCP sessions remain

in high packet sending rates. In this way, some TCP sessions remain in high sending rates in any case, and the link bandwidth can be fully utilized.

Configuring WRED

Configuration Prerequisites

Before configuring WRED, determine the following:

- The parameters to be configured in the WRED table
- The port/port group where the WRED table is to be applied

Configuration Procedure

Follow these steps to configure WRED:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a WRED table		**qos wred queue table** *table-name*	—
Configure drop parameters for the WRED table		**queue** *queue-id* [**drop-level** *drop-level*] **low-limit** *low-limit* [**discard-probability** *discard-prob*]	Optional By default, the *low-limit* argument is 10 and the *discard-prob* argument is 10.
Enter port view or port group view	Enter port view	**interface** *interface-type interface-number*	Use either command The configuration performed in Ethernet interface view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Apply the WRED table		**qos wred apply** *table-name*	Required

Configuration Example

Network requirements

Create a WRED table with the default parameters and apply the WRED table to GigabitEthernet 1/0/1.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create a WRED table with the default parameters.

```
[Sysname] qos wred queue table queue-table1
[Sysname-wred-table-queue-table1] quit
```

Apply the WRED table to GigabitEthernet 1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos wred apply queue-table1
```

Displaying and Maintaining WRED

To do…	Use the command…	Remarks
Display the WRED configuration of a port	**display qos wred interface** [*interface-type interface-number*]	Available in any view
Display the configuration of a WRED table	**display qos wred table** [*table-name*]	Available in any view

7 Traffic Mirroring Configuration

When configuring traffic mirroring, go to these sections for information that you are interested in:

Overview

Traffic mirroring is to replicate the specified packets to the specified destination. It is generally used for testing and troubleshooting the network. .

Depending on different types of mirroring destinations, there are three types of traffic mirroring:

- Mirroring to port: The desired traffic on a mirrored port is replicated and sent to a destination port (that is, a mirroring port).
- Mirroring to CPU: The desired traffic on a mirrored port is replicated and sent to the CPU for further analysis.
- Mirroring to VLAN: The desired traffic on a mirrored port is replicated and sent to a VLAN, where the traffic is broadcast and all the ports (if available) in the VLAN will receive the traffic. If the destination VLAN does not exist, you can still configure the function, and the function will automatically take effect after the VLAN is created and a port is added to it.

Note

On Switch 4800G series Ethernet switches, traffic can only be mirrored to ports and to CPU.

Configuring Traffic Mirroring

To configure traffic mirroring, you must enter the view of an existing traffic behavior.

Follow these steps to configure traffic mirroring to a port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Configure traffic mirroring action in the traffic behavior	**mirror-to** { **cpu** | **interface** *interface-type interface-number* }	Required

Displaying and Maintaining Traffic Mirroring

To do…	Use the command…	Remarks
Display the configuration information about the user-defined traffic behavior	**display traffic behavior user-defined** *behavior-name*	Available in any view
Display the configuration information about the user-defined policy	**display qos policy user-defined** *policy-name*	

Traffic Mirroring Configuration Example

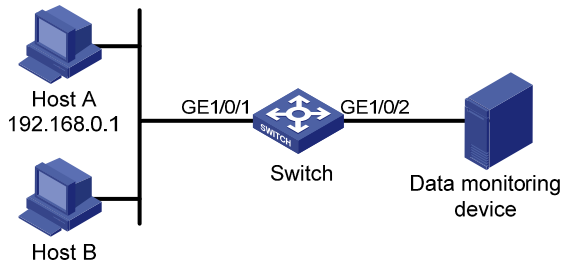
Network Requirements

The user's network is as described below:

- Host A (with the IP address 192.168.0.1) and Host B are connected to GigabitEthernet1/0/1 of the switch.
- The data monitoring device is connected to GigabitEthernet1/0/2 of the switch.

It is required to monitor and analyze packets sent by Host A on the data monitoring device.

Figure 7-1 Network diagram for configuring traffic mirroring to a port



Configuration Procedure

Configure Switch:

Enter system view.

```
<Sysname> system-view
```

Configure basic IPv4 ACL 2000 to match packets with the source IP address 192.168.0.1.

```
[Sysname] acl number 2000
[Sysname-acl-basic-2000] rule permit source 192.168.0.1 0
[Sysname-acl-basic-2000] quit
```

Configure a traffic classification rule to use ACL 2000 for traffic classification.

```
[Sysname] traffic classifier 1
[Sysname-classifier-1] if-match acl 2000
[Sysname-classifier-1] quit
```

Configure a traffic behavior and define the action of mirroring traffic to GigabitEthernet1/0/2 in the traffic behavior.

```
[Sysname] traffic behavior 1
[Sysname-behavior-1] mirror-to interface GigabitEthernet 1/0/2
```

```
[Sysname-behavior-1] quit
```

Configure a QoS policy and associate traffic behavior 1 with classification rule 1.

```
[Sysname] qos policy 1
[Sysname-policy-1] classifier 1 behavior 1
[Sysname-policy-1] quit
```

Apply the policy in the inbound direction of GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos apply policy 1 inbound
```

After the configurations, you can monitor all packets sent from Host A on the data monitoring device.

Table of Contents

1 User Profile Configuration

When configuring user profile, go to these sections for information you are interested in:

User Profile Overview

User profile provides a configuration template to save predefined configurations. Based on different application scenarios, you can configure different items for a user profile, such as Committed Access Rate (CAR), Quality of Service (QoS), and so on.

When accessing the device, users need to be authenticated. During the authentication process, the authentication server sends the user profile name to the device, which then enables the configurations in the user profile. After the users pass the authentication and access the device, the device will restrict the users' access based on these configurations. When the users log out, the device automatically disables the configurations in the user profile, and thus the restrictions on the users are removed. Therefore, user profile is applicable to restricting online users' access; if no users are online (no user access, no users pass the authentication, or users have logged out), user profile does not take effect as it is a predefined configuration.

With user profile, you can:

- Make use of system resources more granularly. For example, without user profile, you can apply a QoS policy based on interface, VLAN, globally and so on. This QoS policy is applicable to a group of users. With user profile, however, you can apply a QoS policy on a per-user basis.
- Restrict users' access to the system resources more flexibly. For example, without user profile, you can perform traffic policing based on CAR, ACL, or for all the traffic of the current interface; when the physical position of users changes (for example, the users access the network using another interface), you need to configure traffic policing on another interface. With user profile, however, you can perform traffic policing on a per-user basis. As long as users are online, the authentication server applies the corresponding user profile (with CAR configured) to the users; when the users are offline, the system automatically removes the corresponding configuration.

User Profile Configuration

User Profile Configuration Task List

Task	Remarks
Creating a User Profile	Required
Applying a QoS Policy to User Profile	Required
Enabling a User Profile	Required

Creating a User Profile

Configuration Prerequisites

Before creating a user profile, you need to configure authentication parameters. User profile supports 802.1X and portal authentications. You can select one of them to authenticate users based on the actual networking when users access the network. However, you need to perform the related configurations (for example, username, password, authentication scheme, domain and binding between a user profile and user) on the client, the device and authentication server.

Creating a User Profile

Follow these steps to create a user profile:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a user profile, and enter the corresponding user profile view	Create a user profile, and enter user-profile DOT1X view	**user-profile** *profile-name* **dot1x**	Use one of the two approaches If the specified user profile already exists, you will directly enter the corresponding user profile view. The two user profile views respectively correspond to the three upper layer authentication types 802.1X and portal. The configuration made in user profile view takes effect when the user profile is enabled and the corresponding users are online.
	Create a user profile, and enter user-profile portal view	**user-profile** *profile-name* **portal**	

Note

Refer to *802.1X Configuration* and *Portal Configuration* in the *Security Volume* for detailed information about 802.1X authentication and portal authentication.

Applying a QoS Policy to User Profile

After a user profile is created, you need to configure detailed items in user profile view to implement restrictions on the online users. Currently supported configurations are as follows:

Follow these steps to apply a QoS policy to traffic of online users:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter user profile view	**user-profile** *profile-name* [**dot1x** | **portal**]	Required The configuration made in user profile view takes effect when the user-profile is active and the corresponding users are online.
Apply the QoS policy	**qos apply policy** *policy-name* { **inbound** | **outbound** }	Required

Note

- When a user profile is active, you cannot configure or remove the QoS policy applied to it.
- The QoS policies applied in user profile view support only the **remark**, **car**, and **filter** actions.
- Do not apply an empty QoS policy in user profile view, because even if you can do that, the user profile cannot be activated.

Enabling a User Profile

A created user profile takes effect only after being enabled.

Follow these steps to enable a user profile:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable a user profile	**user-profile** *profile-name* **enable**	Required A user profile is disabled by default.

Caution

- Only an enabled user profile can be used by a user. You cannot modify or remove the configuration items in a user profile until the user profile is disabled.
- Disabling a user profile logs out the users using the user profile.

Displaying and Maintaining User Profile

To do…	Use the command…
Display information about all the created user profiles	**display user-profile**

Security Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The Security Volume is organized as follows:

Features	Description
AAA	Authentication, Authorization and Accounting (AAA) provide a uniform framework used for configuring these three security functions to implement the network security management. This document describes: • Introduction to AAA, RADIUS and HWTACACS • AAA configuration • RADIUS configuration • HWTACACS configuration
802.1X	IEEE 802.1X (hereinafter simplified as 802.1X) is a port-based network access control protocol that is used as the standard for LAN user access authentication. This document describes: • 802.1X overview • 802.1X configuration • 802.1X Guest-VLAN configuration
HABP	On an HABP-capable switch, HABP packets can bypass 802.1X authentication and MAC authentication, allowing communication among switches in a cluster. This document describes: • Introduction to HABP • HABP configuration
MAC Authentication	MAC authentication provides a way for authenticating users based on ports and MAC addresses; it requires no client software to be installed on the hosts. This document describes: • RADIUS-Based MAC Authentication • Local MAC Authentication
Portal	Portal authentication, as its name implies, helps control access to the Internet. This document describes: • Portal overview • Portal configuration

Features	Description
Port Security	Port security is a MAC address-based security mechanism for network access controlling. It is an extension to the existing 802.1X authentication and MAC authentication. This document describes: • Enabling Port Security • Setting the Maximum Number of Secure MAC Addresses • Setting the Port Security Mode • Configuring Port Security Features • Configuring Secure MAC Addresses • Ignoring Authorization Information from the Server
IP Source Guard	By filtering packets on a per-port basis, IP source guard prevents illegal packets from traveling through, thus improving the network security. This document describes: • Configuring a Static Binding Entry • Configuring Dynamic Binding Function
SSH2.0	SSH ensures secure login to a remote device in a non-secure network environment. By encryption and strong authentication, it protects the device against attacks. This document describes: • Configuring Asymmetric Keys • Configuring the Device as an SSH Server • Configuring the Device as an SSH Client • Configuring an SFTP Server • Configuring an SFTP Client
PKI	The Public Key Infrastructure (PKI) is a hierarchical framework designed for providing information security through public key technologies and digital certificates and verifying the identities of the digital certificate owners. This document describes PKI related configuration.
SSL	Secure Sockets Layer (SSL) is a security protocol providing secure connection service for TCP-based application layer protocols, this document describes SSL related configuration.
Public Key Configuration	This document describes Public Key Configuration.
ACL	An ACL is used for identifying traffic based on a series of preset matching criteria. This document describes: • ACL overview and ACL types • ACL configuration

Table of Contents

1 AAA Configuration

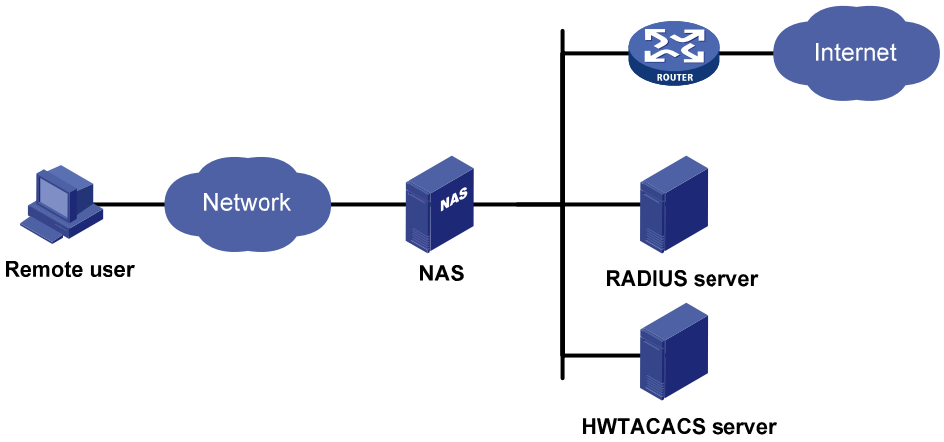
When configuring AAA, go to these sections for information you are interested in:

Introduction to AAA

Authentication, Authorization, and Accounting (AAA) provides a uniform framework for configuring these three security functions to implement network security management.

AAA usually uses a client/server model, where the client runs on the network access server (NAS) and the server maintains user information centrally. In an AAA network, a NAS is a server for users but a client for the AAA servers, as shown in Figure 1-1.

Figure 1-1 AAA networking diagram



When a user tries to establish a connection to the NAS and to obtain the rights to access other networks or some network resources, the NAS authenticates the user or the corresponding connection. The NAS can transparently pass the user's AAA information to the server (RADIUS server or HWTACACS server). The RADIUS/HWTACACS protocol defines how a NAS and a server exchange user information between them.

In the AAA network shown in Figure 1-1, there is a RADIUS server and a HWTACACS server. You can determine the authentication, authorization and accounting methods according to the actual

requirements. For example, you can use the HWTACACS server for authentication and authorization, and the RADIUS server for accounting.

The three security functions are described as follows:

- Authentication: Identifies remote users and judges whether a user is legal.
- Authorization: Grants different users different rights. For example, a user logging into the server can be granted the permission to access and print the files in the server.
- Accounting: Records all network service usage information of users, including the service type, start and end time, and traffic. In this way, accounting can be used for not only charging, but also network security surveillance.

You can use AAA to provide only one or two security functions, if desired. For example, if your company only wants employees to be authenticated before they access specific resources, you only need to configure an authentication server. If network usage information is expected to be recorded, you also need to configure an accounting server.

As described above, AAA provides a uniform framework to implement network security management. It is a security mechanism that enables authenticated and authorized entities to access specific resources and records operations of the entities. The AAA framework thus allows for excellent scalability and centralized user information management.

AAA can be implemented through multiple protocols. Currently, the device supports using RADIUS, HWTACACS for AAA, and RADIUS is often used in practice.

Introduction to RADIUS

Remote Authentication Dial-In User Service (RADIUS) is a distributed information interaction protocol in a client/server model. RADIUS can protect networks against unauthorized access and is often used in network environments where both high security and remote user access are required. Based on UDP, RADIUS uses UDP port 1812 for authentication and 1813 for accounting. RADIUS defines the RADIUS packet format and message transfer mechanism.

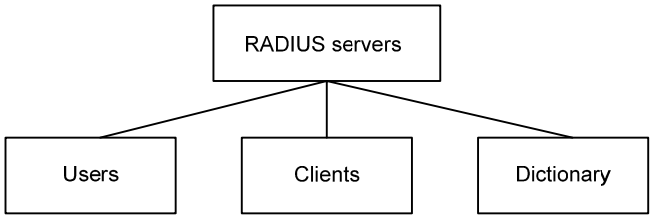
RADIUS was originally designed for dial-in user access. With the diversification of access methods, RADIUS has been extended to support more access methods, for example, Ethernet access and ADSL access. It uses authentication and authorization in providing access services and uses accounting to collect and record usage information of network resources.

Client/Server Model

- Client: The RADIUS client runs on the NASs located throughout the network. It passes user information to designated RADIUS servers and acts on the responses (for example, rejects or accepts user access requests).
- Server: The RADIUS server runs on the computer or workstation at the network center and maintains information related to user authentication and network service access. It listens to connection requests, authenticates users, and returns the processing results (for example, rejecting or accepting the user access request) to the clients.

In general, the RADIUS server maintains three databases, namely, Users, Clients, and Dictionary, as shown in Figure 1-2:

Figure 1-2 RADIUS server components



- Users: Stores user information such as the usernames, passwords, applied protocols, and IP addresses.
- Clients: Stores information about RADIUS clients, such as the shared keys and IP addresses.
- Dictionary: Stores information about the meanings of RADIUS protocol attributes and their values.

Security and Authentication Mechanisms

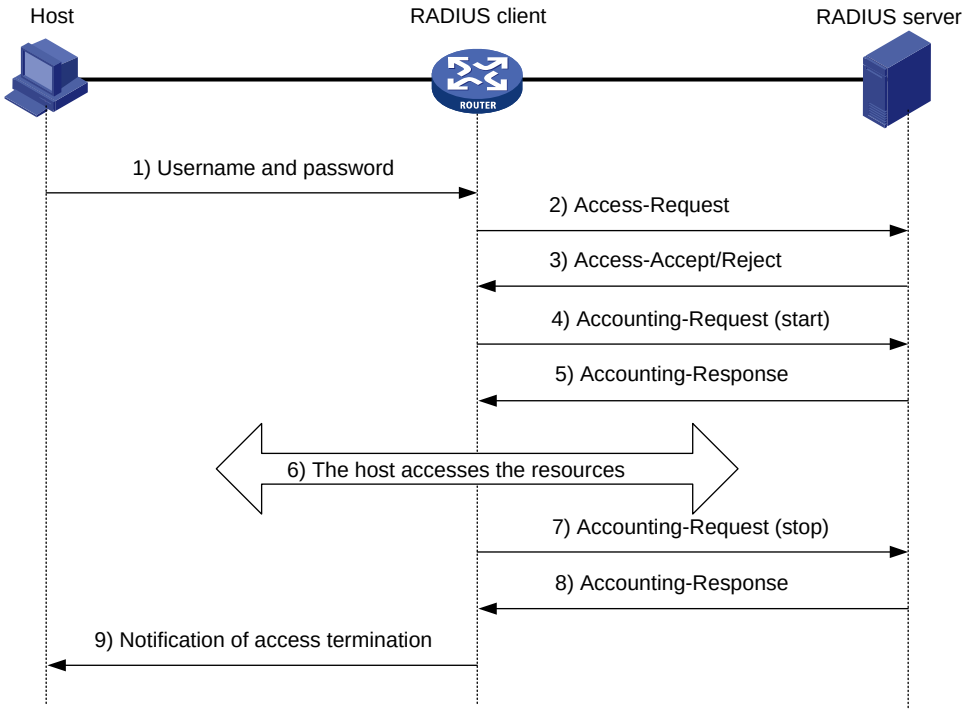
Information exchanged between a RADIUS client and the RADIUS server is authenticated with a shared key, which is never transmitted over the network. This enhances the information exchange security. In addition, to prevent user passwords from being intercepted in non-secure networks, RADIUS encrypts passwords before transmitting them.

A RADIUS server supports multiple user authentication methods, for example, the Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP). Moreover, a RADIUS server can act as the client of another AAA server to provide authentication proxy services.

Basic Message Exchange Process of RADIUS

Figure 1-3 illustrates the interaction of the host, the RADIUS client, and the RADIUS server.

Figure 1-3 Basic message exchange process of RADIUS



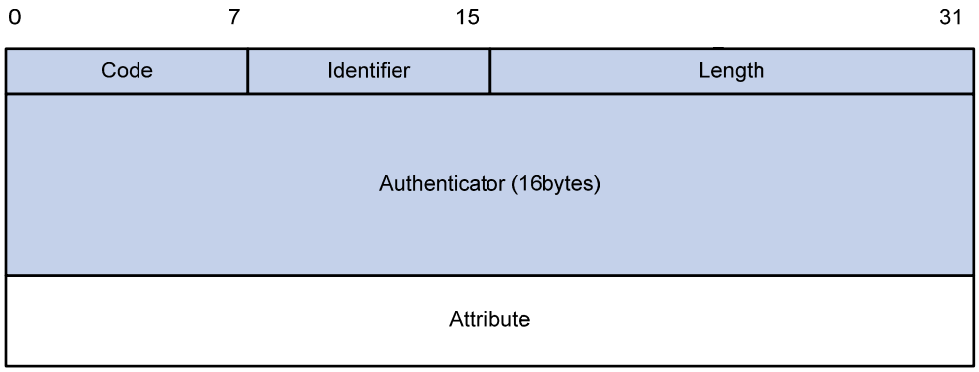
The following is how RADIUS operates:

1) The host initiates a connection request carrying the username and password to the RADIUS client.
2) Having received the username and password, the RADIUS client sends an authentication request (Access-Request) to the RADIUS server, with the user password encrypted by using the Message-Digest 5 (MD5) algorithm and the shared key.
3) The RADIUS server authenticates the username and password. If the authentication succeeds, it sends back an Access-Accept message containing the user's authorization information. If the authentication fails, it returns an Access-Reject message.
4) The RADIUS client permits or denies the user according to the returned authentication result. If it permits the user, it sends a start-accounting request (Accounting-Request) to the RADIUS server.
5) The RADIUS server returns a start-accounting response (Accounting-Response) and starts accounting.
6) The user accesses the network resources.
7) The host requests the RADIUS client to tear down the connection and the RADIUS client sends a stop-accounting request (Accounting-Request) to the RADIUS server.
8) The RADIUS server returns a stop-accounting response (Accounting-Response) and stops accounting for the user.
9) The user stops access to network resources.

RADIUS Packet Format

RADIUS uses UDP to transmit messages. It ensures the smooth message exchange between the RADIUS server and the client through a series of mechanisms, including the timer management mechanism, retransmission mechanism, and slave server mechanism. Figure 1-4 shows the RADIUS packet format.

Figure 1-4 RADIUS packet format



Descriptions of the fields are as follows:

1) The Code field (1-byte long) is for indicating the type of the RADIUS packet. Table 1-1 gives the possible values and their meanings.

Table 1-1 Main values of the Code field

Code	Packet type	Description
1	Access-Request	From the client to the server. A packet of this type carries user information for the server to authenticate the user. It must contain the User-Name attribute and can optionally contain the attributes of NAS-IP-Address, User-Password, and NAS-Port.
2	Access-Accept	From the server to the client. If all the attribute values carried in the Access-Request are acceptable, that is, the authentication succeeds, the server sends an Access-Accept response.

Code	Packet type	Description
3	Access-Reject	From the server to the client. If any attribute value carried in the Access-Request is unacceptable, the server rejects the user and sends an Access-Reject response.
4	Accounting-Request	From the client to the server. A packet of this type carries user information for the server to start/stop accounting for the user. It contains the Acct-Status-Type attribute, which indicates whether the server is requested to start the accounting or to end the accounting.
5	Accounting-Response	From the server to the client. The server sends to the client a packet of this type to notify that it has received the Accounting-Request and has correctly started recording the accounting information.

2) The Identifier field (1-byte long) is for matching request packets and response packets and detecting retransmitted request packets. The request and response packets of the same type have the same identifier.
3) The Length field (2-byte long) indicates the length of the entire packet, including the Code, Identifier, Length, Authenticator, and Attribute fields. The value of the field is in the range 20 to 4096. Bytes beyond the length are considered the padding and are neglected upon reception. If the length of a received packet is less than that indicated by the Length field, the packet is dropped.
4) The Authenticator field (16-byte long) is used to authenticate replies from the RADIUS server, and is also used in the password hiding algorithm. There are two kinds of authenticators: request authenticator and response authenticator.
5) The Attribute field, with a variable length, carries the specific authentication, authorization, and accounting information for defining configuration details of the request or response. This field is represented in triplets of Type, Length, and Value.

- Type: One byte, in the range 1 to 255. It indicates the type of the attribute. Commonly used attributes for RADIUS authentication, authorization and accounting are listed in Table 1-2.
- Length: One byte for indicating the length of the attribute in bytes, including the Type, Length, and Value fields.
- Value: Value of the attribute, up to 253 bytes. Its format and content depend on the Type and Length fields.

Table 1-2 RADIUS attributes

No.	Attribute	No.	Attribute
1	User-Name	45	Acct-Authentic
2	User-Password	46	Acct-Session-Time
3	CHAP-Password	47	Acct-Input-Packets
4	NAS-IP-Address	48	Acct-Output-Packets
5	NAS-Port	49	Acct-Terminate-Cause
6	Service-Type	50	Acct-Multi-Session-Id
7	Framed-Protocol	51	Acct-Link-Count
8	Framed-IP-Address	52	Acct-Input-Gigawords
9	Framed-IP-Netmask	53	Acct-Output-Gigawords
10	Framed-Routing	54	(unassigned)

No.	Attribute	No.	Attribute
11	Filter-ID	55	Event-Timestamp
12	Framed-MTU	56-59	(unassigned)
13	Framed-Compression	60	CHAP-Challenge
14	Login-IP-Host	61	NAS-Port-Type
15	Login-Service	62	Port-Limit
16	Login-TCP-Port	63	Login-LAT-Port
17	(unassigned)	64	Tunnel-Type
18	Reply_Message	65	Tunnel-Medium-Type
19	Callback-Number	66	Tunnel-Client-Endpoint
20	Callback-ID	67	Tunnel-Server-Endpoint
21	(unassigned)	68	Acct-Tunnel-Connection
22	Framed-Route	69	Tunnel-Password
23	Framed-IPX-Network	70	ARAP-Password
24	State	71	ARAP-Features
25	Class	72	ARAP-Zone-Access
26	Vendor-Specific	73	ARAP-Security
27	Session-Timeout	74	ARAP-Security-Data
28	Idle-Timeout	75	Password-Retry
29	Termination-Action	76	Prompt
30	Called-Station-Id	77	Connect-Info
31	Calling-Station-Id	78	Configuration-Token
32	NAS-Identifier	79	EAP-Message
33	Proxy-State	80	Message-Authenticator
34	Login-LAT-Service	81	Tunnel-Private-Group-id
35	Login-LAT-Node	82	Tunnel-Assignment-id
36	Login-LAT-Group	83	Tunnel-Preference
37	Framed-AppleTalk-Link	84	ARAP-Challenge-Response
38	Framed-AppleTalk-Network	85	Acct-Interim-Interval
39	Framed-AppleTalk-Zone	86	Acct-Tunnel-Packets-Lost
40	Acct-Status-Type	87	NAS-Port-Id
41	Acct-Delay-Time	88	Framed-Pool
42	Acct-Input-Octets	89	(unassigned)
43	Acct-Output-Octets	90	Tunnel-Client-Auth-id
44	Acct-Session-Id	91	Tunnel-Server-Auth-id

Note

The attribute types listed in Table 1-2 are defined by RFC 2865, RFC 2866, RFC 2867, and RFC 2568.

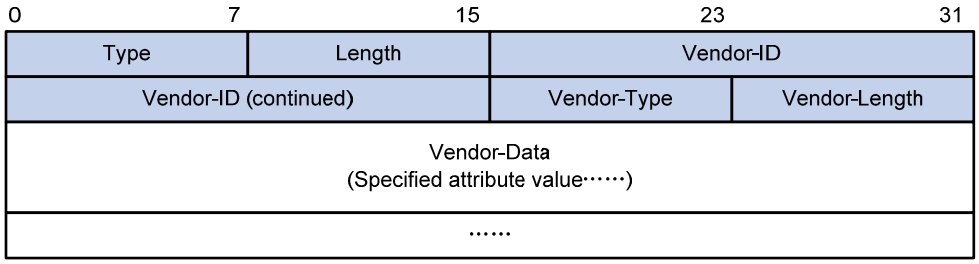
Extended RADIUS Attributes

The RADIUS protocol features excellent extensibility. Attribute 26 (Vender-Specific) defined by RFC 2865 allows a vender to define extended attributes to implement functions that the standard RADIUS protocol does not provide.

A vendor can encapsulate multiple type-length-value (TLV) sub-attributes in RADIUS packets for extension in applications. As shown in Figure 1-5, a sub-attribute that can be encapsulated in Attribute 26 consists of the following four parts:

- Vendor-ID (four bytes): Indicates the ID of the vendor. Its most significant byte is 0 and the other three bytes contain a code complying with RFC 1700.
- Vendor-Type: Indicates the type of the sub-attribute.
- Vendor-Length: Indicates the length of the sub-attribute.
- Vendor-Data: Indicates the contents of the sub-attribute.

Figure 1-5 Segment of a RADIUS packet containing an extended attribute



Introduction to HWTACACS

HW Terminal Access Controller Access Control System (HWTACACS) is an enhanced security protocol based on TACACS (RFC 1492). Similar to RADIUS, it uses a client/server model for information exchange between NAS and HWTACACS server.

HWTACACS is mainly used to provide AAA services for terminal users. In a typical HWTACACS application, a terminal user needs to log into the device for operations, and HWTACACS authenticates, authorizes and keeps accounting for the user. Working as the HWTACACS client, the device sends the username and password to the HWTACACS sever for authentication. After passing authentication and being authorized, the user can log into the device to perform operations.

Differences Between HWTACACS and RADIUS

HWTACACS and RADIUS have many common features, like implementing AAA, using a client/server model, using shared keys for user information security and having good flexibility and extensibility. Meanwhile, they also have differences, as listed in Table 1-3.

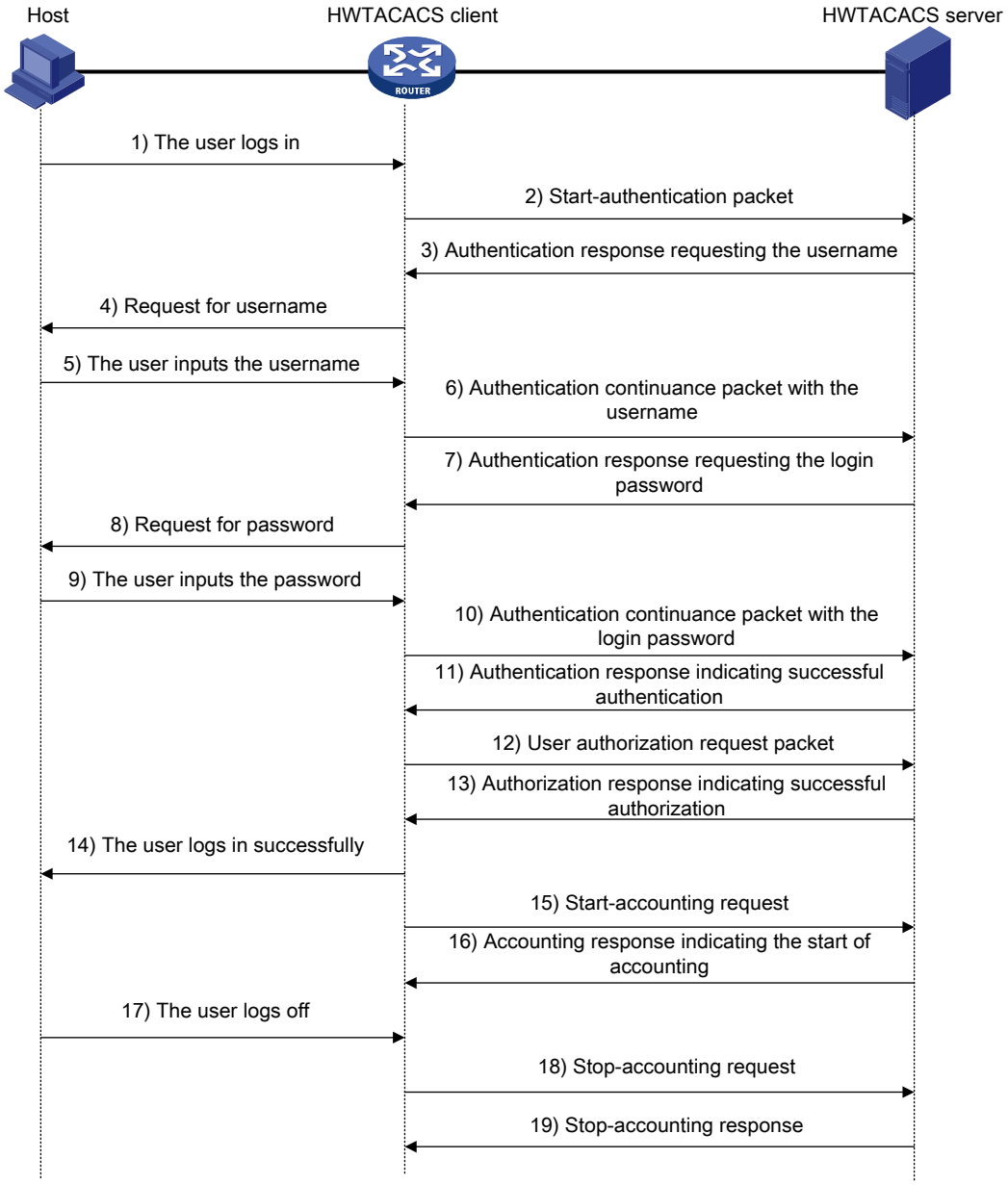
Table 1-3 Primary differences between HWTACACS and RADIUS

HWTACACS	RADIUS
Uses TCP, providing more reliable network transmission.	Uses UDP, providing higher transport efficiency.
Encrypts the entire packet except for the HWTACACS header.	Encrypts only the user password field in an authentication packet.
Protocol packets are complicated and authorization is independent of authentication. Authentication and authorization can be deployed on different HWTACACS servers.	Protocol packets are simple and authorization is combined with authentication.
Supports authorized use of configuration commands. For example, an authenticated login user can be authorized to configure the device.	Does not support authorized use of configuration commands.

Basic Message Exchange Process of HWTACACS

The following takes a Telnet user as an example to describe how HWTACACS performs user authentication, authorization, and accounting. Figure 1-6 illustrates the basic message exchange process of HWTACACS.

Figure 1-6 Basic message exchange process of HWTACACS for a Telnet user



1) A Telnet user sends an access request to the NAS.
2) Upon receiving the request, the HWTACACS client sends a start-authentication packet to the HWTACACS server.
3) The HWTACACS server sends back an authentication response requesting the username.
4) Upon receiving the response, the HWTACACS client asks the user for the username.
5) The user inputs the username.
6) After receiving the username from the user, the HWTACACS client sends to the server a continue-authentication packet carrying the username.
7) The HWTACACS server sends back an authentication response, requesting the login password.
8) Upon receipt of the response, the HWTACACS client asks the user for the login password.
9) The user inputs the password.
10) After receiving the login password, the HWTACACS client sends to the HWTACACS server a continue-authentication packet carrying the login password.
11) The HWTACACS server sends back an authentication response indicating that the user has passed authentication.

12) The HWTACACS client sends the user authorization request packet to the HWTACACS server.
13) The HWTACACS server sends back the authorization response, indicating that the user is authorized now.
14) Knowing that the user is now authorized, the HWTACACS client pushes the configuration interface of the NAS to the user.
15) The HWTACACS client sends a start-accounting request to the HWTACACS server.
16) The HWTACACS server sends back an accounting response, indicating that it has received the start-accounting request.
17) The user logs off.
18) The HWTACACS client sends a stop-accounting request to the HWTACACS server.
19) The HWTACACS server sends back a stop-accounting response, indicating that the stop-accounting request has been received.

Protocols and Standards

The protocols and standards related to AAA, RADIUS, HWTACACS include:

- RFC 2865: Remote Authentication Dial In User Service (RADIUS)
- RFC 2866: RADIUS Accounting
- RFC 2867: RADIUS Accounting Modifications for Tunnel Protocol Support
- RFC 2868: RADIUS Attributes for Tunnel Protocol Support
- RFC 2869: RADIUS Extensions
- RFC 1492: An Access Control Protocol, Sometimes Called TACACS

AAA Configuration Task List

The basic procedure to configure AAA is as follows:

1) Configure the required AAA schemes.
- Local authentication: Configure local users and related attributes, including usernames and passwords of the users to be authenticated.
- Remote authentication: Configure the required RADIUS and/or HWTACACS schemes, and configure user attributes on the servers accordingly.

2) Configure the AAA methods: Reference the configured AAA schemes in the users' ISP domains.
- Authentication method: No authentication (**none**), local authentication (**local**), or remote authentication (**scheme**)
- Authorization method: No authorization (**none**) , local authorization (**local**), or remote authorization (**scheme**)
- Accounting method: No accounting (**none**), local accounting (**local**), or remote accounting (**scheme**)

 **Note**

For login users, it is necessary to configure the authentication mode for logging into the user interface as **scheme**. For detailed information, refer to *Login Configuration* of the *System Volume*.

AAA Configuration Task List

Task	Remarks
Creating an ISP Domain	Required
Configuring ISP Domain Attributes	Optional
Configuring AAA Authentication Methods for an ISP Domain	Required For local authentication, refer to Configuring Local User Attributes. For RADIUS authentication, refer to Configuring RADIUS. For HWTACACS authentication, refer to Configuring HWTACACS.
Configuring AAA Authorization Methods for an ISP Domain	Optional
Configuring AAA Accounting Methods for an ISP Domain	Optional
Configuring Local User Attributes	Optional
Configuring User Group Attributes	Optional
Tearing down User Connections Forcibly	Optional
Displaying and Maintaining AAA	Optional

RADIUS Configuration Task List

Task	Remarks
Creating a RADIUS Scheme	Required
Specifying the RADIUS Authentication/Authorization Servers	Required
Specifying the RADIUS Accounting Servers and Relevant Parameters	Optional
Setting the Shared Key for RADIUS Packets	Required
Setting the Upper Limit of RADIUS Request Retransmission Attempts	Optional
Setting the Supported RADIUS Server Type	Optional
Setting the Status of RADIUS Servers	Optional
Configuring Attributes Related to Data to Be Sent to the RADIUS Server	Optional
Setting Timers Regarding RADIUS Servers	Optional
Specifying Security Policy Servers	Optional
Enabling the Listening Port of the RADIUS Client	Optional
Displaying and Maintaining RADIUS	Optional

HWTACACS Configuration Task List

Task	Remarks
Creating a HWTACACS scheme	Required
Specifying the HWTACACS Authentication Servers	Required
Specifying the HWTACACS Authorization Servers	Optional
Specifying the HWTACACS Accounting Servers	Optional
Setting the Shared Key for HWTACACS Packets	Required
Configuring Attributes Related to the Data Sent to HWTACACS Server	Optional
Setting Timers Regarding HWTACACS Servers	Optional
Displaying and Maintaining HWTACACS	Optional

Configuring AAA

By configuring AAA, you can provide network access service for legal users, protect the networking devices, and avoid unauthorized access and repudiation. In addition, you can configure ISP domains to perform AAA on accessing users.

In AAA, users are divided into LAN users (such as 802.1x users and MAC authentication users), login users (such as SSH, Telnet, FTP, and terminal access users), portal users and command line users (that is, command line authentication users). Except for command line users, you can configure separate authentication/authorization/accounting policies for all the other types of users. Command line users can be configured with authorization policy independently.

Configuration Prerequisites

For remote authentication, authorization, or accounting, you must create the RADIUS or HWTACACS scheme first. For RADIUS scheme configuration, refer to Configuring RADIUS. For HWTACACS scheme configuration, refer to Configuring HWTACACS.

Creating an ISP Domain

An Internet service provider (ISP) domain represents a group of users belonging to it. For a username in the *userid@isp-name* format, the access device considers the *userid* part the username for authentication and the *isp-name* part the domain name.

In a networking scenario with multiple ISPs, an access device may connect users of different ISPs. As users of different ISPs may have different user attributes (such as username and password structure, service type, and rights), you need to configure ISP domains to distinguish the users. In addition, you need to configure different attribute sets including AAA methods for the ISP domains.

For the NAS, each user belongs to an ISP domain. Up to 16 ISP domains can be configured on a NAS. If a user does not provide the ISP domain name, the system considers that the user belongs to the default ISP domain.

Follow these steps to create an ISP domain:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an ISP domain and enter ISP domain view	**domain** *isp-name*	Required
Return to system view	**quit**	—
Specify the default ISP domain	**domain default enable** *isp-name*	Optional By default, the system has a default ISP domain named **system**.

Note

- You cannot delete the default ISP domain unless you change it to a non-default ISP domain (with the **domain default disable** command) first.
- If a user enters a username without an ISP domain name, the device uses the authentication method configured for the default ISP domain to authenticate the user.

Configuring ISP Domain Attributes

Follow these steps to configure ISP domain attributes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an ISP domain and enter ISP domain view	**domain** *isp-name*	Required
Place the ISP domain to the state of active or blocked	**state** { **active** | **block** }	Optional When created, an ISP domain is in the active state by default, and users in the domain can request network services.
Specify the maximum number of active users in the ISP domain	**access-limit enable** *max-user-number*	Optional No limit by default
Configure the idle cut function	**idle-cut enable** *minute*	Optional Disabled by default Currently, this command is effective only for LAN users.
Configure the self-service server localization function	**self-service-url enable** *url-string*	Optional Disabled by default

Note

A self-service RADIUS server, for example, comprehensive access management system (CAMS/iMC), is required for the self-service server localization function to work. With the self-service function, a user can manage and control his or her accounting information or card number. A server with self-service software is a self-service server.

Configuring AAA Authentication Methods for an ISP Domain

In AAA, authentication, authorization, and accounting are separate processes. Authentication refers to the interactive authentication process of username/password/user information during access or service request. The authentication process neither sends authorization information to a supplicant nor triggers any accounting.

AAA supports the following authentication methods:

- No authentication: All users are trusted and no authentication is performed. Generally, this method is not recommended.
- Local authentication: Authentication is performed by the NAS, which is configured with the user information, including the usernames, passwords, and attributes. Local authentication features high speed and low cost, but the amount of information that can be stored is limited by the hardware.
- Remote authentication: The access device cooperates with a RADIUS or HWTACACS server to authenticate users. As for RADIUS, the device can use the standard RADIUS protocol or extended RADIUS protocol in collaboration with systems like CAMS/iMC to implement user authentication. Remote authentication features centralized information management, high capacity, high reliability, and support for centralized authentication for multiple devices. You can configure local authentication as the backup method in case the remote server is not available.

You can configure AAA authentication to work alone without authorization and accounting. By default, an ISP domain uses the local authentication method.

Before configuring authentication methods, complete these three tasks:

- For RADIUS or HWTACACS authentication, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication methods do not require any scheme.
- Determine the access mode or service type to be configured. With AAA, you can configure an authentication method specifically for each access mode and service type, limiting the authentication protocols that can be used for access.
- Determine whether to configure an authentication method for all access modes or service types.

Follow these steps to configure AAA authentication methods for an ISP domain:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an ISP domain and enter ISP domain view	**domain** *isp-name*	Required

To do…	Use the command…	Remarks
Specify the default authentication method for all types of users	**authentication default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional **local** by default
Specify the authentication method for LAN users	**authentication lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authentication method is used by default.
Specify the authentication method for login users	**authentication login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authentication method is used by default.
Specify the authentication method for portal users	**authentication portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authentication method is used by default.

Note

- The authentication method specified with the **authentication default** command is for all types of users and has a priority lower than that for a specific access mode.
- With an authentication method that references a RADIUS scheme, AAA accepts only the authentication result from the RADIUS server. The Access-Accept message from the RADIUS server does include the authorization information, but the authentication process ignores the information.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* **local** keyword and argument combination configured, local authentication is the backup method and is used only when the remote server is not available.
- If the primary authentication method is **local** or **none**, the system performs local authentication or does not perform any authentication, and will not use any RADIUS or HWTACACS authentication scheme.

Configuring AAA Authorization Methods for an ISP Domain

In AAA, authorization is a separate process at the same level as authentication and accounting. Its responsibility is to send authorization requests to the specified authorization server and to send authorization information to users. Authorization method configuration is optional in AAA configuration.

AAA supports the following authorization methods:

- No authorization: Every user is trusted and has the corresponding default rights of the system.
- Local authorization: Users are authorized by the access device according to the attributes configured for them.
- Remote authorization: The access device cooperates with a RADIUS or HWTACACS server to authorize users. RADIUS authorization is bound with RADIUS authentication. RADIUS

authorization can work only after RADIUS authentication is successful, and the authorization information is carried in the Access-Accept message. HWTACACS authorization is separate from HWTACACS authentication, and the authorization information is carried in the authorization response after successful authentication. You can configure local authorization or no authorization as the backup method in case the remote server is not available.

By default, an ISP domain uses the local authorization method. If the no authorization method (**none**) is configured, the users are not required to be authorized, in which case an authenticated user has the default right. The default right is visiting (the lowest one) for EXEC users (that is, console users who use the console, AUX port, or Telnet to connect to the device, such as Telnet or SSH users. Each connection of these types is called an EXEC user). The default right for FTP users is to use the root directory of the device.

Before configuring authorization methods, complete these three tasks:

1) For HWTACACS authorization, configure the HWTACACS scheme to be referenced first. For RADIUS authorization, the RADIUS authorization scheme must be the same as the RADIUS authentication scheme; otherwise, it does not take effect.
2) Determine the access mode or service type to be configured. With AAA, you can configure an authorization scheme specifically for each access mode and service type, limiting the authorization protocols that can be used for access.
3) Determine whether to configure an authorization method for all access modes or service types.

Follow these steps to configure AAA authorization methods for an ISP domain:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an ISP domain and enter ISP domain view	**domain** *isp*-name	Required
Specify the default authorization method for all types of users	**authorization default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional **local** by default
Specify the authorization method for command line users	**authorization command** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local** | **none**] | **local** | **none** }	Optional The default authorization method is used by default.
Specify the authorization method for LAN users	**authorization lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authorization method is used by default.
Specify the authorization method for login users	**authorization login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authorization method is used by default.
Specify the authorization method for portal users	**authorization portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default authorization method is used by default.

Note

- The authorization method specified with the **authorization default** command is for all types of users and has a priority lower than that for a specific access mode.
- RADIUS authorization is special in that it takes effect only when the RADIUS authorization scheme is the same as the RADIUS authentication scheme. In addition, if a RADIUS authorization fails, the error message returned to the NAS says that the server is not responding.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* [**local** | **none**] keyword and argument combination configured, local authorization or no authorization is the backup method and is used only when the remote server is not available.
- If the primary authorization method is **local** or **none**, the system performs local authorization or does not perform any authorization; it will never use the RADIUS or HWTACACS authorization scheme.
- The authorization information of the RADIUS server is sent to the RADIUS client along with the authentication response message; therefore, you cannot specify a separate RADIUS authorization server. If you use RADIUS for authorization and authentication, you must use the same scheme setting for authorization and authentication; otherwise, the system will prompt you with an error message.

Configuring AAA Accounting Methods for an ISP Domain

In AAA, accounting is a separate process at the same level as authentication and authorization. Its responsibility is to send accounting start/update/end requests to the specified accounting server. Accounting is not required, and therefore accounting method configuration is optional.

AAA supports the following accounting methods:

- No accounting: The system does not perform accounting for the users.
- Local accounting: Local accounting is implemented on the access device. It is for collecting statistics on the number of users and controlling the number of local user connections; it does not provide statistics for user charge.
- Remote accounting: The access device cooperates with a RADIUS server or HWTACACS server for accounting of users. You can configure local accounting as the backup method in case the remote server is not available.

By default, an ISP domain uses the local accounting method.

Before configuring accounting methods, complete these three tasks:

1) For RADIUS or HWTACACS accounting, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication methods do not require any scheme.
2) Determine the access mode or service type to be configured. With AAA, you can configure an accounting method specifically for each access mode and service type, limiting the accounting protocols that can be used for access.
3) Determine whether to configure an accounting method for all access modes or service types.

Follow these steps to configure AAA accounting methods for an ISP domain:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an ISP domain and enter ISP domain view	**domain** *isp-name*	Required
Enable the accounting optional feature	**accounting optional**	Optional Disabled by default
Specify the default accounting method for all types of users	**accounting default** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional **local** by default
Specify the accounting method for LAN users	**accounting lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default accounting method is used by default.
Specify the accounting method for login users	**accounting login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default accounting method is used by default.
Specify the accounting method for portal users	**accounting portal** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional The default accounting method is used by default.

Note

- With the **accounting optional** command configured, a user to be disconnected can still use the network resources even when there is no available accounting server or communication with the current accounting server fails.
- The local accounting is not used for accounting implementation, but together with the **attribute access-limit** command for limiting the number of local user connections. However, with the **accounting optional** command configured, the limit on the number of local user connections is not effective.
- The accounting method specified with the **accounting default** command is for all types of users and has a priority lower than that for a specific access mode.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* **local** keyword and argument combination configured, local accounting is the backup method and is used only when the remote server is not available.
- If the primary accounting method is **local** or **none**, the system performs local accounting or does not perform any accounting, and will not use the RADIUS or HWTACACS accounting scheme.
- In login access mode, accounting is not supported for FTP services.

Configuring Local User Attributes

For local authentication, you need to create local users and configure user attributes on the device as needed.

A local user represents a set of user attributes configured on a device, and such a user set is uniquely identified by the username. For a user requesting network service to pass local authentication, you must add an entry as required in the local user database of the device.

Each local user belongs to a local user group and bears all attributes of the group, such as the password control attributes and authorization attributes. For details about local user group, refer to Configuring User Group Attributes.

When configuring local users and local user groups, pay attention to the effective ranges and priority relationship of user group attributes and user attributes:

- Authorization attributes

You can configure an authorization attribute in user group view or local user view, making the attribute effective on all local users of the group or only the local user. An authorization attribute configured in local user view takes precedence over the same attribute configured in user group view.

Follow these steps to configure the attributes for a local user:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the password display mode for all local users	**local-user password-display-mode** { **auto** | **cipher-force** }	Optional **auto** by default, indicating to display the password of a local user in the way indicated by the **password** command.
Add a local user and enter local user view	**local-user** *user-name*	Required No local user exists by default.
Configure a password for the local user	**password** { **cipher** | **simple** } *password*	Optional
Place the local user to the state of active or blocked	**state** { **active** | **block** }	Optional When created, a local user is in the state of active by default, and the user can request network services.
Set the maximum number of user connections using the local user account	**access-limit** *max-user-number*	Optional By default, there is no limit on the maximum number of user connections using the same local user account.
Specify the service types for the local user	**service-type** { **ftp** | **lan-access** | { **ssh** | **telnet** | **terminal** } * | **portal** }	Optional By default, no service is authorized to a user.

To do…	Use the command…	Remarks
Configure the binding attributes for the local user	**bind-attribute** { **call-number** *call-number* [: *subcall-number*] | **ip** *ip-address* | **location port** *slot-number subslot-number port-number* | **mac** *mac-address* | **vlan** *vlan-id* } *	Optional By default, no binding attribute is configured for a local user.
Configure the authorization attributes for the local user	**authorization-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	Optional By default, no authorization attribute is configured for a local user.
Set the expiration time of the user	**expiration-date** *time*	Optional Not set by default
Specify the user group for the local user	**group** *group-name*	Optional By default, a local user belongs to default user group **system**.

Note that:

- With the **local-user password-display-mode cipher-force** command configured, a local user password is always displayed in cipher text, regardless of the configuration of the **password** command. In this case, if you use the **save** command to save the configuration, all existing local user passwords will still be displayed in cipher text after the device restarts, even if you restore the display mode to **auto**.
- The **access-limit** command configured for a local user takes effect only when local accounting is used.
- Local authentication checks the service types of a local user. If the service types are not available, the user cannot pass authentication.
- In the authentication method that requires the username and password, including local authentication, RADIUS authentication and HWTACACS authentication, the commands that a login user can use after logging in depend on the level of the user. In other authentication methods, which commands are available depends on the level of the user interface. For an SSH user using public key authentication, the commands that can be used depend on the level configured on the user interface. For details regarding authentication method and commands accessible to user interface, refer to *Login Configuration* in the *System Volume.*
- Binding attributes are checked upon authentication of a local user. If the checking fails, the user fails the authentication. Therefore, be cautious when deciding which binding attributes should be configured for a local user.
- Every configurable authorization attribute has its definite application environments and purposes. Therefore, when configuring authorization attributes for a local user, consider what attributes are needed.

Configuring User Group Attributes

For simplification of local user configuration and manageability of local users, the concept of user group is introduced. A user group consists of a group of local users and has a set of local user attributes. You can configure local user attributes for a user group to implement centralized management of user

attributes for the local users in the group. Currently, you can configure password control attributes and authorization attributes for a user group.

By default, every newly added local user belongs to the user group of system and bears all attributes of the group. User group system is automatically created by the device.

Follow these steps to configure the attributes for a user group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a user group and enter user group view	**user-group** *group-name*	Required
Configure the authorization attributes for the user group	**authorization-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	Optional By default, no authorization attribute is configured for a user group.

Tearing down User Connections Forcibly

Follow these steps to tear down user connections forcibly:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Tear down AAA user connections forcibly	**cut connection** { **all** | **domain** *isp-name* | **ucibindex** *ucib-index* | **user-name** *user-name* } [**slot** *slot-number*]	Required Applies to only LAN access and portal user connections at present

Displaying and Maintaining AAA

To do…	Use the command…	Remarks
Display the configuration information of a specified ISP domain or all ISP domains	**display domain** [*isp-name*]	Available in any view
Display information about specified or all user connections	**display connection** [**domain** *isp-name* | **ucibindex** *ucib-index* | **user-name** *user-name*] [**slot** *slot-number*]	Available in any view
Display information about specified or all local users	**display local-user** [**idle-cut** { **disable** | **enable** } | **service-type** { **ftp** | **lan-access** | **portal** | **ssh** | **telnet** | **terminal** } | **state** { **active** | **block** } | **user-name** *user-name* | **vlan** *vlan-id*] [**slot** *slot-number*]	Available in any view
Display configuration information about a specified user group or all user groups	**display user-group** [*group-name*]	Available in any view

Configuring RADIUS

The RADIUS protocol is configured on a per scheme basis. After creating a RADIUS scheme, you need to configure the IP addresses and UDP ports of the RADIUS servers for the scheme. The servers include authentication/authorization servers and accounting servers, or primary servers and secondary servers. In other words, the attributes of a RADIUS scheme mainly include IP addresses of primary and secondary servers, shared key, and RADIUS server type.

Actually, the RADIUS protocol configurations only set the parameters necessary for the information interaction between a NAS and a RADIUS server. For these settings to take effect, you must reference the RADIUS scheme containing those settings in ISP domain view. For information about the commands for referencing a scheme, refer to Configuring AAA.

Note

When there are users online, you cannot modify RADIUS parameters other than the retransmission ones and the timers.

Creating a RADIUS Scheme

Before performing other RADIUS configurations, follow these steps to create a RADIUS scheme and enter RADIUS scheme view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default

Note

A RADIUS scheme can be referenced by more than one ISP domain at the same time.

Specifying the RADIUS Authentication/Authorization Servers

Follow these steps to specify the RADIUS authentication/authorization servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default

To do…	Use the command…	Remarks
Specify the primary RADIUS authentication/authorization server	**primary authentication** *ip-address* [*port-number*]	Required Configure at least one of the commands No authentication server by default
Specify the secondary RADIUS authentication/authorization server	**secondary authentication** *ip-address* [*port-number*]	

Note

- It is recommended to specify only the primary RADIUS authentication/authorization server if backup is not required.
- If both the primary and secondary authentication/authorization servers are specified, the secondary one is used when the primary one is unreachable.
- In practice, you may specify two RADIUS servers as the primary and secondary authentication/authorization servers respectively. At one time, a server can be the primary authentication/authorization server for a scheme and the secondary authentication/authorization servers for another scheme.
- The IP addresses of the primary and secondary authentication/authorization servers for a scheme cannot be the same. Otherwise, the configuration fails.

Specifying the RADIUS Accounting Servers and Relevant Parameters

Follow these steps to specify the RADIUS accounting servers and perform related configurations:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Specify the primary RADIUS accounting server	**primary accounting** *ip-address* [*port-number*]	Required Configure at least one of the commands No accounting server by default
Specify the secondary RADIUS accounting server	**secondary accounting** *ip-address* [*port-number*]	
Enable the device to buffer stop-accounting requests getting no responses	**stop-accounting-buffer enable**	Optional Enabled by default
Set the maximum number of stop-accounting request transmission attempts	**retry stop-accounting** *retry-times*	Optional 500 by default
Set the maximum number of accounting request transmission attempts	**retry realtime-accounting** *retry-times*	Optional 5 by default

Note

- It is recommended to specify only the primary RADIUS accounting server if backup is not required.
- If both the primary and secondary accounting servers are specified, the secondary one is used when the primary one is not reachable.
- In practice, you can specify two RADIUS servers as the primary and secondary accounting servers respectively, or specify one server to function as the primary accounting server in a scheme and the secondary accounting server in another scheme. Besides, because RADIUS uses different UDP ports to receive authentication/authorization and accounting packets, the port for authentication/authorization must be different from that for accounting.
- You can set the maximum number of stop-accounting request transmission buffer, allowing the device to buffer and resend a stop-accounting request until it receives a response or the number of transmission retries reaches the configured limit. In the latter case, the device discards the packet.
- You can set the maximum number of accounting request transmission attempts on the device, allowing the device to disconnect a user when the number of accounting request transmission attempts for the user reaches the limit but it still receives no response to the accounting request.
- The IP addresses of the primary and secondary accounting servers cannot be the same. Otherwise, the configuration fails.
- Currently, RADIUS does not support keeping accounts on FTP users.

Setting the Shared Key for RADIUS Packets

The RADIUS client and RADIUS server use the MD5 algorithm to encrypt packets exchanged between them and a shared key to verify the packets. Only when the same key is used can they properly receive the packets and make responses.

Follow these steps to set the shared key for RADIUS packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Set the shared key for RADIUS authentication/authorization or accounting packets	**key** { **accounting** | **authentication** } *string*	Required No key by default

Note

The shared key configured on the device must be the same as that configured on the RADIUS server.

Setting the Upper Limit of RADIUS Request Retransmission Attempts

Because RADIUS uses UDP packets to carry data, the communication process is not reliable. If a NAS receives no response from the RADIUS server before the response timeout timer expires, it is required

to retransmit the RADIUS request. If the number of transmission attempts exceeds the specified limit but it still receives no response, it considers that the authentication has failed.

Follow these steps to set the upper limit of RADIUS request retransmission attempts:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Set the number of retransmission attempts of RADIUS packets	**retry** *retry-times*	Optional 3 by default

Note

- The maximum number of retransmission attempts of RADIUS packets multiplied by the RADIUS server response timeout period cannot be greater than 75.
- Refer to the **timer response-timeout** command in the command manual for configuring RADIUS server response timeout period.

Setting the Supported RADIUS Server Type

Follow these steps to set the supported RADIUS server type:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Specify the RADIUS server type supported by the device	**server-type** { **extended** | **standard** }	Optional By default, the supported RADIUS server type is **standard**.

Note

- If you change the type of RADIUS server, the data stream destined to the original RADIUS server will be restored to the default unit.
- When a third-party RADIUS is used, you can configure the RADIUS server to **standard** or **extended**. When CAMS/iMC server is used, you must configure the RADIUS server to **extended.**

Setting the Status of RADIUS Servers

When a primary server fails, the device automatically tries to communicate with the secondary server.

When both the primary and secondary servers are available, the device sends request packets to the primary server.

Once the primary server fails, the primary server turns into the state of block, and the device turns to the secondary server. In this case:

- If the secondary server is available, the device triggers the primary server quiet timer. After the quiet timer times out, the status of the primary server is active again and the status of the secondary server remains the same.
- If the secondary server fails, the device restores the status of the primary server to active immediately.

If the primary server has resumed, the device turns to use the primary server and stops communicating with the secondary server. After accounting starts, the communication between the client and the secondary server remains unchanged.

Follow these steps to set the status of RADIUS servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Set the status of the primary RADIUS authentication/authorization server	**state primary authentication** { **active** | **block** }	Optional **active** for every server configured with IP address in the RADIUS scheme
Set the status of the primary RADIUS accounting server	**state primary accounting** { **active** | **block** }	
Set the status of the secondary RADIUS authentication/authorization server	**state secondary authentication** { **active** | **block** }	
Set the status of the secondary RADIUS accounting server	**state secondary accounting** { **active** | **block** }	

Note

- If both the primary server and the secondary server are in the blocked state, it is necessary to manually turn the secondary server to the active state so that the secondary server can perform authentication. If the secondary server is still in the blocked state, the primary/secondary switchover cannot take place.
- If one server is in the active state while the other is blocked, the primary/secondary switchover will not take place even if the active server is not reachable.
- The server status set by the **state** command cannot be saved in the configuration file and will be restored to **active** every time the server restarts.

Configuring Attributes Related to Data to Be Sent to the RADIUS Server

Follow these steps to configure the attributes related to data to be sent to the RADIUS server:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable the RADIUS trap function		**radius trap** { **accounting-server-down** | **authentication-server-down** }	Optional Disabled by default
Create a RADIUS scheme and enter RADIUS scheme view		**radius scheme** *radius-scheme-name*	Required Not defined by default
Specify the format of the username to be sent to a RADIUS server		**user-name-format** { **keep-original** | **with-domain** | **without-domain** }	Optional By default, the ISP domain name is included in the username.
Specify the unit for data flows or packets to be sent to a RADIUS server		**data-flow-format** { **data** { **byte** | **giga-byte** | **kilo-byte** | **mega-byte** } | **packet** { **giga-packet** | **kilo-packet** | **mega-packet** | **one-packet** } }*	Optional The defaults are as follows: **byte** for data flows, and **one-packet** for data packets.
Set the source IP address of the device to send RADIUS packets	In RADIUS scheme view	**nas-ip** *ip-address*	Use either command By default, the outbound port serves as the source IP address to send RADIUS packets
	In system view	**quit**	
		radius nas-ip *ip-address*	

Note

- Some earlier RADIUS servers cannot recognize usernames that contain an ISP domain name. In this case, the device must remove the domain name before sending a username including a domain name. You can configure the **user-name-format without-domain** command on the device for this purpose.
- If a RADIUS scheme defines that the username is sent without the ISP domain name, do not apply the RADIUS scheme to more than one ISP domain, thus avoiding the confused situation where the RADIUS server regards two users in different ISP domains but with the same userid as one.
- The unit of data flows sent to the RADIUS server must be consistent with the traffic statistics unit of the RADIUS server. Otherwise, accounting cannot be performed correctly.
- The **nas-ip** command in RADIUS scheme view is only for the current RADIUS scheme, while the **radius nas-ip** command in system view is for all RADIUS schemes. However, the **nas-ip** command in RADIUS scheme view takes precedence over the **radius nas-ip** command.

Setting Timers Regarding RADIUS Servers

When communicating with the RADIUS server, a device can enable the following three timers:

- RADIUS server response timeout (**response-timeout**): If a NAS receives no response from the RADIUS server in a period of time after sending a RADIUS request (authentication/authorization or accounting request), it has to resend the request so that the user has more opportunity to obtain the RADIUS service. The NAS uses the RADIUS server response timeout timer to control the transmission interval.

- Primary server quiet timer (**timer quiet**): If the primary server is not reachable, its state changes to blocked, and the device will turn to the specified secondary server. If the secondary server is reachable, the device starts this timer and communicates with the secondary server. After this timer expires, the device turns the state of the primary server to active and tries to communicate with the primary server while keeping the state of the secondary server unchanged. If the primary server has come back into operation, the device interacts with the primary server and terminates its communication with the secondary server.
- Real-time accounting interval (**realtime-accounting**): This timer defines the interval for performing real-time accounting of users. After this timer is set, the switch will send accounting information of online users to the RADIUS server at the specified interval.

Follow these steps to set timers regarding RADIUS servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required Not defined by default
Set the RADIUS server response timeout timer	**timer response-timeout** *seconds*	Optional 3 seconds by default
Set the quiet timer for the primary server	**timer quiet** *minutes*	Optional 5 minutes by default
Set the real-time accounting interval	**timer realtime-accounting** *minutes*	Optional 12 minutes by default

Note

- The maximum number of retransmission attempts of RADIUS packets multiplied by the RADIUS server response timeout period cannot be greater than 75. This product is also the upper limit of the timeout time of different access modules.
- For an access module, the maximum number of retransmission attempts multiplied by the RADIUS server response timeout period must be smaller than the timeout time. Otherwise, stop-accounting messages cannot be buffered, and the primary/secondary server switchover cannot take place. For example, as the timeout time of voice access is 10 seconds, the product of the two parameters cannot exceed 10 seconds; as the timeout time of Telnet access is 30 seconds, the product of the two parameters cannot exceed 30 seconds. For detailed information about timeout time of a specific access module, refer to the corresponding part in the *Access Volume.*
- To configure the maximum number of retransmission attempts of RADIUS packets, refer to the command **retry** in the command manual.

Specifying Security Policy Servers

The core of the EAD solution is integration and cooperation, and the security policy server system is the management and control center. As a collection of software, the security policy server system can run on Windows and Linux to provide functions such as user management, security policy management, security status assessment, security cooperation control, and security event audit.

Follow these steps to specify a security policy server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a RADIUS scheme and enter its view	**radius scheme** *radius-scheme-name*	Required By default, no RADIUS scheme is present.
Specify a security policy server	**security-policy-server** *ip-address*	Optional Not specified by default

Note

- If more than one interface of the device is configured with user access authentication functions, the interfaces may use different security policy servers. You can specify up to eight security policy servers for a RADIUS scheme.
- If the RADIUS server and the security policy server reside on the same physical device, you do not need to configure the IP address of the security policy server.
- The specified security policy server must be a security policy server or RADIUS server that is correctly configured and working normally. Otherwise, the device will regard it as an illegal server.

Enabling the Listening Port of the RADIUS Client

Follow these steps to enable the listening port of the RADIUS client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the listening port of the RADIUS client	**radius client enable**	Optional Enabled by default

Displaying and Maintaining RADIUS

To do…	Use the command…	Remarks
Display the configuration information of a specified RADIUS scheme or all RADIUS schemes	**display radius scheme** [*radius-scheme-name*] [**slot** *slot-number*]	Available in any view
Display statistics about RADIUS packets	**display radius statistics** [**slot** *slot-number*]	Available in any view
Display information about buffered stop-accounting requests that get no responses	**display stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]	Available in any view
Clear RADIUS statistics	**reset radius statistics** [**slot** *slot-number*]	Available in user view

To do…	Use the command…	Remarks
Clear buffered stop-accounting requests that get no responses	**reset stop-accounting-buffer** { **radius-scheme** *radius-server-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]	Available in user view

Configuring HWTACACS

Note

Different from RADIUS, except for deleting HWTACACS schemes and changing the IP addresses of the HWTACACS servers, you can make any changes to HWTACACS parameters, whether there are users online or not.

Creating a HWTACACS scheme

The HWTACACS protocol is configured on a per scheme basis. Before performing other HWTACACS configurations, follow these steps to create a HWTACACS scheme and enter HWTACACS scheme view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default

Note

- Up to 16 HWTACACS schemes can be configured.
- A scheme can be deleted only when it is not referenced.

Specifying the HWTACACS Authentication Servers

Follow these steps to specify the HWTACACS authentication servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default

To do…	Use the command…	Remarks
Specify the primary HWTACACS authentication server	**primary authentication** *ip-address* [*port-number*]	Required Configure at least one of the commands No authentication server by default
Specify the secondary HWTACACS authentication server	**secondary authentication** *ip-address* [*port-number*]	

Note

- It is recommended to specify only the primary HWTACACS authentication server if backup is not required.
- If both the primary and secondary authentication servers are specified, the secondary one is used when the primary one is not reachable.
- The IP addresses of the primary and secondary authentication servers cannot be the same. Otherwise, the configuration fails.
- You can remove an authentication server only when no active TCP connection for sending authentication packets is using it.

Specifying the HWTACACS Authorization Servers

Follow these steps to specify the HWTACACS authorization servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default
Specify the primary HWTACACS authorization server	**primary authorization** *ip-address* [*port-number*]	Required Configure at least one of the commands No authorization server by default
Specify the secondary HWTACACS authorization server	**secondary authorization** *ip-address* [*port-number*]	

Note

- It is recommended to specify only the primary HWTACACS authorization server if backup is not required.
- If both the primary and secondary authorization servers are specified, the secondary one is used when the primary one is not reachable.
- The IP addresses of the primary and secondary authorization servers cannot be the same. Otherwise, the configuration fails.
- You can remove an authorization server only when no active TCP connection for sending authorization packets is using it.

Specifying the HWTACACS Accounting Servers

Follow these steps to specify the HWTACACS accounting servers and perform related configurations:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default
Specify the primary HWTACACS accounting server	**primary accounting** *ip-address* [*port-number*]	Required Configure at least one of the commands No accounting server by default
Specify the secondary HWTACACS accounting server	**secondary accounting** *ip-address* [*port-number*]	
Enable the device to buffer stop-accounting requests getting no responses	**stop-accounting-buffer enable**	Optional Enabled by default
Set the maximum number of stop-accounting request transmission attempts	**retry stop-accounting** *retry-times*	Optional 100 by default

Note

- It is recommended to specify only the primary HWTACACS accounting server if backup is not required.
- If both the primary and secondary accounting servers are specified, the secondary one is used when the primary one is not reachable.
- The IP addresses of the primary and secondary accounting servers cannot be the same. Otherwise, the configuration fails.
- You can remove an accounting server only when no active TCP connection for sending accounting packets is using it.
- Currently, HWTACACS does not support keeping accounts on FTP users.

Setting the Shared Key for HWTACACS Packets

When using a HWTACACS server as an AAA server, you can set a key to secure the communications between the device and the HWTACACS server.

The HWTACACS client and HWTACACS server use the MD5 algorithm to encrypt packets exchanged between them and a shared key to verify the packets. Only when the same key is used can they properly receive the packets and make responses.

Follow these steps to set the shared key for HWTACACS packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default
Set the shared keys for HWTACACS authentication, authorization, and accounting packets	**key** { **accounting** | **authentication** | **authorization** } *string*	Required No shared key exists by default.

Configuring Attributes Related to the Data Sent to HWTACACS Server

Follow these steps to configure the attributes related to the data sent to the HWTACACS server:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view		**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default
Specify the format of the username to be sent to a HWTACACS server		**user-name-format** { **keep-original** | **with-domain** | **without-domain** }	Optional By default, the ISP domain name is included in the username.
Specify the unit for data flows or packets to be sent to a HWTACACS server		**data-flow-format** { **data** { **byte** | **giga-byte** | **kilo-byte** | **mega-byte** } | **packet** { **giga-packet** | **kilo-packet** | **mega-packet** | **one-packet** } }*	Optional The defaults are as follows: **byte** for data flows, and **one-packet** for data packets.
Set the source IP address of the device to send HWTACACS packets	In HWTACACS scheme view	**nas-ip** *ip-address*	Use either command By default, the outbound port serves as the source IP address to send HWTACACS packets.
	In system view	**quit**	
		hwtacacs nas-ip *ip-address*	

Note

- If a HWTACACS server does not support a username with the domain name, you can configure the device to remove the domain name before sending the username to the server.
- The **nas-ip** command in HWTACACS scheme view is only for the current HWTACACS scheme, while the **hwtacacs nas-ip** command in system view is for all HWTACACS schemes. However, the **nas-ip** command in HWTACACS scheme view overwrites the configuration of the **hwtacacs nas-ip** command.

Setting Timers Regarding HWTACACS Servers

Follow these steps to set timers regarding HWTACACS servers:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a HWTACACS scheme and enter HWTACACS scheme view	**hwtacacs scheme** *hwtacacs-scheme-name*	Required Not defined by default
Set the HWTACACS server response timeout timer	**timer response-timeout** *seconds*	Optional 5 seconds by default
Set the quiet timer for the primary server	**timer quiet** *minutes*	Optional 5 minutes by default
Set the real-time accounting interval	**timer realtime-accounting** *minutes*	Optional 12 minutes by default

Note

- For real-time accounting, a NAS must transmit the accounting information of online users to the HWTACACS accounting server periodically. Note that if the device does not receive any response to the information, it does not disconnect the online users forcibly
- The real-time accounting interval must be a multiple of 3.
- The setting of the real-time accounting interval somewhat depends on the performance of the NAS and the HWTACACS server: a shorter interval requires higher performance.

Displaying and Maintaining HWTACACS

To do…	Use the command…	Remarks
Display configuration information or statistics of the specified or all HWTACACS schemes	**display hwtacacs** [*hwtacacs-server-name* [**statistics**]] [**slot** *slot-number*]	Available in any view

To do…	Use the command…	Remarks
Display information about buffered stop-accounting requests that get no responses	**display stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name* [**slot** *slot-number*]	Available in any view
Clear HWTACACS statistics	**reset hwtacacs statistics** { **accounting** | **all** | **authentication** | **authorization** } [**slot** *slot-number*]	Available in user view
Clear buffered stop-accounting requests that get no responses	**reset stop-accounting-buffer hwtacacs-scheme** *hwtacacs-scheme-name* [**slot** *slot-number*]	Available in user view

AAA Configuration Examples

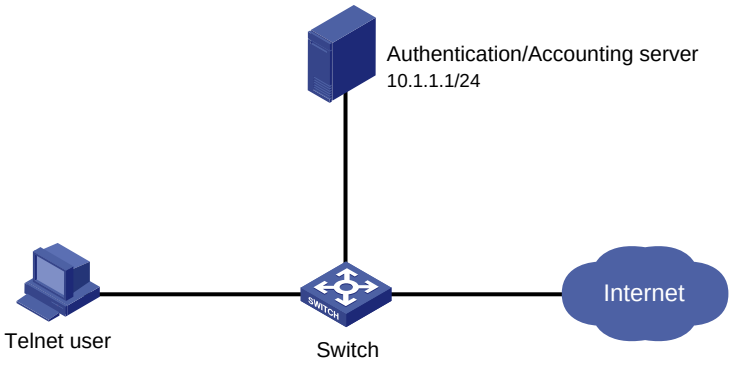
AAA for Telnet Users by a HWTACACS Server

Network requirements

As shown in Figure 1-7, configure the switch to use the HWTACACS server to provide authentication, authorization, and accounting services to login users.

- The HWTACACS server is used for authentication, authentication, and accounting. Its IP address is 10.1.1.1.
- On the switch, set the shared keys for authentication, authorization, and accounting packets to **expert**. Configure the switch to remove the domain name from a user name before sending the user name to the HWTACACS server.
- On the HWTACACS server, set the shared keys for packets exchanged with the switch to **expert**.

Figure 1-7 Configure AAA for Telnet users by a HWTACACS server



Configuration procedure

Configure the IP addresses of the interfaces (omitted).

Enable the Telnet server on the switch.

```
<Switch> system-view
[Switch] telnet server enable
```

Configure the switch to use AAA for Telnet users.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

```
[Switch-ui-vty0-4] quit
```

Configure the HWTACACS scheme.

```
[Switch] hwtacacs scheme hwtac
[Switch-hwtacacs-hwtac] primary authentication 10.1.1.1 49
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.1 49
[Switch-hwtacacs-hwtac] primary accounting 10.1.1.1 49
[Switch-hwtacacs-hwtac] key authentication expert
[Switch-hwtacacs-hwtac] key authorization expert
[Switch-hwtacacs-hwtac] key accounting expert
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Configure the AAA methods for the domain.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login hwtacacs-scheme hwtac
[Switch-isp-bbb] authorization login hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting login hwtacacs-scheme hwtac
[Switch-isp-bbb] quit
```

You can achieve the same result by setting default AAA methods for all types of users.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication default hwtacacs-scheme hwtac
[Switch-isp-bbb] authorization default hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting default hwtacacs-scheme hwtac
```

When telneting into the switch, a user enters username userid@bbb for authentication using domain **bbb**.

AAA for Telnet Users by Separate Servers

Network requirements

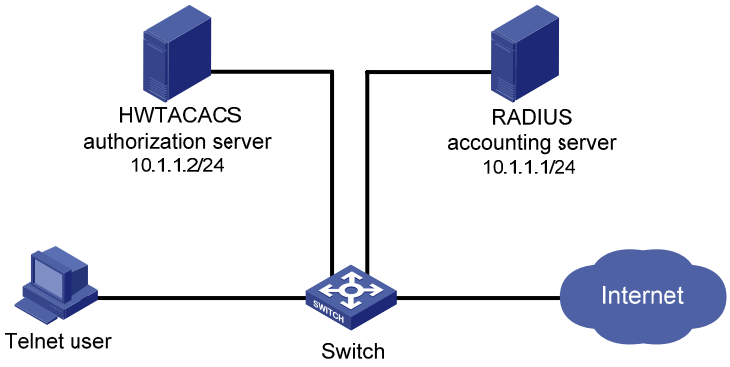
As shown in Figure 1-8, configure the switch to provide local authentication, HWTACACS authorization, and RADIUS accounting services to Telnet users. The user name and the password for Telnet users are both **hello**.

- The HWTACACS server is used for authorization. Its IP address is 10.1.1.2. On the switch, set the shared keys for packets exchanged with the HWTACACS server to **expert**. Configure the switch to remove the domain name from a user name before sending the user name to the HWTACACS server.
- The RADIUS server is used for accounting. Its IP address is 10.1.1.1. On the switch, set the shared keys for packets exchanged with the RADIUS server to **expert**.

Note

Configuration of separate AAA for other types of users is similar to that given in this example. The only difference lies in the access type.

Figure 1-8 Configure AAA by separate servers for Telnet users



Configuration procedure

Configure the IP addresses of various interfaces (omitted).

Enable the Telnet server on the switch.

```
<Switch> system-view
[Switch] telnet server enable
```

Configure the switch to use AAA for Telnet users.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
[Switch-ui-vty0-4] quit
```

Configure the HWTACACS scheme.

```
[Switch] hwtacacs scheme hwtac
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.2 49
[Switch-hwtacacs-hwtac] key authorization expert
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Configure the RADIUS scheme.

```
[Switch] radius scheme rd
[Switch-radius-rd] primary accounting 10.1.1.1 1813
[Switch-radius-rd] key accounting expert
[Switch-radius-rd] server-type extended
[Switch-radius-rd] user-name-format without-domain
[Switch-radius-rd] quit
```

Create a local user named **hello**.

```
[Switch] local-user hello
[Switch-luser-hello] service-type telnet
[Switch-luser-hello] password simple hello
[Switch-luser-hello] quit
```

Configure the AAA methods for the ISP domain.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login local
[Switch-isp-bbb] authorization login hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting login radius-scheme rd
```

```
[Switch-isp-bbb] quit
```

Configure the default AAA methods for all types of users.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication default local
[Switch-isp-bbb] authorization default hwtacacs-scheme hwtac
[Switch-isp-bbb] accounting default radius-scheme cams
```

When telneting into the switch, a user enters username telnet@bbb for authentication using domain **bbb**.

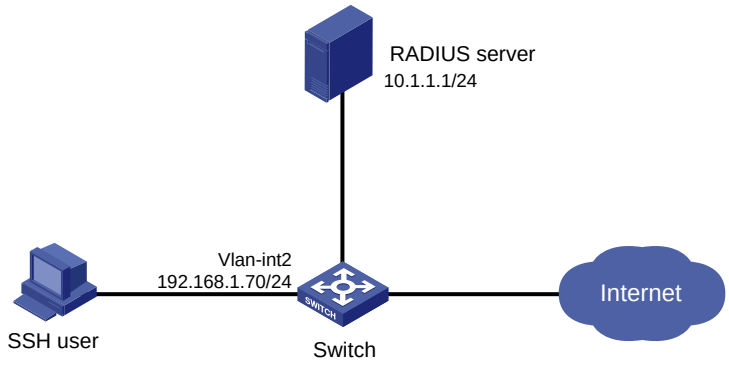
AAA for SSH Users by a RADIUS Server

Network requirements

As shown in Figure 1-9, configure the switch to use the RADIUS server to provide authentication, authorization, and accounting services to SSH users.

- The RADIUS server is responsible for both authentication and accounting. Its IP address is 10.1.1.1.
- On the switch, set both the shared keys for authentication and accounting packets to **expert**; and specify that the usernames sent to the RADIUS server carry the domain name.
- The RADIUS server runs the CAMS server.

Figure 1-9 Configure AAA for SSH users by a RADIUS server



Configuration procedure

1) Configure the RADIUS server.

Note

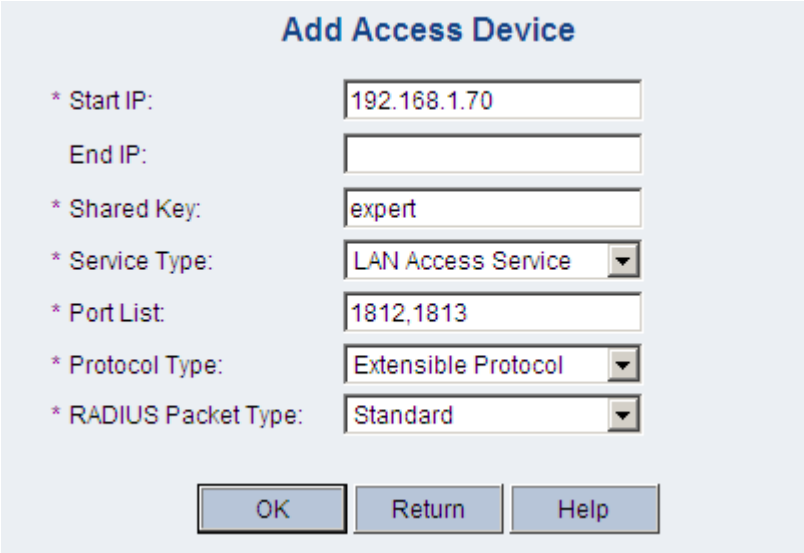
This example assumes that the RADIUS server runs the CAMS server Version 2.10.

Add an access device.

Log into the CAMS management platform and select **System Management** > **System Configuration** from the navigation tree. In the **System Configuration** window, click **Modify** of the **Access Device** item, and then click **Add** to enter the **Add Access Device** window and perform the following configurations:

- Specify the IP address of the switch as 192.168.1.70
- Set both the shared keys for authentication and accounting packets to **expert**
- Select **LAN Access Service** as the service type
- Specify the ports for authentication and accounting as 1812 and 1813 respectively
- Select **Extensible Protocol** as the protocol type
- Select **Standard** as the RADIUS packet type

Figure 1-10 Add an access device

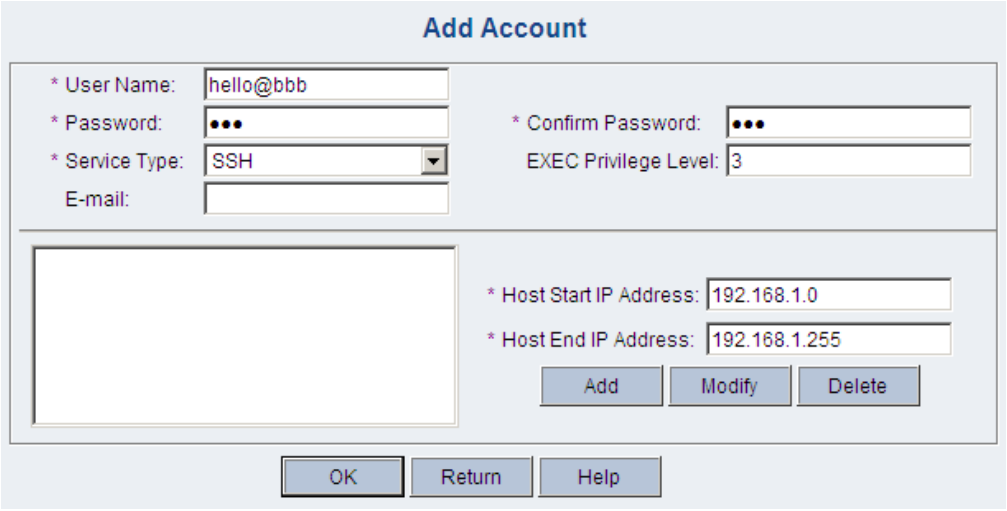


Add a user for device management

From the navigation tree, select **User Management** > **User for Device Management**, and then in the right pane, click **Add** to enter the **Add Account** window and perform the following configurations:

- Add a user named **hello@bbb**, and specify the password
- Select **SSH** as the service type
- Specify the IP address range of the hosts to be managed

Figure 1-11 Add an account for device management



2) Configure the switch

Configure the IP address of VLAN interface 2, through which the SSH user accesses the switch.

```
<Switch> system-view
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.1.70 255.255.255.0
[Switch-Vlan-interface2] quit
```

Generate RSA and DSA key pairs and enable the SSH server.

```
[Switch] public-key local create rsa
[Switch] public-key local create dsa
[Switch] ssh server enable
```

Configure the switch to use AAA for SSH users.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

Configure the user interfaces to support SSH.

```
[Switch-ui-vty0-4] protocol inbound ssh
[Switch-ui-vty0-4] quit
```

Configure the RADIUS scheme.

```
[Switch] radius scheme rad
[Switch-radius-rad] primary authentication 10.1.1.1 1812
[Switch-radius-rad] primary accounting 10.1.1.1 1813
[Switch-radius-rad] key authentication expert
[Switch-radius-rad] key accounting expert
[Switch-radius-rad] user-name-format with-domain
[Switch-radius-rad] quit
```

Configure the AAA methods for the domain.

```
[Switch] domain bbb
[Switch-isp-bbb] authentication login radius-scheme rad
[Switch-isp-bbb] authorization login radius-scheme rad
[Switch-isp-bbb] accounting login radius-scheme rad
[Switch-isp-bbb] quit
```

When using SSH to log in, a user enters a username in the form userid@bbb for authentication using domain **bbb**.

3) Verify the configuration

After the above configuration, the SSH user should be able to use the configured account to access the user interface of the switch. The commands that the user can access depend on the settings for EXEC users on the CAMS server.

Troubleshooting AAA

Troubleshooting RADIUS

Symptom 1: User authentication/authorization always fails.

Analysis:

1) A communication failure exists between the NAS and the RADIUS server.
2) The username is not in the format of userid@isp-name or no default ISP domain is specified for the NAS.
3) The user is not configured on the RADIUS server.

4) The password of the user is incorrect.
5) The RADIUS server and the NAS are configured with different shared key.

Solution:

Check that:

1) The NAS and the RADIUS server can ping each other.
2) The username is in the userid@isp-name format and a default ISP domain is specified on the NAS.
3) The user is configured on the RADIUS server.
4) The correct password is entered.
5) The same shared key is configured on both the RADIUS server and the NAS.

Symptom 2: RADIUS packets cannot reach the RADIUS server.

Analysis:

1) The communication link between the NAS and the RADIUS server is down (at the physical layer and data link layer).
2) The NAS is not configured with the IP address of the RADIUS server.
3) The UDP ports for authentication/authorization and accounting are not correct.
4) The port numbers of the RADIUS server for authentication, authorization and accounting are being used by other applications.

Solution:

Check that:

1) The communication links between the NAS and the RADIUS server work well at both physical and link layers.
2) The IP address of the RADIUS server is correctly configured on the NAS.
3) UDP ports for authentication/authorization/accounting configured on the NAS are the same as those configured on the RADIUS server.
4) The port numbers of the RADIUS server for authentication, authorization and accounting are available.

Symptom 3: A user is authenticated and authorized, but accounting for the user is not normal.

Analysis:

1) The accounting port number is not correct.
2) Configuration of the authentication/authorization server and the accounting server are not correct on the NAS. For example, one server is configured on the NAS to provide all the services of authentication/authorization and accounting, but in fact the services are provided by different servers.

Solution:

Check that:

1) The accounting port number is correctly set.
2) The authentication/authorization server and the accounting server are correctly configured on the NAS.

Troubleshooting HWTACACS

Refer to Troubleshooting RADIUS if you encounter a HWTACACS fault.

Table of Contents

1 802.1X Configuration

When configuring 802.1X, go to these sections for information you are interested in:

802.1X Overview

The 802.1X protocol was proposed by IEEE802 LAN/WAN committee for security of wireless LANs (WLAN). It has been widely used on Ethernet as a common port access control mechanism.

As a port-based access control protocol, 802.1X authenticates and controls accessing devices at the port level. A device connected to an 802.1X-enabled port of an access control device can access the resources on the LAN only after passing authentication.

Note

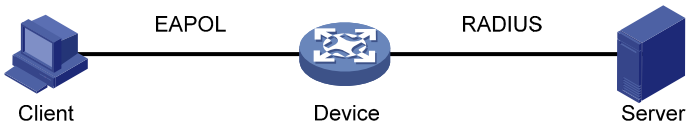
The port security feature provides rich security modes that combine or extend 802.1X and MAC address authentication. In a networking environment that requires flexible use of 802.1X and MAC address authentication, you are recommended to configure the port security feature. In a network environment that requires only 802.1X authentication, you are recommended to configure the 802.1X directly rather than configure the port security feature for simplicity sake. For how to use the port security feature, refer to *Port Security Configuration* in the *Security Volume*.

To get more information about 802.1X, go to these topics:

Architecture of 802.1X

802.1X operates in the typical client/server model and defines three entities: client, device, and server, as shown in Figure 1-1.

Figure 1-1 Architecture of 802.1X



- Client: An entity to be authenticated by the device residing on the same LAN. A client is usually a user-end device and initiates 802.1X authentication through 802.1X client software supporting the EAP over LANs (EAPOL) protocol.
- Device: The entity that authenticates connected clients residing on the same LAN. A device is usually an 802.1X-enabled network device and provides ports (physical or logical) for clients to access the LAN.
- Server: The entity providing authentication, authorization, and accounting services for the device. The server usually runs the Remote Authentication Dial-in User Service (RADIUS).

Authentication Modes of 802.1X

The 802.1X authentication system employs the Extensible Authentication Protocol (EAP) to exchange authentication information between the client, device, and authentication server.

- Between the client and the device, EAP protocol packets are encapsulated using EAPOL to be transferred on the LAN.
- Between the device and the RADIUS server, EAP protocol packets can be handled in two modes: EAP relay and EAP termination. In EAP relay mode, EAP protocol packets are encapsulated by using the EAP over RADIUS (EAPOR) and then relayed to the RADIUS server. In EAP termination mode, EAP protocol packets are terminated at the device, repackaged in the Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP) attributes of RADIUS packets, and then transferred to the RADIUS server.

Basic Concepts of 802.1X

These basic concepts are involved in 802.1X: controlled port/uncontrolled port, authorized state/unauthorized state, and control direction.

Controlled port and uncontrolled port

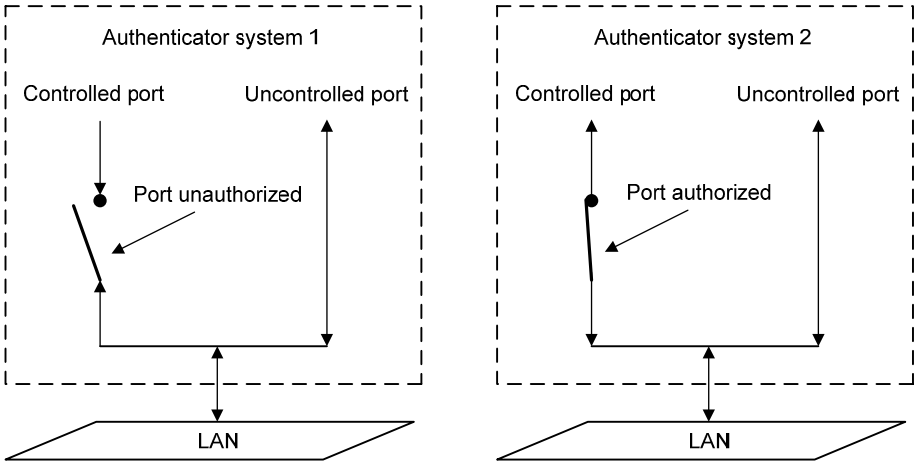
A device provides ports for clients to access the LAN. Each port can be regarded as a unity of two logical ports: a controlled port and an uncontrolled port.

- The uncontrolled port is always open in both the inbound and outbound directions to allow EAPOL protocol frames to pass, guaranteeing that the client can always send and receive authentication frames.
- The controlled port is open to allow data traffic to pass only when it is in the authorized state.
- The controlled port and uncontrolled port are two parts of the same port. Any frames arriving at the port are visible to both of them.

Authorized state and unauthorized state

The device uses the authentication server to authenticate a client trying to access the LAN and controls the status of the controlled port depending on the authentication result, putting the controlled port in the authorized state or unauthorized state, as shown in Figure 1-2.

Figure 1-2 Authorized/unauthorized status of a controlled port



You can set the access control mode of a specified port to control the authorization status. The access control modes include:

- **authorized-force**: Places the port in the authorized state, allowing users of the ports to access the network without authentication.
- **unauthorized-force**: Places the port in the unauthorized state, denying any access requests from users of the ports.
- **auto**: Places the port in the unauthorized state initially to allow only EAPOL frames to pass, and turns the ports into the authorized state to allow access to the network after the users pass authentication. This is the most common choice.

Control direction

In the unauthorized state, the controlled port can be set to deny traffic to and from the client or just the traffic from the client.

Note

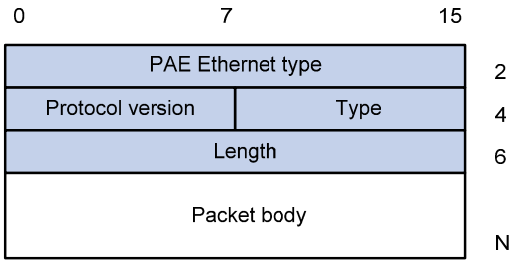
Currently, your device can only be set to deny traffic from the client.

EAP over LANs

EAPOL frame format

EAPOL, defined in 802.1X, is intended to carry EAP protocol packets between clients and devices over LANs. Figure 1-3 shows the EAPOL frame format.

Figure 1-3 EAPOL frame format



- PAE Ethernet type: Protocol type. It takes the value 0x888E.
- Protocol version: Version of the EAPOL protocol supported by the EAPOL frame sender.
- Type: Type of the EAPOL frame. Table 1-1 lists the types that the device currently supports.

Table 1-1 Types of EAPOL frames

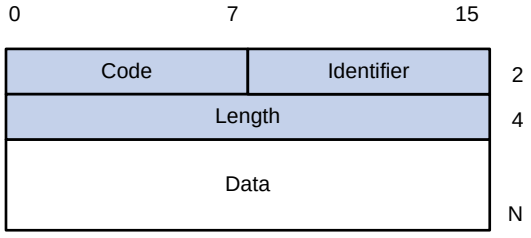
Type	Description
EAP-Packet (a value of 0x00)	Frame for carrying authentication information, present between a device and the authentication server. A frame of this type is repackaged and transferred by RADIUS to get through complex networks to reach the authentication server.
EAPOL-Start (a value of 0x01)	Frame for initiating authentication, present between a client and a device.
EAPOL-Logoff (a value of 0x02)	Frame for logoff request, present between a client and a device.

- Length: Length of the data, that is, length of the Packet body field, in bytes. If the value of this field is 0, no subsequent data field is present.
- Packet body: Content of the packet. The format of this field varies with the value of the Type field.

EAP Packet Format

An EAPOL frame of the type of EAP-Packet carries an EAP packet in its Packet body field. The format of the EAP packet is shown in Figure 1-4.

Figure 1-4 EAP packet format

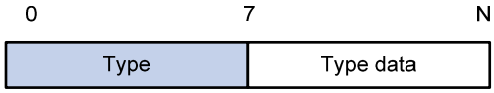


- Code: Type of the EAP packet, which can be Request, Response, Success, or Failure.

An EAP packet of the type of Success or Failure has no Data field, and has a length of 4.

An EAP packet of the type of Request or Response has a Data field in the format shown in Figure 1-5. The Type field indicates the EAP authentication type. A value of 1 represents Identity, indicating that the packet is for querying the identity of the client. A value of 4 represents MD5-Challenge, which corresponds closely to the PPP CHAP protocol.

Figure 1-5 Format of the Data field in an EAP request/response packet



- Identifier: Allows matching of responses with requests.
- Length: Length of the EAP packet, including the Code, Identifier, Length, and Data fields, in bytes.
- Data: Content of the EAP packet. This field is zero or more bytes and its format is determined by the Code field.

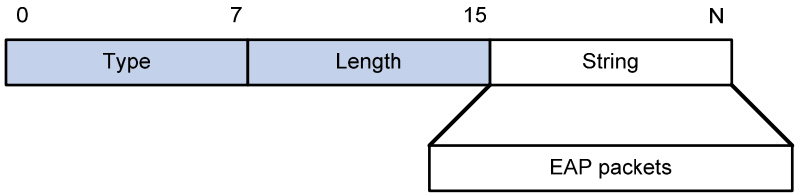
EAP over RADIUS

Two attributes of RADIUS are intended for supporting EAP authentication: EAP-Message and Message-Authenticator. For information about RADIUS packet format, refer to *AAA Configuration* in the *Security Volume*.

EAP-Message

The EAP-Message attribute is used to encapsulate EAP packets. Figure 1-6 shows its encapsulation format. The value of the Type field is 79. The String field can be up to 253 bytes. If the EAP packet is longer than 253 bytes, it can be fragmented and encapsulated into multiple EAP-Message attributes.

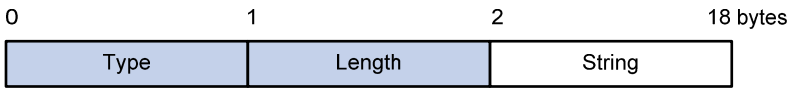
Figure 1-6 Encapsulation format of the EAP-Message attribute



Message-Authenticator

Figure 1-7 shows the encapsulation format of the Message-Authenticator attribute. The Message-Authenticator attribute is used to prevent access requests from being snooped during EAP or CHAP authentication. It must be included in any packet with the EAP-Message attribute; otherwise, the packet will be considered invalid and get discarded.

Figure 1-7 Encapsulation format of the Message-Authenticator attribute



802.1X Authentication Triggering

802.1X authentication can be initiated by either a client or the device.

Unsolicited triggering of a client

A client initiates authentication by sending an EAPOL-Start frame to the device. The destination address of the frame is 01-80-C2-00-00-03, the multicast address specified by the IEEE 802.1X protocol.

Some devices in the network may not support multicast packets with the above destination address, causing the authentication device unable to receive the authentication request of the client. To solve the problem, the device also supports EAPOL-Start frames whose destination address is a broadcast MAC address. In this case, the H3C iNode 802.1X client is required.

Unsolicited triggering of the device

The device can trigger authentication by sending EAP-Request/Identity packets to unauthenticated clients periodically (every 30 seconds by default). This method can be used to authenticate clients which cannot send EAPOL-Start frames and therefore cannot trigger authentication, for example, the 802.1X client provided by Windows XP.

Authentication Process of 802.1X

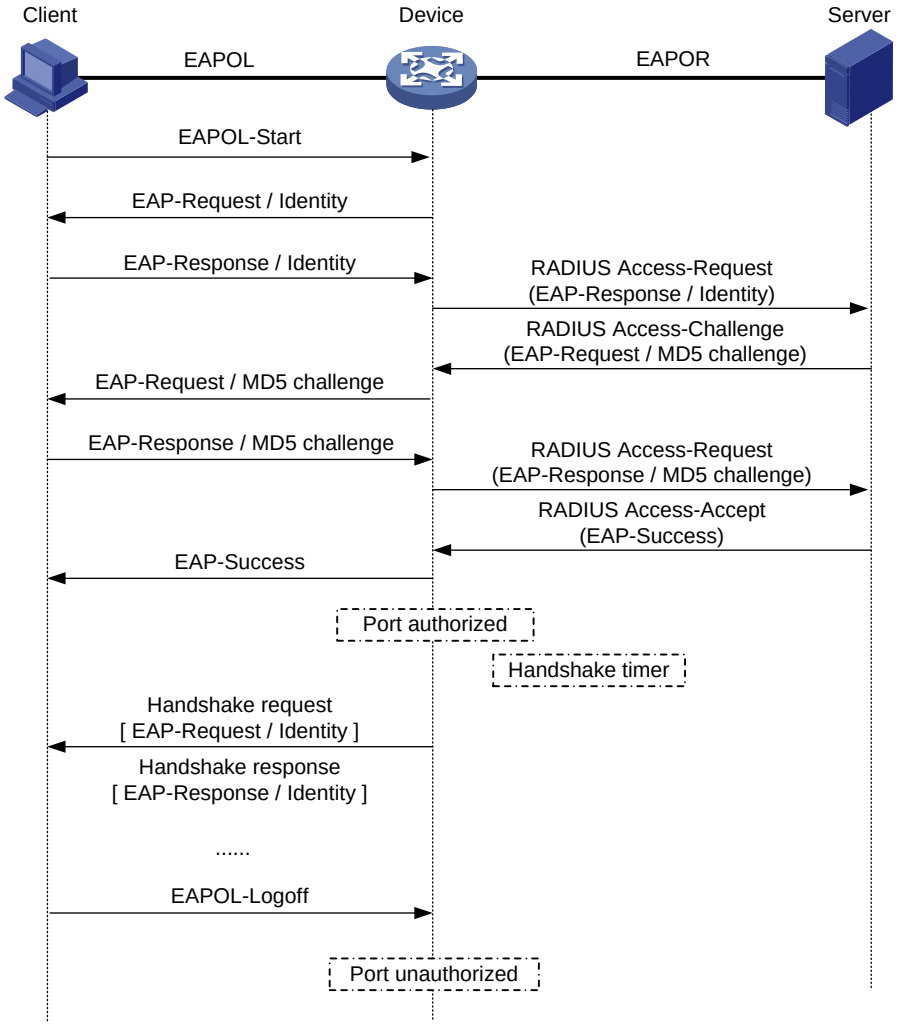
An 802.1X device communicates with a remotely located RADIUS server in two modes: EAP relay and EAP termination. The following description takes the EAP relay as an example to show the 802.1X authentication process.

EAP relay

EAP relay is an IEEE 802.1X standard mode. In this mode, EAP packets are carried in an upper layer protocol, such as RADIUS, so that they can go through complex networks and reach the authentication server. Generally, EAP relay requires that the RADIUS server support the EAP attributes of EAP-Message and Message-Authenticator, which are used to encapsulate EAP packets and protect RADIUS packets carrying the EAP-Message attribute respectively.

Figure 1-8 shows the message exchange procedure with EAP-MD5.

Figure 1-8 Message exchange in EAP relay mode



1) When a user launches the 802.1X client software and enters the registered username and password, the 802.1X client software generates an EAPOL-Start frame and sends it to the device to initiate an authentication process.
2) Upon receiving the EAPOL-Start frame, the device responds with an EAP-Request/Identity packet for the username of the client.
3) When the client receives the EAP-Request/Identity packet, it encapsulates the username in an EAP-Response/Identity packet and sends the packet to the device.
4) Upon receiving the EAP-Response/Identity packet, the device relays the packet in a RADIUS Access-Request packet to the authentication server.
5) When receiving the RADIUS Access-Request packet, the RADIUS server compares the identify information against its user information table to obtain the corresponding password information. Then, it encrypts the password information using a randomly generated challenge, and sends the challenge information through a RADIUS Access-Challenge packet to the device.
6) After receiving the RADIUS Access-Challenge packet, the device relays the contained EAP-Request/MD5 Challenge packet to the client.
7) When receiving the EAP-Request/MD5 Challenge packet, the client uses the offered challenge to encrypt the password part (this process is not reversible), creates an EAP-Response/MD5 Challenge packet, and then sends the packet to the device.
8) After receiving the EAP-Response/MD5 Challenge packet, the device relays the packet in a RADIUS Access-Request packet to the authentication server.

9) When receiving the RADIUS Access-Request packet, the RADIUS server compares the password information encapsulated in the packet with that generated by itself. If the two are identical, the authentication server considers the user valid and sends to the device a RADIUS Access-Accept packet.
10) Upon receiving the RADIUS Access-Accept packet, the device opens the port to grant the access request of the client. After the client gets online, the device periodically sends handshake requests to the client to check whether the client is still online. By default, if two consecutive handshake attempts end up with failure, the device concludes that the client has gone offline and performs the necessary operations, guaranteeing that the device always knows when a client goes offline.
11) The client can also send an EAPOL-Logoff frame to the device to go offline unsolicitedly. In this case, the device changes the status of the port from authorized to unauthorized and sends an EAP-Failure frame to the client.

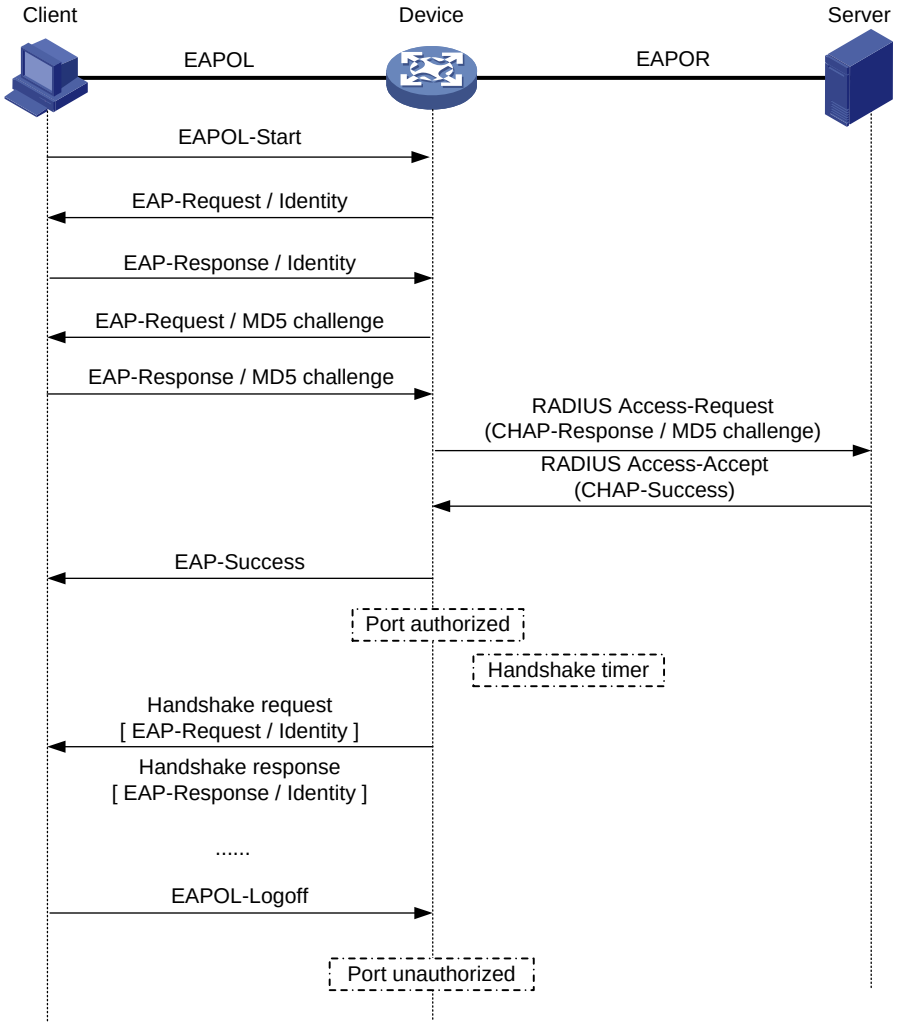
Note

In EAP relay mode, a client must use the same authentication method as that of the RADIUS server. On the device, however, you only need to execute the **dot1x authentication-method eap** command to enable EAP relay.

EAP termination

In EAP termination mode, EAP packets are terminated at the device and then repackaged into the PAP or CHAP attributes of RADIUS and transferred to the RADIUS server for authentication, authorization, and accounting. Figure 1-9 shows the message exchange procedure with CHAP authentication.

Figure 1-9 Message exchange in EAP termination mode



Different from the authentication process in EAP relay mode, it is the device that generates the random challenge for encrypting the user password information in EAP termination authentication process. Consequently, the device sends the challenge together with the username and encrypted password information from the client to the RADIUS server for authentication.

802.1X Timers

This section describes the timers used on an 802.1X device to guarantee that the client, the device, and the RADIUS server can interact with each other in a reasonable manner.

- Username request timeout timer (tx-period): The device starts this timer when it sends an EAP-Request/Identity frame to a client. If it receives no response before this timer expires, the device retransmits the request. When cooperating with a client that sends EAPOL-Start requests only when requested, the device multicasts EAP-Request/Identity frames to the client at an interval set by this timer.
- Client timeout timer (supp-timeout): Once a device sends an EAP-Request/MD5 Challenge frame to a client, it starts this timer. If this timer expires but it receives no response from the client, it retransmits the request.
- Server timeout timer (server-timeout): Once a device sends a RADIUS Access-Request packet to the authentication server, it starts this timer. If this timer expires but it receives no response from the server, it retransmits the request.

- Handshake timer (handshake-period): After a client passes authentication, the device sends to the client handshake requests at this interval to check whether the client is online. If the device receives no response after sending the allowed maximum number of handshake requests, it considers that the client is offline.
- Quiet timer (quiet-period): When a client fails the authentication, the device refuses further authentication requests from the client in this period of time.

Extensions to 802.1X

The devices extend and optimize the mechanism that the 802.1X protocol specifies by:

- Allowing multiple users to access network services through the same physical port.
- Supporting two authentication methods: **portbased** and **macbased**. With the **portbased** method, after the first user of a port passes authentication, all other users of the port can access the network without authentication, and when the first user goes offline, all other users get offline at the same time. With the **macbased** method, each user of a port must be authenticated separately, and when an authenticated user goes offline, no other users are affected.

Note

After an 802.1X client passes authentication, the authentication server sends authorization information to the device. If the authorization information contains VLAN authorization information, the device adds the port connecting the client to the assigned VLAN. This neither changes nor affects the configurations of the port. The only result is that the assigned VLAN takes precedence over the manually configured one, that is, the assigned VLAN takes effect. After the client goes offline, the configured one takes effect.

Features Working Together with 802.1X

VLAN assignment

After an 802.1X user passes the authentication, the server will send an authorization message to the device. If the server is enabled with the VLAN assignment function, the assigned VLAN information will be included in the message. The device, depending on the link type of the port used to log in, adds the port to the assigned VLAN according to the following rules:

- If the port link type is Access, the port leaves its initial VLAN, that is, the VLAN configured for it and joins the assigned VLAN.
- If the port link type is Trunk, the assigned VLAN is allowed to pass the current trunk port. The default VLAN ID of the port is that of the assigned VLAN.
- If the port link type is Hybrid, the assigned VLAN is allowed to pass the current port without carrying the tag. The default VLAN ID of the port is that of the assigned VLAN. Note that if the Hybrid port is assigned a MAC-based VLAN, the device will dynamically create a MAC-based VLAN according to the VLAN assigned by the authentication server, and remain the default VLAN ID of the port unchanged.

The assigned VLAN neither changes nor affects the configuration of a port. However, as the assigned VLAN has higher priority than the initial VLAN of the port, it is the assigned VLAN that takes effect after a user passes authentication. After the user goes offline, the port returns to the initial VLAN of the port.

For details about VLAN configuration, refer to *VLAN Configuration* in the *Access Volume*.

Note

- With a Hybrid port, the VLAN assignment will fail if you have configured the assigned VLAN to carry tags.
- With a Hybrid port, you cannot configure an assigned VLAN to carry tags after the VLAN has been assigned.

Guest VLAN

Guest VLAN allows unauthenticated users and users failing the authentication to access a specified VLAN, where the users can, for example, download or upgrade the client software, or execute some user upgrade programs. This VLAN is called the guest VLAN.

Currently, on the S4800G series Ethernet switches, a guest VLAN can be only a port-based guest VLAN (PGV), which is supported on a port that uses the access control method of **portbased**.

With PGV configured on a port, if no users are successfully authenticated on the port in a certain period of time (90 seconds by default), the port will be added to the guest VLAN and all users accessing the port will be authorized to access the resources in the guest VLAN.

The device adds a PGV-configured port into the guest VLAN according to the port's link type in the similar way as described in VLAN assignment. When a user of a port in the guest VLAN initiates an authentication, if the authentication is not successful, the port stays in the guest VLAN; if the authentication is successful, the port leaves the guest VLAN, and:

- If the authentication server assigns a VLAN, the port joins the assigned VLAN. After the user goes offline, the port returns to its initial VLAN, that is, the VLAN specified for it during port configuration, or, in other words, the VLAN it was in before it joined the guest VLAN.
- If the authentication server does not assign any VLAN, the port returns to its initial VLAN. After the client goes offline, the port just stays in its initial VLAN.

ACL assignment

ACLs provide a way of controlling access to network resources and defining access rights. When a user logs in through a port, and the RADIUS server is configured with authorization ACLs, the device will permit or deny data flows traversing through the port according to the authorization ACLs. Before specifying authorization ACLs on the server, you need to configure the ACL rules on the device. You can change the access rights of users by modifying authorization ACL settings on the RADIUS server or changing the corresponding ACL rules on the device.

Mandatory authentication domain for a specified port

The mandatory authentication domain function provides a security control mechanism for 802.1X access. With a mandatory authentication domain specified for a port, the system uses the mandatory

authentication domain for authentication, authorization, and accounting of all 802.1X users on the port. In this way, users accessing the port cannot use any account in other domains.

Meanwhile, for EAP relay mode 802.1X authentication that uses certificates, the certificate of a user determines the authentication domain of the user. However, you can specify different mandatory authentication domains for different ports even if the user certificates are from the same certificate authority (that is, the user domain names are the same). This allows you to deploy 802.1X access policies flexibly.

Configuring 802.1X

Configuration Prerequisites

802.1X provides a user identity authentication scheme. However, 802.1X cannot implement the authentication scheme solely by itself. RADIUS or local authentication must be configured to work with 802.1X.

- Configure the ISP domain to which the 802.1X user belongs and the AAA scheme to be used (that is, local authentication or RADIUS).
- For remote RADIUS authentication, the username and password information must be configured on the RADIUS server.
- For local authentication, the username and password information must be configured on the device and the service type must be set to **lan-access**.

For detailed configuration of the RADIUS client, refer to *AAA Configuration* in the *Security Volume*.

Configuring 802.1X Globally

Follow these steps to configure 802.1X globally:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable 802.1X globally		**dot1x**	Required Disabled by default
Set the authentication method		**dot1x authentication-method** { **chap** | **eap** | **pap** }	Optional CHAP by default
Set the port access control parameters	Set the port access control mode for specified or all ports	**dot1x port-control** { **authorized-force** | **auto** | **unauthorized-force** } [**interface** *interface-list*]	Optional **auto** by default
	Set the port access control method for specified or all ports	**dot1x port-method** { **macbased** | **portbased** } [**interface** *interface-list*]	Optional **macbased** by default
	Set the maximum number of users for specified or all ports	**dot1x max-user** *user-number* [**interface** *interface-list*]	Optional 256 by default

To do…	Use the command…	Remarks
Set the maximum number of attempts to send an authentication request to a client	**dot1x retry** *max-retry-value*	Optional 2 by default
Set timers	**dot1x timer** { **handshake-period** *handshake-period-value* | **quiet-period** *quiet-period-value* | **server-timeout** *server-timeout-value* | **supp-timeout** *supp-timeout-value* | **tx-period** *tx-period-value* }	Optional The defaults are as follows: 15 seconds for the handshake timer, 60 seconds for the quiet timer, 100 seconds for the server timeout timer, 30 seconds for the client timeout timer, and 30 seconds for the username request timeout timer.
Enable the quiet timer	**dot1x quiet-period**	Optional Disabled by default

Note that:

- For 802.1X to take effect on a port, you must enable it both globally in system view and for the port in system view or Ethernet interface view.
- You can also enable 802.1X and set port access control parameters (that is, the port access control mode, port access method, and the maximum number of users) for a port in Ethernet interface view. For detailed configuration, refer to Configuring 802.1X for a Port. The only difference between configuring 802.1X globally and configuring 802.1X for a port lies in the applicable scope. If both a global setting and a local setting exist for an argument of a port, the last configured one is in effect.
- 802.1X timers only need to be changed in special or extreme network environments. For example, you can give the client timeout timer a higher value in a low-performance network, give the quiet timer a higher value in a vulnerable network or a lower value for quicker authentication response, or adjust the server timeout timer to suit the performance of the authentication server.

Configuring 802.1X for a Port

Enabling 802.1X for a port

Follow these steps to enable 802.1X for a port:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable 802.1X for one or more ports	In system view	**dot1x interface** *interface-list*	Required Use either approach. Disabled by default
	In Ethernet interface view	**interface** *interface-type interface-number*	
		dot1x	

Configuring 802.1X parameters for a port

Follow these steps to configure 802.1X parameters for a port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Set the port access control mode for the port	**dot1x port-control** { **authorized-force** | **auto** | **unauthorized-force** }	Optional **auto** by default
Set the port access control method for the port	**dot1x port-method** { **macbased** | **portbased** }	Optional **macbased** by default
Set the maximum number of users for the port	**dot1x max-user** *user-number*	Optional 256 by default
Enable online user handshake	**dot1x handshake**	Optional Enabled by default
Enable multicast trigger	**dot1x multicast-trigger**	Optional Enabled by default
Specify the mandatory authentication domain for the port	**dot1x mandatory-domain** *domain-name*	Optional No mandatory authentication domain is specified by default.

Note that:

- Enabling 802.1X on a port is mutually exclusive with adding the port to an aggregation group and adding the port to a service loopback group.
- In EAP relay authentication mode, the device encapsulates the 802.1X user information in the EAP attributes of RADIUS packets and sends the packets to the RADIUS server for authentication. In this case, you can configure the **user-name-format** command but it does not take effect. For information about the **user-name-format** command, refer to *AAA Commands* in the *Security Volume.*
- If the username of a client contains the version number or one or more blank spaces, you can neither retrieve information nor disconnect the client by using the username. However, you can use items such as IP address and connection index number to do so.
- Once enabled with the 802.1X multicast trigger function, a port sends multicast trigger messages to the client periodically to initiate authentication.
- For a user-side device sending untagged traffic, the voice VLAN function and 802.1X are mutually exclusive and cannot be configured together on the same port. For details about voice VLAN, refer to *VLAN Configuration* in the *Access Volume*.

Configuring an 802.1X Port-based Guest VLAN

Configuration prerequisites

- Enable 802.1X.
- Create the VLAN to be specified as the guest VLAN.
- Set the port access control method to **portbased**.
- Ensure that the 802.1X multicast trigger function is enabled.

Configuration procedure

Follow these steps to configure a port-based guest VLAN:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Configure the guest VLAN for specified or all ports	In system view	**dot1x guest-vlan** *guest-vlan-id* [**interface** *interface-list*]	Required Use either approach. By default, a port is configured with no guest VLAN.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		dot1x guest-vlan *vlan-id*	

Note

- Different ports can be configured with different guest VLANs, but a port can be configured with only one guest VLAN.
- You cannot configure both the guest VLAN function and the free IP function in EAD fast deployment.

Caution

If the data flows from a user-side device carry VLAN tags, and 802.1X and guest VLAN are enabled on the access port, you are recommended to configure different VLAN IDs for the voice VLAN, the default port VLAN, and the guest VLAN of 802.1X.

Displaying and Maintaining 802.1X

To do…	Use the command…	Remarks
Display 802.1X session information, statistics, or configuration information of specified or all ports	**display dot1x** [**sessions** | **statistics**] [**interface** *interface-list*]	Available in any view
Clear 802.1X statistics	**reset dot1x statistics** [**interface** *interface-list*]	Available in user view

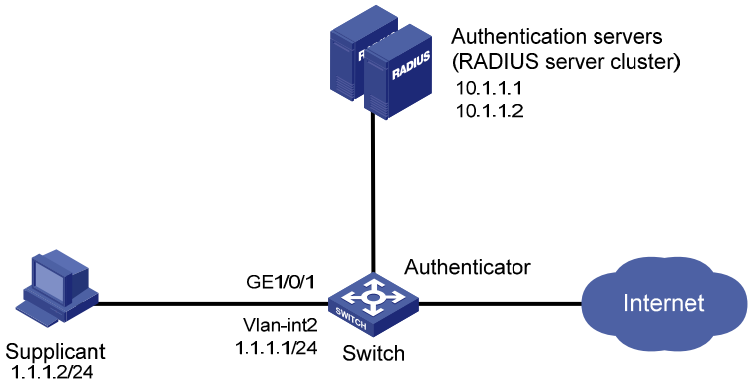
802.1X Configuration Example

Network requirements

- The access control method of **macbased** is required on the port GigabitEthernet 1/0/1 to control clients.

- All clients belong to default domain aabbcc.net, which can accommodate up to 30 users. RADIUS authentication is performed at first, and then local authentication when no response from the RADIUS server is received. If the RADIUS accounting fails, the device gets users offline.
- A server group with two RADIUS servers is connected to the device. The IP addresses of the servers are 10.1.1.1 and 10.1.1.2 respectively. Use the former as the primary authentication/secondary accounting server, and the latter as the secondary authentication/primary accounting server.
- Set the shared key for the device to exchange packets with the authentication server as name, and that for the device to exchange packets with the accounting server as money.
- Specify the device to try up to five times at an interval of 5 seconds in transmitting a packet to the RADIUS server until it receives a response from the server, and to send real time accounting packets to the accounting server every 15 minutes.
- Specify the device to remove the domain name from the username before passing the username to the RADIUS server.
- Set the username of the 802.1X user as **localuser** and the password as **localpass** and specify to use clear text mode. Enable the idle cut function to get the user offline whenever the user remains idle for over 20 minutes.

Figure 1-10 Network diagram for 802.1X configuration



Configuration procedure

Note

The following configuration procedure covers most AAA/RADIUS configuration commands for the device, while configuration on the 802.1X client and RADIUS server are omitted. For information about AAA/RADIUS configuration commands, refer to *AAA Configuration* in the *Security Volume*.

Configure the IP addresses for each interface. (Omitted)

Add local access user localuser, enable the idle cut function, and set the idle cut interval.

```
<Device> system-view
[Device] local-user localuser
[Device-luser-localuser] service-type lan-access
[Device-luser-localuser] password simple localpass
[Device-luser-localuser] attribute idle-cut 20
```

```
[Device-luser-localuser] quit
```

Create RADIUS scheme **radius1** and enter its view.

```
[Device] radius scheme radius1
```

Configure the IP addresses of the primary authentication and accounting RADIUS servers.

```
[Device-radius-radius1] primary authentication 10.1.1.1
[Device-radius-radius1] primary accounting 10.1.1.2
```

Configure the IP addresses of the secondary authentication and accounting RADIUS servers.

```
[Device-radius-radius1] secondary authentication 10.1.1.2
[Device-radius-radius1] secondary accounting 10.1.1.1
```

Specify the shared key for the device to exchange packets with the authentication server.

```
[Device-radius-radius1] key authentication name
```

Specify the shared key for the device to exchange packets with the accounting server.

```
[Device-radius-radius1] key accounting money
```

Set the interval for the device to retransmit packets to the RADIUS server and the maximum number of transmission attempts.

```
[Device-radius-radius1] timer response-timeout 5
[Device-radius-radius1] retry 5
```

Set the interval for the device to send real time accounting packets to the RADIUS server.

```
[Device-radius-radius1] timer realtime-accounting 15
```

Specify the device to remove the domain name of any username before passing the username to the RADIUS server.

```
[Device-radius-radius1] user-name-format without-domain
[Device-radius-radius1] quit
```

Create domain **aabbcc.net** and enter its view.

```
[Device] domain aabbcc.net
```

Set **radius1** as the RADIUS scheme for users of the domain and specify to use local authentication as the secondary scheme.

```
[Device-isp-aabbcc.net] authentication default radius-scheme radius1 local
[Device-isp-aabbcc.net] authorization default radius-scheme radius1 local
[Device-isp-aabbcc.net] accounting default radius-scheme radius1 local
```

Set the maximum number of users for the domain as 30.

```
[Device-isp-aabbcc.net] access-limit enable 30
```

Enable the idle cut function and set the idle cut interval.

```
[Device-isp-aabbcc.net] idle-cut enable 20
[Device-isp-aabbcc.net] quit
```

Configure **aabbcc.net** as the default domain.

```
[Device] domain default enable aabbcc.net
```

Enable 802.1X globally.

```
[Device] dot1x
```

Enable 802.1X for port GigabitEthernet 1/0/1.

```
[Device] interface GigabitEthernet 1/0/1
[Device-GigabitEthernet1/0/1] dot1x
[Device-GigabitEthernet1/0/1] quit
```

Set the port access control method. (Optional. The default settings meet the requirement.)

```
[Device] dot1x port-method macbased interface GigabitEthernet 1/0/1
```

Guest VLAN and VLAN Assignment Configuration Example

Network requirements

As shown in Figure 1-11:

- A host is connected to port GigabitEthernet 1/0/2 of the device and must pass 802.1X authentication to access the Internet. GigabitEthernet 1/0/2 is in VLAN 1.
- The authentication server runs RADIUS and is in VLAN 2.
- The update server, which is in VLAN 10, is for client software download and upgrade.
- Port GigabitEthernet 1/0/3 of the device, which is in VLAN 5, is for accessing the Internet.

As shown in Figure 1-12:

- On port GigabitEthernet 1/0/2, enable 802.1X and set VLAN 10 as the guest VLAN of the port. If the device sends an EAP-Request/Identity packet from the port for the maximum number of times but still receives no response, the device adds the port to its guest VLAN. In this case, the host and the update server are both in VLAN 10, so that the host can access the update server and download the 802.1X client.

As shown in Figure 1-13:

- After the host passes the authentication and logs in, the host is added to VLAN 5. In this case, the host and GigabitEthernet 1/0/3 are both in VLAN 5, so that the host can access the Internet.

Figure 1-11 Network diagram for guest VLAN configuration

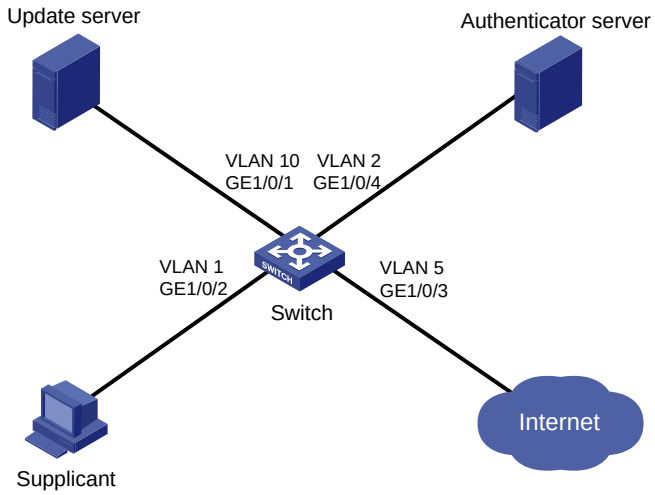


Figure 1-12 Network diagram with the port in the guest VLAN

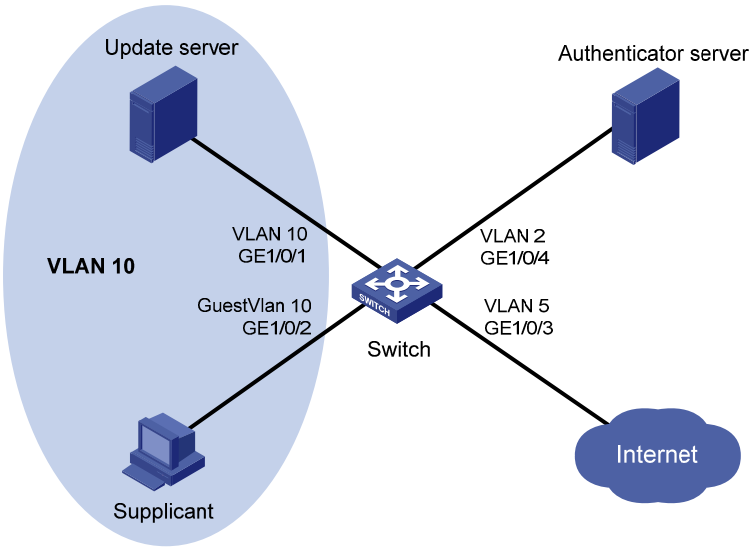
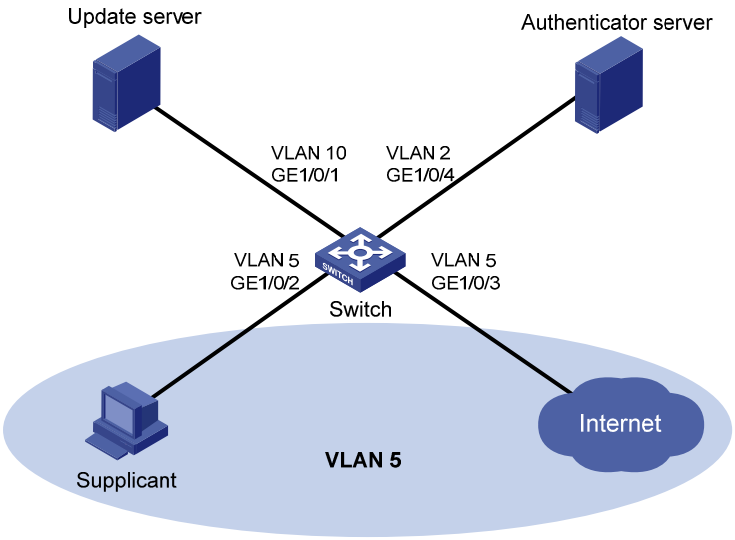


Figure 1-13 Network diagram when the client passes authentication



Configuration procedure

Note

- The following configuration procedure uses many AAA/RADIUS commands. For detailed configuration of these commands, refer to *AAA Configuration* in the *Security Volume*.
- Configurations on the 802.1X client and RADIUS server are omitted.

Configure RADIUS scheme 2000.

```
<Device> system-view
[Device] radius scheme 2000
```

```
[Device-radius-2000] primary authentication 10.11.1.1 1812
[Device-radius-2000] primary accounting 10.11.1.1 1813
[Device-radius-2000] key authentication abc
[Device-radius-2000] key accounting abc
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

Configure authentication domain **system** and specify to use RADIUS scheme 2000 for users of the domain.

```
[Device] domain system
[Device-isp-system] authentication default radius-scheme 2000
[Device-isp-system] authorization default radius-scheme 2000
[Device-isp-system] accounting default radius-scheme 2000
[Device-isp-system] quit
```

Enable 802.1X globally.

```
[Device] dot1x
```

Enable 802.1X for port GigabitEthernet 1/0/2.

```
[Device] interface GigabitEthernet 1/0/2
[Device-GigabitEthernet1/0/2] dot1x
```

Set the port access control method to **portbased**.

```
[Device-GigabitEthernet1/0/2] dot1x port-method portbased
```

Set the port access control mode to **auto**.

```
[Device-GigabitEthernet1/0/2] dot1x port-control auto
[Device-GigabitEthernet1/0/2] quit
```

Create VLAN 10.

```
[Device] vlan 10
[Device-vlan10] quit
```

Specify port GigabitEthernet 1/0/2 to use VLAN 10 as its guest VLAN.

```
[Device] dot1x guest-vlan 10 interface GigabitEthernet 1/0/2
```

You can use the **display current-configuration** or **display interface GigabitEthernet 1/0/2** command to view your configuration. You can also use the **display vlan 10** command in the following cases to verify whether the configured guest VLAN functions:

- When no users log in.
- When a user fails the authentication.
- When a user goes offline.

After a user passes the authentication successfully, you can use the **display interface GigabitEthernet 1/0/2** command to verity that port GigabitEthernet 1/0/2 has been added to the assigned VLAN 5.

ACL Assignment Configuration Example

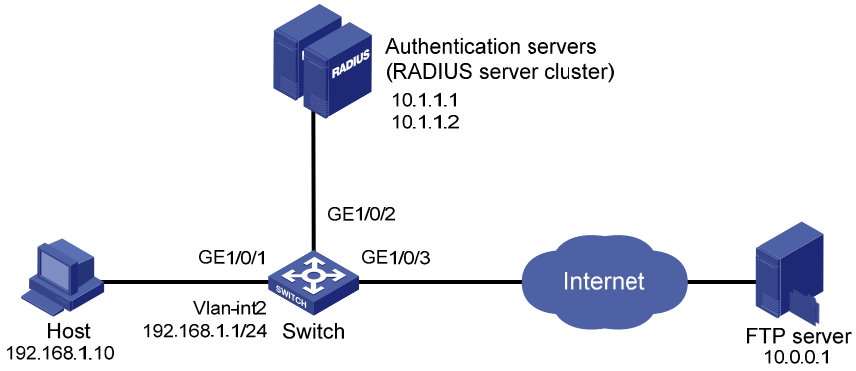
Network requirements

As shown in Figure 1-14, a host is connected to port GigabitEthernet 1/0/1 of the device and must pass 802.1X authentication to access the Internet.

- Configure the RADIUS server to assign ACL 3000.
- Enable 802.1X authentication on port GigabitEthernet 1/0/1 of the device, and configure ACL 3000.

After the host passes 802.1X authentication, the RADIUS server assigns ACL 3000 to port GigabitEthernet 1/0/1. As a result, the host can access the Internet but cannot access the FTP server, whose IP address is 10.0.0.1.

Figure 1-14 Network diagram for ACL assignment



Configuration procedure

Configure the IP addresses of the interfaces. (Omitted)

Configure the RADIUS scheme.

```
<Device> system-view
[Device] radius scheme 2000
[Device-radius-2000] primary authentication 10.1.1.1 1812
[Device-radius-2000] primary accounting 10.1.1.2 1813
[Device-radius-2000] key authentication abc
[Device-radius-2000] key accounting abc
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

Create an ISP domain and specify the AAA schemes.

```
[Device] domain 2000
[Device-isp-2000] authentication default radius-scheme 2000
[Device-isp-2000] authorization default radius-scheme 2000
[Device-isp-2000] accounting default radius-scheme 2000
[Device-isp-2000] quit
```

Configure ACL 3000 to deny packets destined for 10.0.0.1.

```
[Device] acl number 3000
[Device-acl-adv-3000] rule 0 deny ip destination 10.0.0.1 0
```

Enable 802.1X globally.

```
[Device] dot1x
```

Enable 802.1X for port GigabitEthernet 1/0/1.

```
[Device] interface GigabitEthernet 1/0/1
[Device-GigabitEthernet1/0/1] dot1x
```

After logging in successfully, a user can use the **ping** command to verify whether the ACL 3000 assigned by the RADIUS server functions.

```
[Device] ping 10.0.0.1
  PING 10.0.0.1: 56  data bytes, press CTRL_C to break
    Request time out
    Request time out
    Request time out
    Request time out
    Request time out

  --- 10.0.0.1 ping statistics ---
    5 packet(s) transmitted
    0 packet(s) received
    100.00% packet loss
```

2 EAD Fast Deployment Configuration

When configuring EAD fast deployment, go to these sections for information you are interested in:

EAD Fast Deployment Overview

Overview

Endpoint Admission Defense (EAD) is an integrated endpoint access control solution. By allowing the security clients, access devices, security policy servers, and third-party servers in the network to collaborate with each other, it can improve the overall defense capability of a network and implement centralized management of users.

Normally, to use EAD on your network, you need to manually deploy the EAD client on each device, which tends to be time consuming and inefficient. To address the issue, quick EAD deployment was developed. In conjunction with 802.1X, it can have an access switch to force all attached devices to download and install the EAD client before permitting them to access the network.

EAD Fast Deployment Implementation

To support the fast deployment of EAD schemes, 802.1X provides the following two mechanisms:

1) Limit on accessible network resources

Before successful 802.1X authentication, a user can access only a specific IP segment, which may have one or more servers. Users can download EAD client software or obtain dynamic IP address from the servers.

2) URL redirection

Before successful 802.1X authentication, a user using a Web browser to access the network is automatically redirected to a specified URL, for example, the EAD client software download page. The server that provides the URL redirection must be in the specific network segment that users can access before passing 802.1X authentication.

Configuring EAD Fast Deployment

Note

Currently, MAC authentication and port security cannot work together with EAD fast deployment. Once MAC authentication or port security is enabled globally, the EAD fast deployment is disabled automatically.

Configuration Prerequisites

- Enable 802.1X globally.
- Enable 802.1X on the specified port, and set the access control mode to **auto**.

Configuration Procedure

Configuring a freely accessible network segment

A freely accessible network segment, also called a free IP, is a network segment that users can access before passing 802.1X authentication.

Once a free IP is configured, the fast deployment of EAD is enabled.

Follow these steps to configure a freely accessible network segment:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a freely accessible network segment	**dot1x free-ip** *ip-address* { *mask-address* | *mask-length* }	Required No freely accessible network segment is configured by default.

Note

- You cannot configure both the free IP and the 802.1X guest VLAN function.
- If no freely accessible network segment is configured, a user cannot obtain a dynamic IP address before passing 802.1X authentication. To solve this problem, you can configure a freely accessible network segment that is on the same network segment with the DHCP server.

Configuring the IE redirect URL

Follow these steps to configure the IE redirect URL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the IE redirect URL	**dot1x url** *url-string*	Required No redirect URL is configured by default.

Note

The redirect URL and the freely accessible network segment must belong to the same network segment. Otherwise, the specified URL is unaccessible.

Setting the EAD rule timeout time

With the EAD fast deployment function, a user is authorized by an EAD rule (generally an ACL rule) to access the freely accessible network segment before passing authentication. After successful authentication, the occupied ACL will be released. If a large amount of users access the freely accessible network segment but fail the authentication, ACLs will soon be used up and new users will be rejected.

An EAD rule timeout timer is designed to solve this problem. When a user accesses the network, this timer is started. If the user neither downloads client software nor performs authentication before the timer expires, the occupied ACL will be released so that other users can use it. When there are a large number of users, you can shorten the timeout time to improve the ACL usage efficiency.

Follow these steps to set the EAD rule timeout time:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set EAD rule timeout time	**dot1x timer ead-timeout** *ead-timeout-value*	Optional 30 minutes by default

Displaying and Maintaining EAD Fast Deployment

To do…	Use the command…	Remarks
Display 802.1X session information, statistics, or configuration information	**display dot1x** [**sessions** | **statistics**] [**interface** *interface-list*]	Available in any view

EAD Fast Deployment Configuration Example

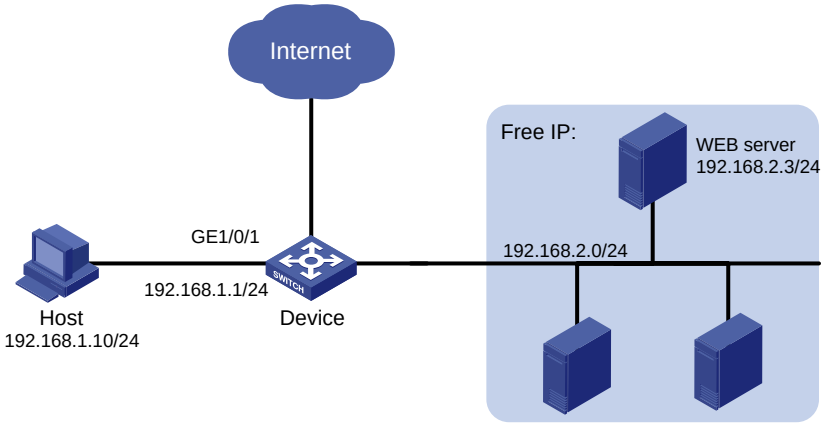
Network requirements

As shown in Figure 2-1, the host is connected to the device, and the device is connected to the freely accessible network segment and outside network.

It is required that:

- Before successful 802.1 authentication, the host using IE to access outside network will be redirected to the WEB server, and it can download and install 802.1X client software.
- After successful 802.1X authentication, the host can access outside network.

Figure 2-1 Network diagram for EAD fast deployment



Configuration procedure

1) Configure the WEB server

Before using the EAD fast deployment function, you need to configure the WEB server to provide the download service of 802.1X client software.

2) Configure the device to support EAD fast deployment

Configure the IP addresses of the interfaces (omitted).

Configure the free IP.

```
<Device> system-view
[Device] dot1x free-ip 192.168.2.0 24
```

Configure the redirect URL for client software download.

```
[Device] dot1x url http://192.168.2.3
```

Enable 802.1X globally.

```
[Device] dot1x
```

Enable 802.1X on the port.

```
[Device] interface GigabitEthernet 1/0/1
[Device -GigabitEthernet1/0/1] dot1x
```

3) Verify your configuration

Use the **ping** command to ping an IP address within the network segment specified by free IP to check that the user can access that segment before passing 802.1X authentication.

```
C:\>ping 192.168.2.3

Pinging 192.168.2.3 with 32 bytes of data:

Reply from 192.168.2.3: bytes=32 time<1ms TTL=128
Reply from 192.168.2.3: bytes=32 time<1ms TTL=128
Reply from 192.168.2.3: bytes=32 time<1ms TTL=128
Reply from 192.168.2.3: bytes=32 time<1ms TTL=128

Ping statistics for 192.168.2.3:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 0ms, Average = 0ms
```

Besides, if the user uses IE to access any external website, the user will be taken to the WEB server, which provides the client software download service.

Troubleshooting EAD Fast Deployment

Users Cannot be Redirected Correctly

Symptom

When a user enters an external website address in the IE browser, the user is not redirected to the specified URL.

Analysis

- The address is in the string format. In this case, the operating system of the host regards the string a website name and tries to have it resolved. If the resolution fails, the operating system sends an ARP request with the address in the format other than X.X.X.X. The redirection function does redirect this kind of ARP request.
- The address is within the freely accessible network segment. In this case, the device regards that the user is trying to access a host in the freely accessible network segment, and redirection will not take place, even if no host is present with the address.
- The redirect URL is not in the freely accessible network segment, no server is present with that URL, or the server with the URL does not provide WEB services.

Solution

- Enter an IP address that is not within the freely accessible network segment in dotted decimal notation (X.X.X.X).
- Ensure that the device and the server are configured correctly.

Table of Contents

1 HABP Configuration

When configuring HABP, go to these sections for the information you are interested in:

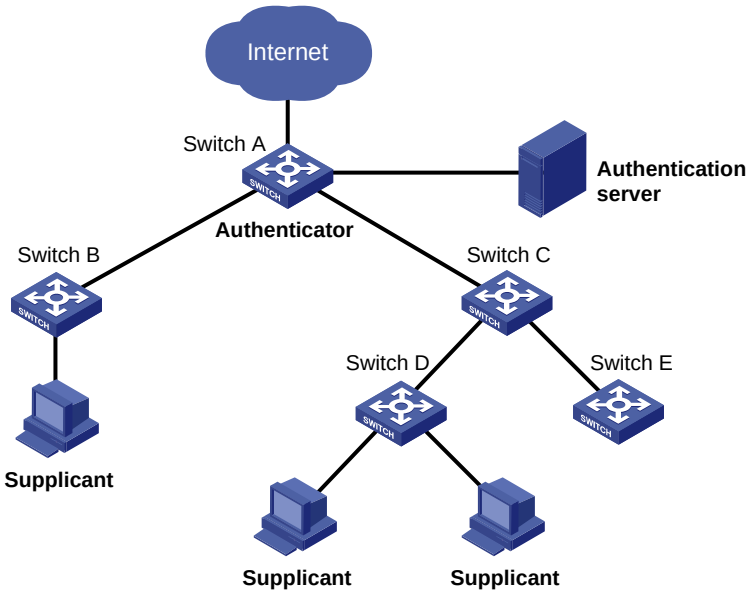
Introduction to HABP

The HW Authentication Bypass Protocol (HABP) is used to enable the downstream network devices of an 802.1X or MAC authentication enabled access device to bypass 802.1X authentication and MAC authentication.

HABP is usually adopted at the access layer of a campus or enterprise network. This feature is useful when 802.1X authentication or MAC address authentication is adopted on the management switch of a cluster, in which case you must configure HABP to allow the packets between the member devices of the cluster to bypass 802.1X authentication because network devices usually do not support 802.1 client. Otherwise, the management device will fail to perform centralized management of the cluster member devices. For more information about the cluster function, refer to *Cluster Configuration* in the *System Volume*.

As shown in Figure 1-1, 802.1X authenticator Switch A has two switches attached to it: Switch B and Switch C. On Switch A, 802.1X authentication is enabled globally and on the ports connecting the downstream network devices. The end-user devices (the supplicants) run the 802.1X client software for 802.1X authentication. For Switch B and Switch D, where 802.1X client is not supported (which is typical of network devices), the communication between them will fail because they cannot pass 802.1X authentication and their packets will be blocked on Switch A. To allow the two switches to communicate, you can use HABP.

Figure 1-1 Network diagram for HABP application



HABP is a link layer protocol that works above the MAC layer. It is built on the client-server model. Generally, the HABP server is assumed by the management device (such as Switch A in the above example), and the attached switches function as the HABP clients, such as Switch B through Switch E in the example. No device can function as both an HABP server and a client at the same time. Typically, the HABP server sends HABP requests to all its clients periodically to collect their MAC addresses, and the clients respond to the requests. After the server learns the MAC addresses of all the clients, it registers the MAC addresses as HABP entries. Then, link layer frames exchanged between the clients can bypass the 802.1X authentication on ports of the server without affecting the normal operation of the whole network. All HABP packets must travel in a VLAN, which is called the management VLAN. Communication between the HABP server and the HABP clients is implemented through the management VLAN.

Configuring HABP

Complete the following tasks to configure HABP:

- Configuring the HABP Server
- Configuring an HABP Client

Configuring the HABP Server

With the HABP server function enabled, the administrative device starts to send HABP requests to the attached switches. The HABP responses include the MAC addresses of the attached switches. This makes it possible for the administrative device to manage the attached switches.

You can configure the interval of sending HABP requests on the administrative device.

Follow these steps to configure an HABP server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable HABP	**habp enable**	Optional Enabled by default

To do…	Use the command…	Remarks
Configure HABP to work in server mode	**habp server vlan** *vlan-id*	Required HABP works in client mode by default.
Set the interval to send HABP requests	**habp timer** *interval*	Optional 20 seconds by default

Configuring an HABP Client

Configure the HABP client function on each device that is attached to the administrative device and needs to be managed. As the HABP client function is enabled by default, this configuration task is optional.

Follow these steps to configure an HABP client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable HABP	**habp enable**	Optional Enabled by default
Configure HABP to work in client mode	**undo habp server**	Optional HABP works in client mode by default.

Displaying and Maintaining HABP

To do…	Use the command…	Remarks
Display HABP configuration information	**display habp**	Available in any view
Display HABP MAC address table entries	**display habp table**	Available in any view
Display HABP packet statistics	**display habp traffic**	Available in any view

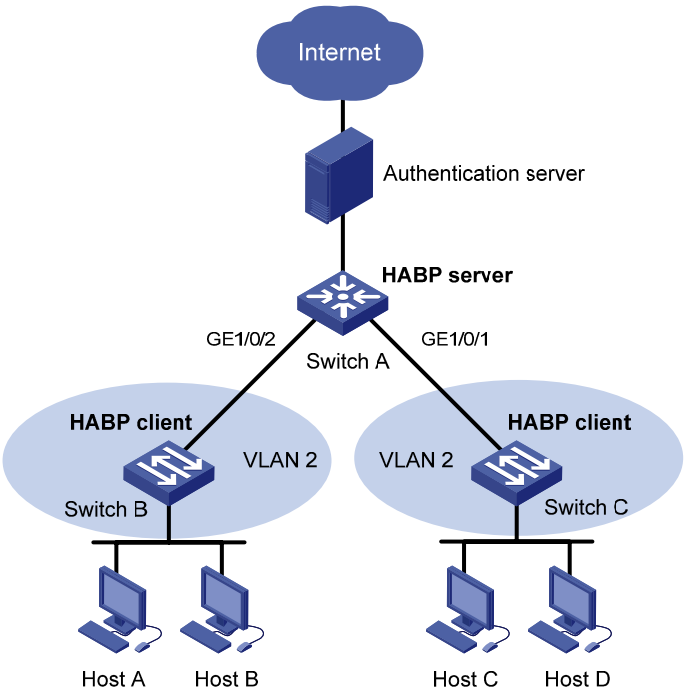
HABP Configuration Example

Network requirements

Switch A is the administrative device and connects two access devices: Switch B and Switch C. Configure HABP so that Switch A can manage Switch B and Switch C.

- Configure Switch A as the HABP server, allowing HABP packets to be transmitted in VLAN 2.
- Enable HABP client on Switch B and Switch C.
- On switch A, set the interval to send HABP request packets to 50 seconds.

Figure 1-2 Network diagram for HABP configuration



Configuration procedure

1) Configure Switch A

Enable HABP.

```
<SwitchA> system-view
[SwitchA] habp enable
```

Configure HABP to work in server mode, allowing HABP packets to be transmitted in VLAN 2.

```
[SwitchA] habp server vlan 2
```

Set the interval to send HABP request packets to 50 seconds.

```
[SwitchA] habp timer 50
```

2) Configure Switch B and Switch C

Configure Switch B and Switch C to work in HABP client mode. This configuration is usually unnecessary because HABP is enabled and works in client mode by default.

3) Verify your configuration

Display HABP configuration information.

```
<SwitchA> display habp
Global HABP information:
    HABP Mode: Server
    Sending HABP request packets every 50 seconds
    Bypass VLAN: 2
```

Display HABP MAC address table entries.

```
<SwitchA> display habp table
MAC             Holdtime  Receive Port
001f-3c00-0030  53        GigabitEthernet 1/0/2
001f-3c00-0031  53        GigabitEthernet 1/0/1
```

Table of Contents

1 MAC Authentication Configuration

When configuring MAC authentication, go to these sections for information you are interested in:

MAC Authentication Overview

MAC authentication provides a way for authenticating users based on ports and MAC addresses. Once detecting a new MAC address, the device initiates the authentication process. MAC authentication requires neither client software to be installed on the hosts, nor any username or password to be entered by users during authentication.

Currently, the device supports two MAC authentication modes: Remote Authentication Dial-In User Service (RADIUS) based MAC authentication and local MAC authentication. For detailed information about RADIUS authentication and local authentication, refer to *AAA Configuration* of the *Security Volume*.

MAC authentication supports two types of usernames:

- MAC address, where the MAC address of a user serves as both the username and password.
- Fixed username, where all users use the same preconfigured username and password for authentication, regardless of the MAC addresses.

RADIUS-Based MAC Authentication

In RADIUS-based MAC authentication, the device serves as a RADIUS client and requires a RADIUS server to cooperate with it.

- If the type of username is MAC address, the device forwards a detected MAC address as the username and password to the RADIUS server for authentication of the user.
- If the type of username is fixed username, the device sends the same username and password configured locally to the RADIUS server for authentication of each user.

If the authentication succeeds, the user will be granted permission to access the network resources.

Local MAC Authentication

In local MAC authentication, the device performs authentication of users locally and different items need to be manually configured for users on the device according to the specified type of username:

- If the type of username is MAC address, a local user must be configured for each user on the device, using the MAC address of the accessing user as both the username and password.
- If the type of username is fixed username, a single username and optionally a single password are required for the device to authenticate all users.

Related Concepts

MAC Authentication Timers

The following timers function in the process of MAC authentication:

- Offline detect timer: At this interval, the device checks to see whether there is traffic from a user. Once detecting that there is no traffic from a user within this interval, the device logs the user out and sends to the RADIUS server a stop accounting request.
- Quiet timer: Whenever a user fails MAC authentication, the device does not perform MAC authentication of the user during such a period.
- Server timeout timer: During authentication of a user, if the device receives no response from the RADIUS server in this period, it assumes that its connection to the RADIUS server has timed out and forbids the user to access the network.

Quiet MAC Address

When a user fails MAC authentication, the MAC address becomes a quiet MAC address, which means that any packets from the MAC address will be discarded silently by the device until the quiet timer expires. This prevents the device from authenticating an illegal user repeatedly in a short time.

 **Caution**

If a quiet MAC address is the same as a static MAC address configured or an MAC address that has passed another type of authentication, the quiet function does not take effect.

VLAN Assigning

For separation of users from restricted network resources, users and restricted resources are usually put into different VLANs. After a user passes identity authentication, the authorization server assigns to the user the VLAN where the restricted resources reside as an authorized VLAN, and the port through which the user accesses the device will be assigned to the authorized VLAN. As a result, the user can access those restricted network resources.

ACL Assigning

ACLs assigned by an authorization server are referred to as authorization ACLs, which are designed to control access to network resources. If the RADIUS server is configured with authorization ACLs, the device will permit or deny data flows traversing through the port through which a user accesses the device according to the authorization ACLs. You can change access rights of users by modifying authorization ACL settings on the RADIUS server.

Configuring MAC Authentication

Configuration Prerequisites

- Create and configure an ISP domain.
- For local authentication, create the local users and configure the passwords.

- For RADIUS authentication, ensure that a route is available between the device and the RADIUS server, and add the usernames and passwords on the server.

Caution

When adding usernames and passwords on the device or server, ensure that:

- The type of username and password must be consistent with that used for MAC authentication.
- All the letters in the MAC address to be used as the username and password must be in lower case.
- The service type of the local users must be configured as **lan-access**.

Configuration Procedure

Follow these steps to configure MAC authentication:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable MAC authentication globally	**mac-authentication**	Required Disabled by default
Enable MAC authentication for specified ports	**mac-authentication interface** *interface-list*	Required Use either approach. Disabled by default
	interface *interface-type interface-number* **mac-authentication** **quit**	
Specify the ISP domain for MAC authentication	**mac-authentication domain** *isp-name*	Optional The default ISP domain is used by default.
Set the offline detect timer	**mac-authentication timer offline-detect** *offline-detect-value*	Optional 300 seconds by default
Set the quiet timer	**mac-authentication timer quiet** *quiet-value*	Optional 60 seconds by default
Set the server timeout timer	**mac-authentication timer server-timeout** *server-timeout-value*	Optional 100 seconds by default
Configure the username and password for MAC authentication	**mac-authentication user-name-format** { **fixed** [**account** *name*] [**password** { **cipher** | **simple** } *password*] | **mac-address** [**with-hyphen** | **without-hyphen**] }	Optional By default, the user's source MAC address serves as the username and password, with "-" in the MAC address.

Note

- You can configure MAC authentication for ports first. However, the configuration takes effect only after you enable MAC authentication globally.
- Enabling MAC authentication on a port is mutually exclusive with adding the port to an aggregation group and adding the port to a service loopback group.
- For details about the default ISP domain, refer to *AAA Configuration* in the *Security Volume*.

Displaying and Maintaining MAC Authentication

To do…	Use the command…	Remarks
Display the global MAC authentication information or the MAC authentication information about specified ports	**display mac-authentication** [**interface** *interface-list*]	Available in any view
Clear the MAC authentication statistics	**reset mac-authentication statistics** [**interface** *interface-list*]	Available in user view

MAC Authentication Configuration Examples

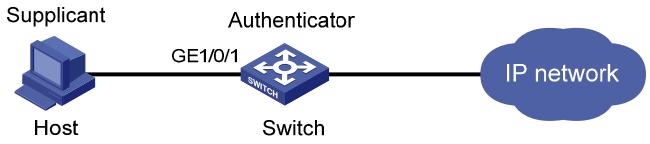
Local MAC Authentication Configuration Example

Network requirements

As illustrated in Figure 1-1, a supplicant is connected to the device through port GigabitEthernet 1/0/1.

- Local MAC authentication is required on every port to control user access to the Internet.
- All users belong to domain aabbcc.net.
- Local users use their MAC addresses as the usernames and passwords for authentication.
- Set the offline detect timer to 180 seconds and the quiet timer to 3 minutes.

Figure 1-1 Network diagram for local MAC authentication



Configuration procedure

1) Configure MAC authentication on the device

Add a local user, setting the username and password as 00-e0-fc-12-34-56, the MAC address of the user.

```
<Device> system-view
[Device] local-user 00-e0-fc-12-34-56
[Device-luser-00-e0-fc-12-34-56] password simple 00-e0-fc-12-34-56
[Device-luser-00-e0-fc-12-34-56] service-type lan-access
```

```
[Device-luser-00-e0-fc-12-34-56] quit
```

Configure ISP domain **aabbcc.net**, and specify that the users in the domain use local authentication.

```
[Device] domain aabbcc.net
[Device-isp-aabbcc.net] authentication lan-access local
[Device-isp-aabbcc.net] quit
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Enable MAC authentication for port GigabitEthernet 1/0/1.

```
[Device] mac-authentication interface GigabitEthernet 1/0/1
```

Specify the ISP domain for MAC authentication.

```
[Device] mac-authentication domain aabbcc.net
```

Set the MAC authentication timers.

```
[Device] mac-authentication timer offline-detect 180
[Device] mac-authentication timer quiet 180
```

Specify the MAC authentication username format as MAC address, that is, using the MAC address (with hyphens) of a user as the username and password for MAC authentication of the user.

```
[Device] mac-authentication user-name-format mac-address with-hyphen
```

2) Verify the configuration

Display global MAC authentication information.

```
<Device> display mac-authentication
MAC address authentication is enabled.
User name format is MAC address, like xx-xx-xx-xx-xx-xx
 Fixed username:mac
 Fixed password:not configured
          Offline detect period is 180s
          Quiet period is 180s.
          Server response timeout value is 100s
          The max allowed user number is 1024 per slot
          Current user number amounts to 1
          Current domain is aabbcc.net
Silent Mac User info:
         MAC ADDR              From Port           Port Index
GigabitEthernet1/0/1 is link-up
  MAC address authentication is enabled
  Authenticate success: 1, failed: 0
  Current online user number is 1
    MAC ADDR         Authenticate state           AuthIndex
    00e0-fc12-3456   MAC_AUTHENTICATOR_SUCCESS     29
```

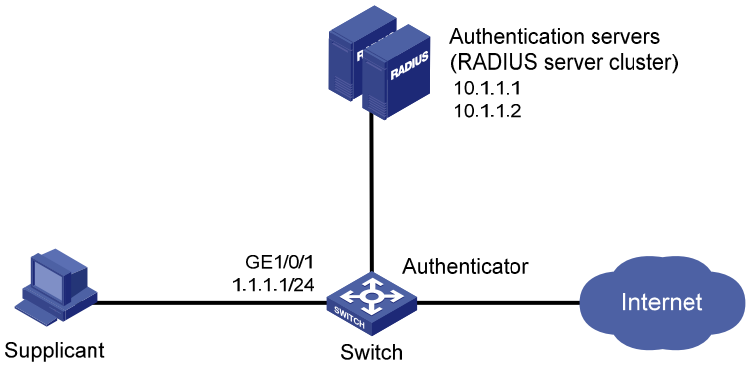
RADIUS-Based MAC Authentication Configuration Example

Network requirements

As illustrated in Figure 1-2, a host is connected to the device through port GigabitEthernet 1/0/1. The device authenticates, authorizes and keeps accounting on the host through the RADIUS server.

- MAC authentication is required on every port to control user access to the Internet.
- Set the offline detect timer to 180 seconds and the quiet timer to 3 minutes.
- All users belong to ISP domain 2000.
- The username type of fixed username is used for authentication, with the username being **aaa** and password being **123456**.

Figure 1-2 Network diagram for MAC authentication using RADIUS



Configuration procedure

Note

It is required that the RADIUS server and the device are reachable to each other and the username and password are configured on the server.

1) Configure MAC authentication on the device

Configure a RADIUS scheme.

```
<Device> system-view
[Device] radius scheme 2000
[Device-radius-2000] primary authentication 10.1.1.1 1812
[Device-radius-2000] primary accounting 10.1.1.2 1813
[Device-radius-2000] key authentication abc
[Device-radius-2000] key accounting abc
[Device-radius-2000] user-name-format without-domain
[Device-radius-2000] quit
```

Specify the AAA schemes for the ISP domain.

```
[Device] domain 2000
[Device-isp-2000] authentication default radius-scheme 2000
[Device-isp-2000] authorization default radius-scheme 2000
[Device-isp-2000] accounting default radius-scheme 2000
[Device-isp-2000] quit
```

Enable MAC authentication globally.

```
[Device] mac-authentication
```

Enable MAC authentication for port GigabitEthernet 1/0/1.

```
[Device] mac-authentication interface GigabitEthernet 1/0/1
```

Specify the ISP domain for MAC authentication.

```
[Device] mac-authentication domain 2000
```

Set the MAC authentication timers.

```
[Device] mac-authentication timer offline-detect 180
[Device] mac-authentication timer quiet 180
```

Specify to use the username **aaa** and password 123456 for MAC authentication of all users.

```
[Device] mac-authentication user-name-format fixed account aaa password simple 123456
```

2) Verify the configuration

Display global MAC authentication information.

```
<Device> display mac-authentication
MAC address authentication is enabled.
User name format is fixed account
 Fixed username:aaa
 Fixed password:123456
          Offline detect period is 180s
          Quiet period is 180s.
          Server response timeout value is 100s
          The max allowed user number is 1024 per slot
          Current user number amounts to 1
          Current domain is 2000
Silent Mac User info:
         MAC ADDR              From Port           Port Index
GigabitEthernet1/0/1 is link-up
  MAC address authentication is enabled
  Authenticate success: 1, failed: 0
  Current online user number is 1
    MAC ADDR         Authenticate state           AuthIndex
    00e0-fc12-3456   MAC_AUTHENTICATOR_SUCCESS     29
```

ACL Assignment Configuration Example

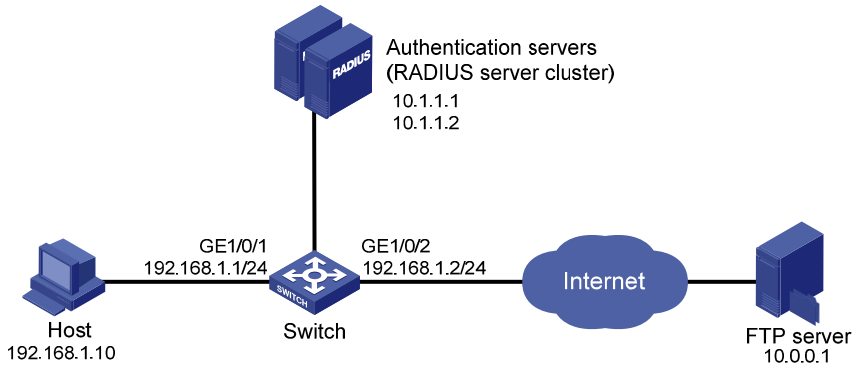
Network requirements

As shown in Figure 1-3, a host is connected to port GigabitEthernet 1/0/1 of the switch and must pass MAC authentication to access the Internet.

- Specify to use the MAC address of a user as the username and password for MAC authentication of the user.
- Configure the RADIUS server to assign ACL 3000.
- On port GigabitEthernet 1/0/1 of the switch, enable MAC authentication and configure ACL 3000.

After the host passes MAC authentication, the RADIUS server assigns ACL 3000 to port GigabitEthernet 1/0/1 of the switch. As a result, the host can access the Internet but cannot access the FTP server, whose IP address is 10.0.0.1.

Figure 1-3 Network diagram for ACL assignment



Configuration procedure

Note

- Make sure that there is a route available between the RADIUS server and the switch.
- In this example, the switch uses the default username type (user MAC address) for MAC authentication. Therefore, you need to add the username and password of each user on the RADIUS server correctly.
- You need to configure the RADIUS server to assign ACL 3000 as the authorization ACL.

Configure the RADIUS scheme.

```
<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.1.1.1 1812
[Sysname-radius-2000] primary accounting 10.1.1.2 1813
[Sysname-radius-2000] key authentication abc
[Sysname-radius-2000] key accounting abc
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit
```

Create an ISP domain and specify the AAA schemes.

```
[Sysname] domain 2000
[Sysname-isp-2000] authentication default radius-scheme 2000
[Sysname-isp-2000] authorization default radius-scheme 2000
[Sysname-isp-2000] accounting default radius-scheme 2000
[Sysname-isp-2000] quit
```

Configure ACL 3000 to deny packets destined for 10.0.0.1.

```
[Sysname] acl number 3000
[Sysname-acl-adv-3000] rule 0 deny ip destination 10.0.0.1 0
[Sysname-acl-adv-3000] quit
```

Enable MAC authentication globally.

```
[Sysname] mac-authentication
```

Specify the ISP domain for MAC authentication users.

```
[Sysname] mac-authentication domain 2000
```

Specify the MAC authentication username type as MAC address, that is, using the MAC address of a user as the username and password for MAC authentication of the user.

```
[Sysname] mac-authentication user-name-format mac-address
```

Enable MAC authentication for port GigabitEthernet 1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] mac-authentication
```

After completing the above configurations, you can use the **ping** command to verify whether the ACL 3000 assigned by the RADIUS server functions.

```
[Sysname] ping 10.0.0.1
  PING 10.0.0.1: 56  data bytes, press CTRL_C to break
    Request time out
    Request time out
    Request time out
    Request time out
    Request time out

  --- 10.0.0.1 ping statistics ---
    5 packet(s) transmitted
    0 packet(s) received
    100.00% packet loss
```

Table of Contents

1 Portal Configuration

When configuring portal, go to these sections for information you are interested in:

Portal Overview

This section covers these topics:

Introduction to Portal

Portal authentication, as its name implies, helps control access to the Internet. Portal authentication is also called web authentication and a website implementing portal authentication is called a portal website.

With portal authentication, an access device forces all users to log into the portal website at first. Every user can access the free services provided on the portal website; but to access the Internet, a user must pass portal authentication on the portal website.

A user can access a known portal website, enter username and password for authentication. This authentication mode is called active authentication. There is still another authentication mode, namely forced authentication, in which the access device forces a user trying to access the Internet through HTTP to log in to a portal website for authentication.

The portal feature provides the flexibility for Internet service providers (ISPs) to manage services. A portal website can, for example, present advertisements, and deliver community services and personalized services. In this way, broadband network providers, equipment providers, and content service providers form an industrial ecological system.

Introduction to Extended Portal Functions

By forcing users to implement patching and anti-virus policies, extended portal functions help users to defend against viruses. The main extended functions are described as follows:

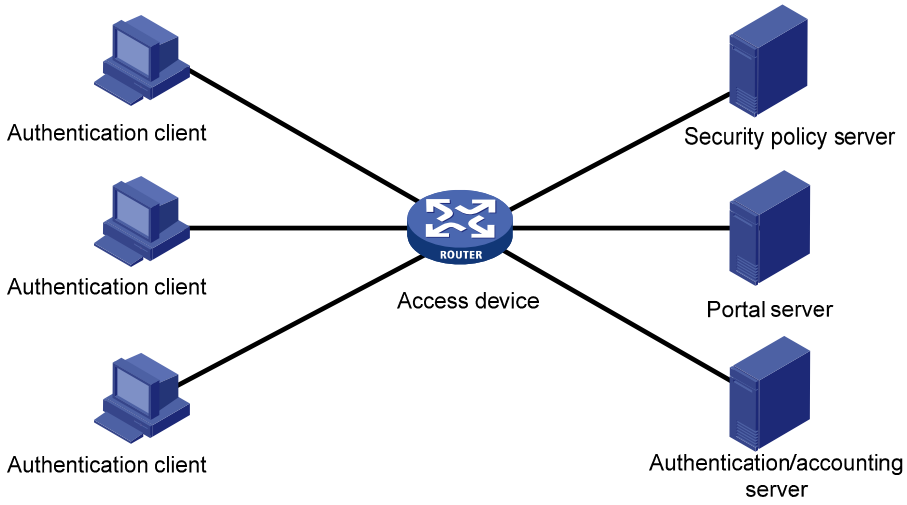
- Security authentication mechanism: The security authentication mechanism works after the identity authentication process to check that the required anti-virus software, virus definition updates and OS patches are installed, and no unauthorized software is installed on the terminal of a user.

- Resource access limit: A user passing identity authentication can access only network resources like the anti-virus server or OS patch server, which are called the restricted resources. Only users passing security authentication can access more network resources, which are called the unrestricted resources.

Portal System Components

As shown in Figure 1-1, a typical portal system consists of five basic components: authentication client, access device, portal server, authentication/accounting server, and security policy server.

Figure 1-1 Portal system components



Authentication client

Client system of a user to be authenticated. It can be a browser using the Hypertext Transfer Protocol (HTTP/HTTPS), or a host running the portal client software. The security authentication of a client depends on the communications between the portal client and the security policy server.

Access device

Device for broadband access. It can be a switch or a router that provides the following three functions:

- Before authentication, redirecting all HTTP requests from users in the subnet to be authenticated to the portal server.
- During authentication, interacting with the portal server, security policy server and the authentication/accounting server for identity authentication, security authentication and accounting.
- After authentication, allowing users to access granted Internet resources.

Portal server

Server that listens to authentication requests from portal clients and exchanges client authentication information with the access device. It provides free portal services and a web-based authentication interface.

Authentication/accounting server

Server that implements user authentication and accounting through interaction with the access device.

Security policy server

Server that interacts with portal clients and access devices for security authentication and resource authorization.

The above five components interact in the following procedure:

1) When an unauthenticated user enters a website address in the address bar of the IE to access the Internet, an HTTP request is created and sent to the access device, which redirects the HTTP request to the web authentication homepage of the portal server. For extended portal functions, authentication clients must run the portal client.
2) On the authentication homepage/authentication dialog box, the user enters and submits the authentication information, which the portal server then transfers to the access device.
3) Upon receipt of the authentication information, the access device communicates with the authentication/accounting server for authentication and accounting.
4) After successful authentication, the access device checks whether there is corresponding security policy for the user. If not, it allows the user to access the Internet. Otherwise, the client, the access device and the security policy server communicates to perform security authentication of the user, and the security policy server authorizes the user to access resources depending on the security authentication result.

Note

- Since a portal client uses an IP address as its ID, ensure that there is no Network Address Translation (NAT) device between the authentication client, access device, portal server, and authentication/accounting server when deploying portal authentication. This is to avoid authentication failure due to NAT operations.
- Currently, only a RADIUS server can serve as the authentication/accounting server in a portal system.
- Currently, security authentication requires the cooperation of the H3C iNode client.

Portal Authentication Modes

Portal authentication supports two modes: non-Layer 3 authentication and Layer 3 authentication.

Non-Layer 3 authentication

Non-Layer 3 authentication falls into two categories: direct authentication and Re-DHCP authentication.

- Direct authentication

Before authentication, a user manually configures an IP address or directly obtains a public IP address through DHCP, and can access only the portal server and predefined free websites. After passing authentication, the user can access the network resources. The process of direct authentication is simpler than that of re-DHCP authentication.

- Re-DHCP authentication

Before authentication, a user gets a private IP address through DHCP and can access only the portal server and predefined free websites. After passing authentication, the user is allocated a public IP address and can access the network resources. No public IP address is allocated to those who fails

authentication. This solves the problem about IP address planning and allocation and proves to be useful. For example, a service provider can allocate public IP addresses to broadband users only when they access networks beyond the residential community network.

Layer 3 authentication

Layer 3 portal authentication is similar to direct authentication. However, in Layer-3 portal authentication mode, Layer 3 forwarding devices can be present between the authentication client and the access device.

Differences between Layer 3 and non-Layer 3 authentication modes

- Networking mode

From this point of view, the difference between these two authentication modes lies in whether or not a Layer 3 forwarding device can be present between the authentication client and the access device. The former supports Layer 3 forwarding devices, while the latter does not.

- User identifier

In Layer 3 authentication mode, a client is uniquely identified by an IP address. This is because the mode supports Layer 3 forwarding devices between the authentication client and the access device but the access device does not learn the MAC address of the authentication client. In non-Layer 3 authentication mode, a client is uniquely identified by the combination of its IP address and MAC address because the access device can learn the MAC address of the authentication client.

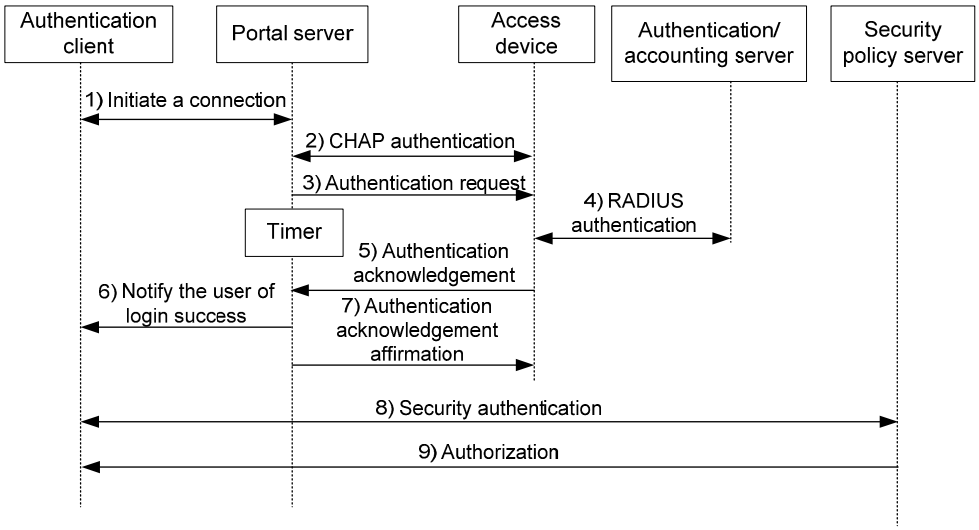
Due to the above differences, when the MAC address of an authentication client remains the same but the IP address changes, a new portal authentication will be triggered in Layer-3 authentication mode but will not be triggered in non-Layer 3 authentication mode. In non-Layer 3 authentication mode, a new portal authentication will be triggered only when both the MAC and IP address of the authentication client are changed.

Portal Authentication Process

Direct authentication and Layer 3 authentication share the same authentication process, while re-DHCP authentication has a different process because of the presence of two address allocation procedures.

Direct authentication/Layer 3 authentication process

Figure 1-2 Direct authentication/Layer 3 authentication process



The direct authentication/Layer 3 authentication process is as follows:

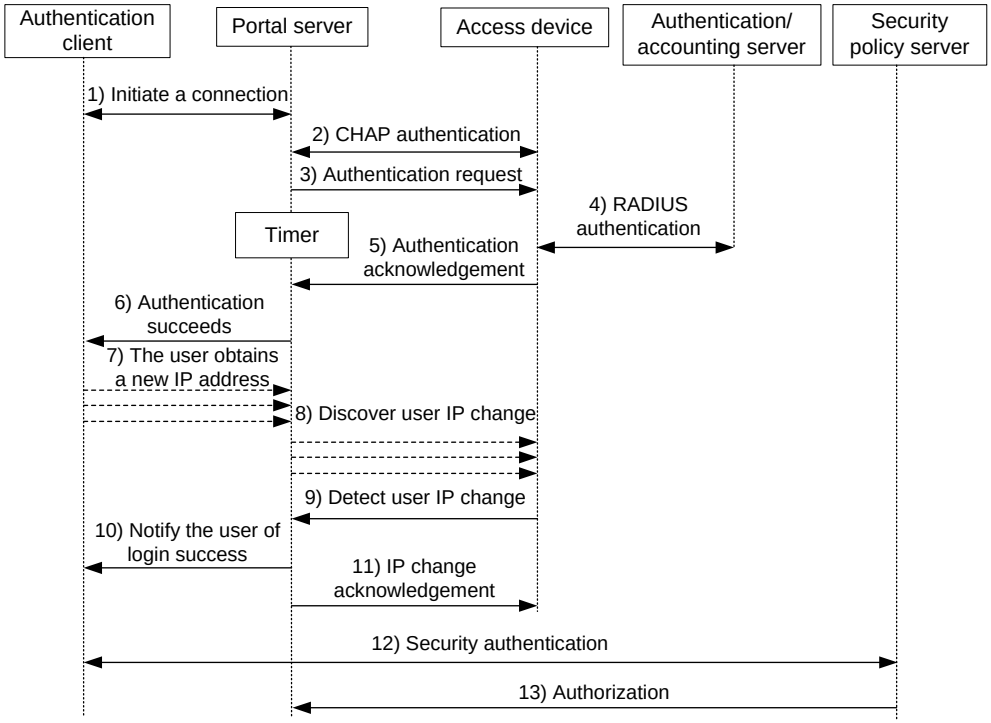
1) A portal user initiates an authentication request through HTTP. When the HTTP packet arrives at the access device, the access device allows it to pass if it is destined for the portal server or a predefined free website, or redirects it to the portal server if it is destined for other websites. The portal server provides a web page for the user to enter the username and password.
2) The portal server and the access device exchange Challenge Handshake Authentication Protocol (CHAP) messages. For Password Authentication Protocol (PAP) authentication, this step is skipped.
3) The portal server assembles the username and password into an authentication request message and sends it to the access device. Meanwhile, the portal server starts a timer to wait for an authentication acknowledgment message.
4) The access device and the RADIUS server exchange RADIUS packets to authenticate the user.
5) If the user passes authentication, the access device sends an authentication acknowledgment message to the portal server.
6) The portal server sends an authentication acknowledgment message to the authentication client to notify it of logon success.
7) The portal server sends an affirmation message to the access device.

With extended portal functions, the process includes two additional steps:

8) The security policy server exchanges security authentication information with the client to check whether the authentication client meets the security requirements.
9) The security policy server authorizes the user to access unrestricted resources based on the security configuration for the user. The authorization information is stored on the access device and used by the access device to control user access.

Re-DHCP authentication process

Figure 1-3 Re-DHCP authentication process



The re-DHCP authentication process is as follows:

Step 1 through step 6 are the same as those in the direct authentication/Layer 3 portal authentication process.

7) After receiving an authentication acknowledgment message, the authentication client obtains a new public IP address through DHCP and notifies the portal server that it has obtained a public IP address.
8) The portal server notifies the access device that the authentication client has obtained a new public IP address.
9) Detecting the change of the IP address by examining ARP packets received, the access device notifies the portal server of the change.
10) The portal server notifies the authentication client of logon success.
11) The portal server sends a user IP address change acknowledgment message to the access device.

With extended portal functions, the process includes two additional steps:

12) The security policy server exchanges security authentication information with the client to check whether the authentication client meets the security requirements.
13) The security policy server authorizes the user to access unrestricted resources based on the security configuration for the user. The authorization information is stored on the access device and used by the access device to take control of user access.

Portal Configuration Task List

Complete these tasks to configure portal authentication:

Task	Remarks
Basic Portal Configuration	Required
Configuring a Portal-Free Rule	Optional
Configuring an Authentication Subnet	Optional
Logging out Users	Optional
Specifying a Mandatory Authentication Domain	Optional

Basic Portal Configuration

Configuration Prerequisites

The portal feature provides a solution for user authentication and security authentication. However, the portal feature cannot implement this solution by itself. Currently, RADIUS authentication needs to be configured on the access device to cooperate with the portal feature to complete user authentication.

The prerequisites for portal authentication are as follows:

- The portal-enabled interfaces of the access device are configured with valid IP addresses or have obtained valid IP addresses through DHCP.
- The portal server and the RADIUS server have been installed and configured properly.
- With re-DHCP authentication, the invalid IP address check function of DHCP relay is enabled on the access device, and the DHCP server is installed and configured properly.
- With RADIUS authentication, usernames and passwords of the users are configured on the RADIUS server, and the RADIUS client configurations are performed on the access device. For information about RADIUS client configuration, refer to *AAA Configuration* in the *Security Volume*.
- To implement extended portal functions, you need install and configure the security policy server and ensure that the ACLs configured on the access device correspond to those specified for restricted resources and unrestricted resources on the security policy server respectively. For information about security policy server configuration, refer to *AAA Configuration* in the *Security Volume*.

Note

- For configuration about the security policy server, refer to *CAMS EAD Security Policy Component User Manual*.
- The ACL for restricted resources and that for unrestricted resources correspond to isolation ACL and security ACL on the security policy server respectively.
- You can modify the authorized ACL on the access device. However, the new ACL takes effect only for portal users logging on after the modification.

Configuration Procedure

Basic Portal configurations include configuring the Portal server and enabling Portal on an interface. To configure a portal server, you need to specify the IP address of the portal server on the access device.

Follow these steps to perform basic portal configuration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a portal server	**portal server** *server-name* **ip** *ip-address* [**key** *key-string* | **port** *port-id* | **url** *url-string*] *	Required By default, no portal server is configured.
Enter interface view	**interface** *interface-type interface-number*	—
Enable portal authentication on the interface	**portal server** *server-name* **method** { **direct** | **layer3** | **redhcp** }	Required Disabled by default

Caution

- Enabling portal authentication on a Layer 3 port is mutually exclusive with adding the port to an aggregation group.
- The destination port number that the device uses for sending packets to the portal server unsolicitedly must be the same as that the remote portal server actually uses.
- The portal server and its parameters can be deleted or modified only when the portal server is not referenced by any interface.
- The portal server to be referenced must exist.
- Only Layer 3 authentication mode can be used in applications with Layer 3 forwarding devices present between the authentication clients and the access device. However, Layer-3 authentication does not require any Layer-3 forwarding devices between the access device and the authentication clients.
- In re-DHCP authentication mode, a user is allowed to send packets using a public IP address before portal authentication, but the corresponding response packets are restricted.

Configuring a Portal-Free Rule

A portal-free rule allows specified users to access specified external websites without portal authentication. Packets matching a portal-free rule will not trigger portal authentication and the users can directly access the specified external websites.

Follow these steps to configure a portal-free rule:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a portal-free rule	**portal free-rule** *rule-number* { **destination** { **any** | **ip** { *ip-address* **mask** { *mask-length* | *netmask* } | **any** } } | **source** { **any** | [**interface** *interface-type interface-number* | **ip** { *ip-address* **mask** { *mask-length* | *mask* } | **any** } | **mac** *mac-address* | **vlan** *vlan-id*] * } } *	Required

Note

- If you specify both a VLAN and an interface in a portal-free rule, the interface must belong to the VLAN.
- You cannot configure two or more portal-free rules with the same filtering conditions. Otherwise, the system prompts that the rule already exists.
- No matter whether portal authentication is enabled, you can only add or remove a portal-free rule, rather than modifying it.

Configuring an Authentication Subnet

By configuring authentication subnets, you specify that only packets from users on the authentication subnets trigger portal authentication. Packets that are neither from portal-free users nor from authentication subnets are discarded.

Follow these steps to configure an authentication subnet:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure an authentication subnet	**portal auth-network** *network-address* { *mask-length* | *mask* }	Optional By default, the authentication subnet is 0.0.0.0/0, which means that users with any source IP addresses are to be authenticated.

Note

- Configuration of authentication subnets applies to only Layer 3 portal authentication.
- In direct authentication mode, the authentication subnet is 0.0.0.0/0.
- In re-DHCP authentication mode, the authentication subnet of an interface is the subnet to which the private IP address of the interface belongs.

Logging out Users

Logging out a user terminates the authentication process for the user or removes the user.

Follow these steps to log out users:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Log out users	**portal delete-user** { *ip-address* | **all** | **interface** *interface-type interface-number* }	Required

Specifying a Mandatory Authentication Domain

After you specify a mandatory authentication domain for an interface, the device will use the mandatory authentication domain for authentication, authorization, and accounting (AAA) of the portal users on the interface, ignoring the domain names carried in the usernames. Thereby, you can specify different authentication domains for different interfaces as needed.

Follow these steps to specify an authentication domain for an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Specify an authentication domain for the interface	**portal domain** *domain-name*	Required By default, no authentication domain is specified for an interface.

Note

The device selects the authentication domain for a portal user on an interface in this order: the ISP domain specified for the interface, the ISP domain carried in the username, and the system default ISP domain. For descriptions on the default ISP domain, refer to *AAA Configuration* in the *Security Volume*.

Displaying and Maintaining Portal

To do…	Use the command…	Remarks
Display the ACLs on a specified interface	**display portal acl** { **all** | **dynamic** | **static** } **interface** *interface-type interface-number*	Available in any view
Display portal connection statistics on a specified interface or all interfaces	**display portal connection statistics** { **all** | **interface** *interface-type interface-number* }	Available in any view
Display information about a portal-free rule or all portal-free rules	**display portal free-rule** [*rule-number*]	Available in any view
Display the portal configuration of a specified interface	**display portal interface** *interface-type interface-number*	Available in any view
Display information about a specified portal server or all portal servers	**display portal server** [*server-name*]	Available in any view
Display portal server statistics on a specified interface or all interfaces	**display portal server statistics** { **all** | **interface** *interface-type interface-number* }	Available in any view
Display TCP spoofing statistics	**display portal tcp-cheat statistics**	Available in any view
Display information about portal users on a specified interface or all interfaces	**display portal user** { **all** | **interface** *interface-type interface-number* }	Available in any view

To do…	Use the command…	Remarks
Clear portal connection statistics on a specified interface or all interfaces	**reset portal connection statistics** {**all** | **interface** *interface-type interface-number* }	Available in user view
Clear portal server statistics on a specified interface or all interfaces	**reset portal server statistics** { **all** | **interface** *interface-type interface-number* }	Available in user view
Clear TCP spoofing statistics	**reset portal tcp-cheat statistics**	Available in user view

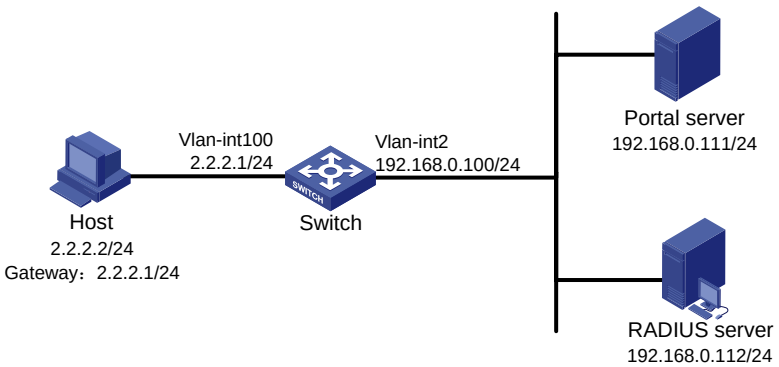
Portal Configuration Examples

Configuring Direct Portal Authentication

Network requirements

- The host is directly connected to the switch and the switch is configured for direct authentication. The host is assigned with a public network IP address manually or automatically by a DHCP server. Before portal authentication, users using the host can access only the portal server. After passing portal authentication, they can access unrestricted Internet resources.
- A RADIUS server serves as the authentication/accounting server.

Figure 1-4 Configure direct portal authentication



Configuration procedure

Note

You need to configure IP addresses for the devices as shown in Figure 1-4 and ensure that routes are available between devices.

Configure the switch:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Set the server type to **extended**.

```
[Switch-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.112
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key authentication radius
[Switch-radius-rs1] key accounting radius
```

Specify that the ISP domain name should not be included in the username sent to the RADIUS server.

```
[Switch-radius-rs1] user-name-format without-domain
[Switch-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[Switch] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[Switch] domain default enable dm1
```

3) Configure portal authentication

Configure the portal server as follows:

- Name: newpt
- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[Switch] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

Enable portal authentication on the interface connecting the host.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 2.2.2.1 255.255.255.0
[Switch–Vlan-interface100] portal server newpt method direct
[Switch] quit
```

Configure the IP address of the interface connected with the portal server.

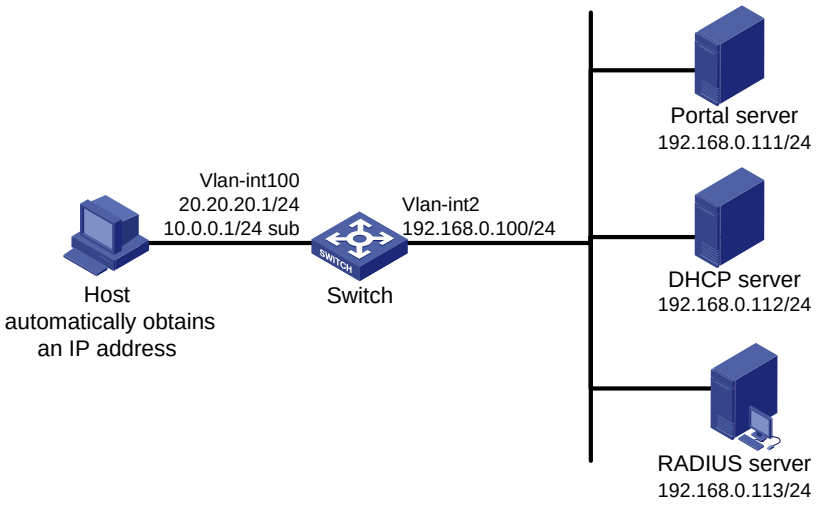
```
[Switch] interface vlan-interface 2
[Switch–Vlan-interface2] ip address 192.168.0.100 255.255.255.0
[Switch–Vlan-interface2] quit
```

Configuring Re-DHCP Portal Authentication

Network requirements

- The host is directly connected to the switch and the switch is configured for re-DHCP authentication. The host is assigned with an IP address through the DHCP server. Before portal authentication, the host uses an assigned private IP address. After passing portal authentication, it can get a public IP address and then users using the host can access unrestricted Internet resources.
- A RADIUS server serves as the authentication/accounting server.

Figure 1-5 Configure re-DHCP portal authentication



Configuration procedure

Note

- For re-DHCP authentication, you need to configure a public address pool (20.20.20.0/24, in this example) and a private address pool (10.0.0.0/24, in this example) on the DHCP server. The configuration steps are omitted. For DHCP configuration information, refer to *DHCP Configuration* in the *IP Services Volume*.
- For re-DHCP authentication, the switch must be configured as a DHCP relay agent (instead of a DHCP server) and the portal-enabled interface must be configured with a primary IP address (a public IP address) and a secondary IP address (a private IP address).
- You need to configure IP addresses for the devices as shown in Figure 1-5 and ensure that routes are available between devices.

Configure the switch:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Set the server type to **extended**.

```
[Switch-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.113
[Switch-radius-rs1] primary accounting 192.168.0.113
[Switch-radius-rs1] key authentication radius
[Switch-radius-rs1] key accounting radius
```

Specify that the ISP domain name should not be included in the username sent to the RADIUS server.

```
[Switch-radius-rs1] user-name-format without-domain
[Switch-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[Switch] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[Switch] domain default enable dm1
```

3) Configure portal authentication

Configure the portal server as follows:

- Name: newpt
- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[Switch] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

Configure the switch as a DHCP relay agent, and enable the invalid address check function.

```
[Switch] dhcp enable
[Switch] dhcp relay server-group 0 ip 192.168.0.112
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 20.20.20.1 255.255.255.0
[Switch–Vlan-interface100] ip address 10.0.0.1 255.255.255.0 sub
[Switch-Vlan-interface100] dhcp select relay
[Switch-Vlan-interface100] dhcp relay server-select 0
[Switch-Vlan-interface100] dhcp relay address-check enable
```

Enable re-DHCP portal authentication on the interface connecting the host.

```
[Switch–Vlan-interface100] portal server newpt method redhcp
[Switch–Vlan-interface100] quit
```

Configure the IP address of the interface connected with the portal server.

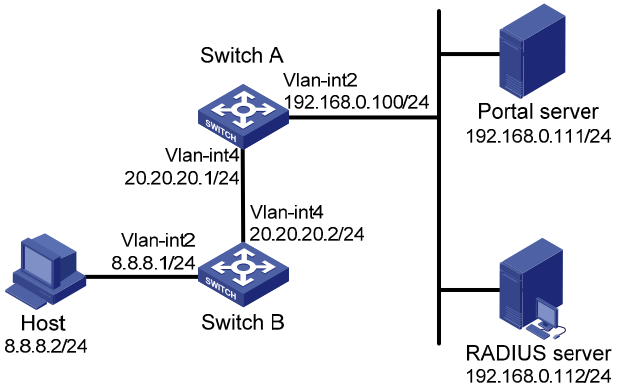
```
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.0.100 255.255.255.0
[Switch-Vlan-interface2] quit
```

Configuring Layer 3 Portal Authentication

Network requirements

- Switch A is configured for Layer 3 portal authentication. Before portal authentication, users can access only the portal server. After passing portal authentication, they can access unrestricted Internet resources.
- The host accesses Switch A through Switch B
- A RADIUS server serves as the authentication/accounting server.

Figure 1-6 Configure Layer 3 portal authentication



Configuration procedure

Note

You need to configure IP addresses for the devices as shown in Figure 1-6 and ensure that routes are available between devices..

Configure Switch A:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<SwitchA> system-view
[SwitchA] radius scheme rs1
```

Set the server type to **extended**.

```
[SwitchA-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[SwitchA-radius-rs1] primary authentication 192.168.0.112
```

```
[SwitchA-radius-rs1] primary accounting 192.168.0.112
[SwitchA-radius-rs1] key authentication radius
[SwitchA-radius-rs1] key accounting radius
```

Specify that the ISP domain name should not be included in the username sent to the RADIUS server.

```
[SwitchA-radius-rs1] user-name-format without-domain
[SwitchA-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[SwitchA] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[SwitchA-isp-dm1] authentication portal radius-scheme rs1
[SwitchA-isp-dm1] authorization portal radius-scheme rs1
[SwitchA-isp-dm1] accounting portal radius-scheme rs1
[SwitchA-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[SwitchA] domain default enable dm1
```

3) Configure portal authentication

Configure the portal server as follows:

- Name: newpt
- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[SwitchA] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

Enable portal authentication on the interface connecting Switch B.

```
[SwitchA] interface vlan-interface 4
[SwitchA–Vlan-interface4] ip address 20.20.20.1 255.255.255.0
[SwitchA–Vlan-interface4] portal server newpt method layer3
[SwitchA–Vlan-interface4] quit
```

Configure the IP address of the interface connected with the portal server.

```
[SwitchA] interface vlan-interface 2
[SwitchA–Vlan-interface2] ip address 192.168.0.100 255.255.255.0
[SwitchA–Vlan-interface2] quit
```

On Switch B, you need to configure a default route to subnet 192.168.0.0/24, setting the next hop as 20.20.20.1. The configuration steps are omitted.

Configuring Direct Portal Authentication with Extended Functions

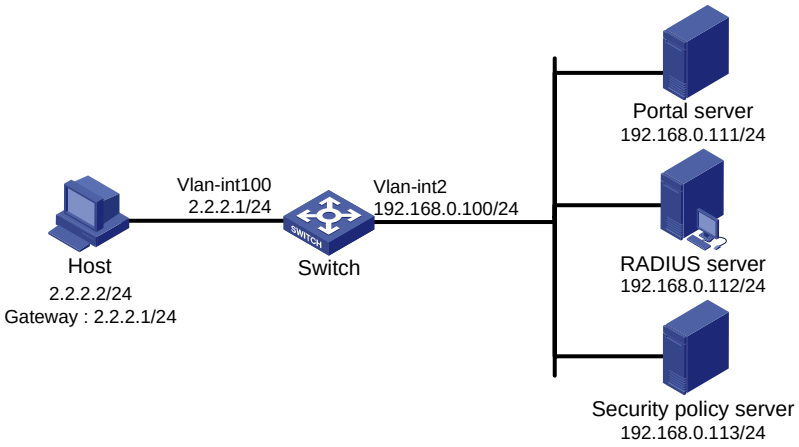
Network requirements

- The host is directly connected to the switch and the switch is configured for direct extended portal authentication. The host is assigned with a public network IP address manually or automatically by a DHCP server. When users using the host have passed identity authentication but have not

passed security authentication, they can access only subnet 192.168.0.0/24. After passing security authentication, they can access unrestricted Internet resources.

- A RADIUS server serves as the authentication/accounting server.

Figure 1-7 Configure direct portal authentication with extended functions



Configuration procedure

Note

You need to configure IP addresses for the devices as shown in Figure 1-7 and ensure that routes are available between devices.

Configure the switch:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Set the server type to **extended**.

```
[Switch-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.112
[Switch-radius-rs1] primary accounting 192.168.0.112
[Switch-radius-rs1] key accounting radius
[Switch-radius-rs1] key authentication radius
[Switch-radius-rs1] user-name-format without-domain
```

Configure the IP address of the security policy server.

```
[Switch-radius-rs1] security-policy-server 192.168.0.113
[Switch-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[Switch] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[Switch] domain default enable dm1
```

3) Configure the ACL (ACL 3000) for restricted resources and the ACL (ACL 3001) for unrestricted resources

---

Note

On the security policy server, you need to specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[Switch] acl number 3000
[Switch-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-adv-3000] quit
[Switch] acl number 3001
[Switch-acl-adv-3001] rule permit ip
[Switch-acl-adv-3001] quit
```

4) Configure portal authentication

Configure the portal server as follows:

- Name: newpt
- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[Switch] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

Enable portal authentication on the interface connecting the host.

```
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 2.2.2.1 255.255.255.0
[Switch–Vlan-interface100] portal server newpt method direct
[Switch] quit
```

Configure the IP address of the interface connected with the portal server.

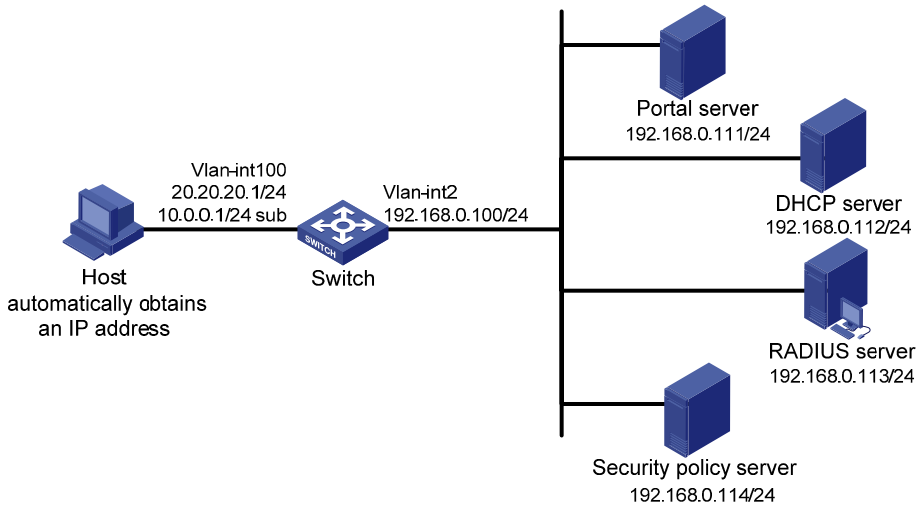
```
[Switch] interface vlan-interface 2
[Switch–Vlan-interface2] ip address 192.168.0.100 255.255.255.0
```

Configuring Re-DHCP Portal Authentication with Extended Functions

Network requirements

- The host is directly connected to the switch and the switch is configured for re-DHCP authentication. The host is assigned with an IP address through the DHCP server. Before portal authentication, the host uses an assigned private IP address. After passing portal authentication, it can get a public IP address.
- When users using the host have passed identity authentication but have not passed security authentication, they can access only subnet 192.168.0.0/24. After passing the security authentication, they can access unrestricted Internet resources.
- A RADIUS server serves as the authentication/accounting server.

Figure 1-8 Configure re-DHCP portal authentication with extended functions



Configuration procedure

Note

- For re-DHCP authentication, you need to configure a public address pool (20.20.20.0/24, in this example) and a private address pool (10.0.0.0/24, in this example) on the DHCP server. The configuration steps are omitted. For DHCP configuration information, refer to *DHCP Configuration* in the *IP Services Volume*.
- For re-DHCP authentication, the switch must be configured as a DHCP relay agent (instead of a DHCP server) and the portal-enabled interface must be configured with a primary IP address (a public IP address) and a secondary IP address (a private IP address).
- You need to configure IP addresses for the devices as shown in Figure 1-8 and ensure that routes are available between devices.

Configure the switch:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<Switch> system-view
[Switch] radius scheme rs1
```

Set the server type to **extended**.

```
[Switch-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[Switch-radius-rs1] primary authentication 192.168.0.113
[Switch-radius-rs1] primary accounting 192.168.0.113
[Switch-radius-rs1] key accounting radius
[Switch-radius-rs1] key authentication radius
[Switch-radius-rs1] user-name-format without-domain
```

Configure the IP address of the security policy server.

```
[Switch-radius-rs1] security-policy-server 192.168.0.114
[Switch-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[Switch] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[Switch-isp-dm1] authentication portal radius-scheme rs1
[Switch-isp-dm1] authorization portal radius-scheme rs1
[Switch-isp-dm1] accounting portal radius-scheme rs1
[Switch-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[Switch] domain default enable dm1
```

3) Configure the ACL (ACL 3000) for restricted resources and the ACL (ACL 3001) for unrestricted resources

---

Note

On the security policy server, you need to specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[Switch] acl number 3000
[Switch-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[Switch-acl-adv-3000] quit
[Switch] acl number 3001
[Switch-acl-adv-3001] rule permit ip
[Switch-acl-adv-3001] quit
```

4) Configure portal authentication

Configure the portal server as follows:

- Name: newpt

- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[Switch] portal server newpt ip 192.168.0.111 key portal port 50100
url http://192.168.0.111/portal
```

Configure the switch as a DHCP relay agent, and enable the invalid address check function.

```
[Switch] dhcp enable
[Switch] dhcp relay server-group 0 ip 192.168.0.112
[Switch] interface vlan-interface 100
[Switch–Vlan-interface100] ip address 20.20.20.1 255.255.255.0
[Switch–Vlan-interface100] ip address 10.0.0.1 255.255.255.0 sub
[Switch-Vlan-interface100] dhcp select relay
[Switch-Vlan-interface100] dhcp relay server-select 0
[Switch-Vlan-interface100] dhcp relay address-check enable
```

Enable re-DHCP portal authentication on the interface connecting the host.

```
[Switch–Vlan-interface100] portal server newpt method redhcp
[Switch–Vlan-interface100] quit
```

Configure the IP address of the interface connected with the portal server.

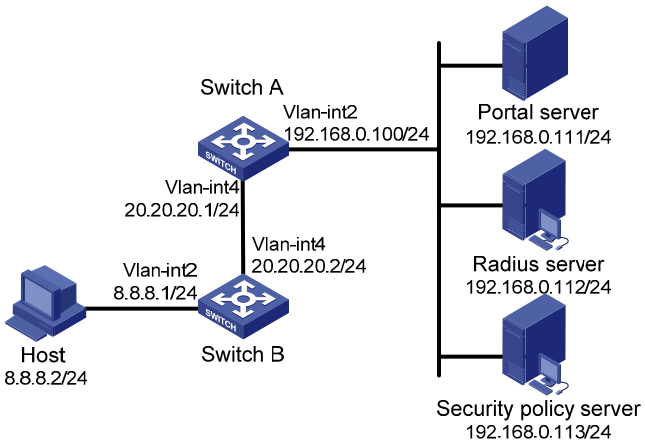
```
[Switch] interface vlan-interface 2
[Switch–Vlan-interface2] ip address 192.168.0.100 255.255.255.0
[Switch–Vlan-interface2] quit
```

Configuring Layer 3 Portal Authentication with Extended Functions

Network requirements

- Switch A is configured for Layer 3 extended portal authentication. When users have passed identity authentication but have not passed security authentication, they can access only subnet 192.168.0.0/24. After passing security authentication, they can access unrestricted Internet resources.
- The host accesses Switch A through Switch B.
- A RADIUS server serves as the authentication/accounting server.

Figure 1-9 Configure Layer 3 portal authentication with extended functions



Configuration procedure

Note

You need to configure IP addresses for the devices as shown in Figure 1-9 and ensure that routes are available between devices.

Configure Switch A:

1) Configure a RADIUS scheme

Create a RADIUS scheme named **rs1** and enter its view.

```
<SwitchA> system-view
[SwitchA] radius scheme rs1
```

Set the server type to **extended**.

```
[SwitchA-radius-rs1] server-type extended
```

Specify the primary authentication server and primary accounting server, and configure the keys for communication with the servers.

```
[SwitchA-radius-rs1] primary authentication 192.168.0.112
[SwitchA-radius-rs1] primary accounting 192.168.0.112
[SwitchA-radius-rs1] key accounting radius
[SwitchA-radius-rs1] key authentication radius
[SwitchA-radius-rs1] user-name-format without-domain
```

Configure the IP address of the security policy server.

```
[SwitchA-radius-rs1] security-policy-server 192.168.0.113
[SwitchA-radius-rs1] quit
```

2) Configure an authentication domain

Create an ISP domain named dm1 and enter its view.

```
[SwitchA] domain dm1
```

Configure the ISP domain to use RADIUS scheme rs1.

```
[SwitchA-isp-dm1] authentication portal radius-scheme rs1
[SwitchA-isp-dm1] authorization portal radius-scheme rs1
[SwitchA-isp-dm1] accounting portal radius-scheme rs1
[SwitchA-isp-dm1] quit
```

Configure dm1 as the default ISP domain, allowing all users to share the authentication and accounting methods of the default domain.

```
[SwitchA] domain default enable dm1
```

3) Configure the ACL (ACL 3000) for restricted resources and the ACL (ACL 3001) for unrestricted resources

Note

On the security policy server, you need to specify ACL 3000 as the isolation ACL and ACL 3001 as the security ACL.

```
[SwitchA] acl number 3000
[SwitchA-acl-adv-3000] rule permit ip destination 192.168.0.0 0.0.0.255
[SwitchA-acl-adv-3000] quit
[SwitchA] acl number 3001
[SwitchA-acl-adv-3001] rule permit ip
[SwitchA-acl-adv-3001] quit
```

4) Configure portal authentication

Configure the portal server as follows:

- Name: newpt
- IP address: 192.168.0.111
- Key: portal
- Port number: 50100
- URL: http://192.168.0.111/portal.

```
[SwitchA] portal server newpt ip 192.168.0.111 key portal port 50100 url http://192.168.0.111/portal
```

Enable portal authentication on the interface connecting Switch B.

```
[SwitchA] interface vlan-interface 4
[SwitchA–Vlan-interface4] ip address 20.20.20.1 255.255.255.0
[SwitchA–Vlan-interface4] portal server newpt method layer3
[SwitchA–Vlan-interface4] quit
```

Configure the IP address of the interface connected with the portal server.

```
[SwitchA] interface vlan-interface 2
[SwitchA–Vlan-interface2] ip address 192.168.0.100 255.255.255.0
[SwitchA–Vlan-interface2] quit
```

On Switch B, you need to configure a default route to subnet 192.168.0.0/24, setting the next hop as 20.20.20.1. The configuration steps are omitted.

Troubleshooting Portal

Inconsistent Keys on the Access Device and the Portal Server

Symptom

When a user is forced to access the portal server, the portal server displays neither the portal authentication page nor any error message. What the user sees is a blank web page.

Analysis

The keys configured on the access device and the portal server are inconsistent, causing CHAP message exchange failure. As a result, the portal server does not display the authentication page.

Solution

- Use the **display portal server** command to display the key for the portal server on the access device and view the key for the access device on the portal server.
- Use the **portal server** command to modify the key on the access device or modify the key for the access device on the portal server to ensure that the keys are consistent.

Incorrect Server Port Number on the Access Device

Symptom

After a user passes the portal authentication, you cannot force the user to log out by executing the **portal delete-user** command on the access device, but the user can log out by using the **disconnect** attribute on the authentication client.

Analysis

When you execute the **portal delete-user** command on the access device to force the user to log out, the access device actively sends a REQ_LOGOUT message to the portal server. The default listening port of the portal server is 50100. However, if the listening port configured on the access device is not 50100, the destination port of the REQ_LOGOUT message is not the actual listening port on the server. Thus, the portal server cannot receive the REQ_LOGOUT message. As a result, you cannot force the user to log out the portal server.

When the user uses the **disconnect** attribute on the client to log out, the portal server actively sends a REQ_LOGOUT message to the access device. The source port is 50100 and the destination port of the ACK_LOGOUT message from the access device is the source port of the REQ_LOGOUT message so that the portal server can receive the ACK_LOGOUT message correctly, no matter whether the listening port is configured on the access device. Therefore, the user can log out the portal server.

Solution

Use the **display portal server** command to display the listening port of the portal server on the access device and use the **portal server** command in the system view to modify it to ensure that it is the actual listening port of the portal server.

Table of Contents

1 Port Security Configuration

When configuring port security, go to these sections for information you are interested in:

Introduction to Port Security

Port Security Overview

Port security is a MAC address-based security mechanism for network access controlling. It is an extension to the existing 802.1X authentication and MAC authentication. It controls the access of unauthorized devices to the network by checking the source MAC address of an inbound frame and the access to unauthorized devices by checking the destination MAC address of an outbound frame.

With port security, you can define various port security modes to make a device learn only legal source MAC addresses, so that you can implement different network security management as needed. When a port security-enabled device detects an illegal frame, it triggers the corresponding port security feature and takes a pre-defined action automatically. This reduces your maintenance workload and greatly enhances system security.

The following types of frames are classified as illegal:

- Received frames with unknown source MAC addresses when MAC address learning is disabled.
- Received frames with unknown source MAC addresses when the number of MAC addresses learned by the port has already reached the upper limit.
- Frames from unauthenticated users.

Note

The security modes of the port security feature provide extended and combined use of 802.1X authentication and MAC authentication and therefore apply to scenarios that require both 802.1X authentication and MAC authentication. For scenarios that require only 802.1X authentication or MAC authentication for access control, however, you are recommended to configure the 802.1X authentication or MAC authentication for simplicity. For information about 802.1X and MAC authentication, refer to *802.1X Configuration* and *MAC Authentication Configuration* in the *Security Volume.*

Port Security Features

NTK

The need to know (NTK) feature checks the destination MAC addresses in outbound frames and allows frames to be sent to only devices passing authentication, thus preventing illegal devices from intercepting network traffic.

Intrusion protection

The intrusion protection feature checks the source MAC addresses in inbound frames and takes a pre-defined action accordingly upon detecting illegal frames. The action may be disabling the port temporarily, disabling the port permanently, or blocking frames from the MAC address for three minutes (unmodifiable).

Trap

The trap feature enables the device to send trap messages upon detecting specified frames that result from, for example, intrusion or user login/logout operations, helping you monitor special activities.

Port Security Modes

Table 1-1 details the port security modes.

Table 1-1 Port security modes

Security mode	Description	Features
noRestrictions	Port security is disabled on the port and access to the port is not restricted.	In this mode, neither the NTK nor the intrusion protection feature is triggered.
autoLearn	In this mode, a port can learn a specified number of MAC addresses and save those addresses as secure MAC addresses. It permits only frames whose source MAC addresses are secure MAC addresses or static MAC addresses configured by using the **mac-address static** command. When the number of secure MAC addresses reaches the upper limit, the port changes to work in secure mode.	In either mode, the device will trigger NTK and intrusion protection upon detecting an illegal frame.
secure	In this mode, a port is disabled from learning MAC addresses and permits only frames whose source MAC addresses are secure MAC addresses or static MAC addresses configured by using the **mac-address static** command.	
userLogin	In this mode, a port performs 802.1X authentication of users in **portbased** mode. A port in this mode can service multiple 802.1X users, but allows only one at a moment.	In this mode, neither NTK nor intrusion protection will be triggered.

Security mode	Description	Features
userLoginSecure	In this mode, a port performs 802.1X authentication of users in **portbased** mode and services only one user passing 802.1X authentication.	In any of these modes, the device will trigger NTK and intrusion protection upon detecting an illegal frame.
userLoginWithOUI	Similar to the userLoginSecure mode, a port in this mode performs 802.1X authentication of users and services only one user passing 802.1X authentication. The port also permits frames from a user whose MAC address contains a specified OUI (organizationally unique identifier).	
macAddressWithRadius	In this mode, a port performs MAC authentication of users.	
macAddressOrUserLoginSecure	This mode is the combination of the userLoginSecure and macAddressWithRadius modes, with 802.1X authentication having a higher priority The port performs MAC authentication upon receiving non-8021.x frames and performs 802.1X authentication upon receiving 802.1X frames.	
macAddressElseUserLoginSecure	This mode is the combination of the macAddressWithRadius and userLoginSecure modes, with MAC authentication having a higher priority. • Upon receiving a non-802.1X frame, a port in this mode performs only MAC authentication. • Upon receiving an 802.1X frame, the port performs MAC authentication and then, if MAC authentication fails, 802.1X authentication.	
userLoginSecureExt	In this mode, a port performs 802.1X authentication of users in macbased mode and supports multiple 802.1X users.	
macAddressOrUserLoginSecureExt	This mode is similar to the macAddressOrUserLoginSecure mode, except that it supports multiple 802.1X and MAC authentication users on the port.	
macAddressElseUserLoginSecureExt	This mode is similar to the macAddressElseUserLoginSecure mode, except that it supports multiple 802.1X and MAC authentication users on the port.	

Note

- Currently, port security supports two authentication methods: 802.1X and MAC authentication. Different port security modes employ different authentication methods or different combinations of authentication methods.
- The maximum number of users a port supports is the lesser of the maximum number of secure MAC addresses or the maximum number of authenticated users the security mode supports. For example, in userLoginSecureExt mode, the maximum number of users a port supports is the lesser of the maximum number of secure MAC addresses configured or the maximum number of users that 802.1X supports.

Tip

These security mode naming rules may help you remember the modes:

- **userLogin** specifies port-based 802.1X authentication.
- **macAddress** specifies MAC address authentication.
- **Else** specifies that the authentication method before **Else** is applied first. If the authentication fails, the protocol type of the authentication request determines whether to turn to the authentication method following the **Else**.
- In a security mode with **Or**, the protocol type of the authentication request determines which authentication method is to be used. However, 802.1X authentication is preferred by wireless users.
- **userLogin with Secure** specifies MAC-based 802.1X authentication.
- **Ext** indicates allowing multiple 802.1X users to be authenticated and get online. A security mode without **Ext** allows only one 802.1X user to be authenticated and get online.

Port Security Configuration Task List

Complete the following tasks to configure port security:

Task		Remarks
Enabling Port Security		Required
Setting the Maximum Number of Secure MAC Addresses		Optional
Setting the Port Security Mode		Required
Configuring Port Security Features	Configuring NTK	Optional
	Configuring Intrusion Protection	Choose one or more features as required.
	Configuring Trapping	
Configuring Secure MAC Addresses		Optional
Ignoring Authorization Information from the Server		Optional

Enabling Port Security

Configuration Prerequisites

Before enabling port security, you need to disable 802.1X and MAC authentication globally.

Configuration Procedure

Follow these steps to enable port security:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable port security	**port-security enable**	Required Disabled by default

Note that:

1) Enabling port security resets the following configurations on a port to the bracketed defaults. Then, values of these configurations cannot be changed manually; the system will adjust them based on the port security mode automatically:

- 802.1X (disabled), port access control method (macbased), and port access control mode (auto)
- MAC authentication (disabled)

2) Disabling port security resets the following configurations on a port to the bracketed defaults:

- Port security mode (noRestrictions)
- 802.1X (disabled), port access control method (macbased), and port access control mode (auto)
- MAC authentication (disabled)

3) Port security cannot be disabled if there is any user present on a port.

Note

- For detailed 802.1X configuration, refer to *802.1X Configuration* in the *Security Volume*.
- For detailed MAC-based authentication configuration, refer to *MAC Authentication Configuration* in the *Security Volume*.

Setting the Maximum Number of Secure MAC Addresses

With port security enabled, more than one authenticated user is allowed on a port. The number of authenticated users allowed, however, cannot exceed the specified upper limit.

By setting the maximum number of secure MAC addresses allowed on a port, you can:

- Control the maximum number of users who are allowed to access the network through the port.
- Control the number of secure MAC addresses that can be added with port security.

Follow these steps to set the maximum number of secure MAC addresses allowed on a port:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Set the maximum number of secure MAC addresses allowed on a port	**port-security max-mac-count** *count-value*	Required Not limited by default

Note

This configuration is different from that of the maximum number of MAC addresses that can be leaned by the port in MAC address management.

Setting the Port Security Mode

Configuration Prerequisites

Before setting the port security mode, ensure that:

- 802.1X is disabled, the port access control method is **macbased**, and the port access control mode is **auto**.
- MAC authentication is disabled.
- The port does not belong to any aggregation group or service loopback group.

The above requirements must be all met. Otherwise, you will see an error message and your configuration will fail. On the other hand, after setting the port security mode on a port, you cannot change any configurations of the first three requirements.

Note

- With port security disabled, you can configure the port security mode, but your configuration does not take effect.
- You cannot change the port security mode of a port when any user is present on the port.
- Before configuring the port to operate in autoLearn mode, set the maximum number of secure MAC addresses allowed on a port.

Configuring Procedure

Follow these steps to enable any other port security mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set an OUI value for user authentication	**port-security oui** *oui-value* **index** *index-value*	Optional Not configured by default. The command is required for the **userlogin-withoui** mode.
Enter interface view	**interface** *interface-type interface-number*	—
Set the port security mode	**port-security port-mode** { **autolearn** | **mac-authentication** | **mac-else-userlogin-secure** | **mac-else-userlogin-secure-ext** | **secure** | **userlogin** | **userlogin-secure** | **userlogin-secure-ext** | **userlogin-secure-or-mac** | **userlogin-secure-or-mac-ext** | **userlogin-withoui** }	Required By default, a port operates in noRestrictions mode.

Note

- You cannot change the maximum number of secure MAC addresses allowed on a port that operates in autoLearn mode.
- OUI, defined by IEEE, is the first 24 bits of the MAC address and uniquely identifies a device vendor.
- You can configure multiple OUI values. However, a port in userLoginWithOUI mode allows only one 802.1X user and one user whose MAC address contains a specified OUI.
- After enabling port security, you can change the port security mode of a port only when the port is operating in noRestrictions mode, the default mode. To change the port security mode of a port operating in any other mode, use the **undo port-security port-mode** command to restore the default port security mode at first.
- You cannot change the port security mode of a port with users online.

Configuring Port Security Features

Configuring NTK

The need to know (NTK) feature checks the destination MAC addresses in outbound frames to allow frames to be forwarded to only devices passing authentication. The NTK feature supports three modes:

- **ntkonly**: Forwards only frames destined for authenticated MAC addresses.
- **ntk-withbroadcasts**: Forwards only frames destined for authenticated MAC addresses or the broadcast address.
- **ntk-withmulticasts**: Forwards only frames destined for authenticated MAC addresses, multicast addresses, or the broadcast address.

By default, NTK is disabled on a port and the port forwards all frames. With NTK configured, a port will discard any unicast packet with an unknown MAC address no matter in which mode it operates.

Follow these steps to configure the NTK feature:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the NTK feature	**port-security ntk-mode** { **ntk-withbroadcasts** | **ntk-withmulticasts** | **ntkonly** }	Required By default, NTK is disabled on a port and all frames are allowed to be sent.

Note

Support for the NTK feature depends on the port security mode.

Configuring Intrusion Protection

The intrusion protection enables a device to perform either of the following security policies when it detects illegal frames:

- **blockmac**: Adds the source MAC addresses of illegal frames to the blocked MAC addresses list and discards frames with blocked source MAC addresses. A blocked MAC address is restored to normal after being blocked for three minutes, which is fixed and cannot be changed.
- **disableport**: Disables the port permanently.
- **disableport-temporarily**: Disables the port for a specified period of time. Use the **port-security timer disableport** command to set the period.

Follow these steps to configure the intrusion protection feature:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure the intrusion protection feature	**port-security intrusion-mode** { **blockmac** | **disableport** | **disableport-temporarily** }	Required By default, intrusion protection is disabled.
Return to system view	**quit**	—
Set the silence timeout during which a port remains disabled	**port-security timer disableport** *time-value*	Optional 20 seconds by default

Note

On a port operating in either the macAddressElseUserLoginSecure mode or the macAddressElseUserLoginSecureExt mode, intrusion protection is triggered only after both MAC authentication and 802.1X authentication for the same frame fail.

Configuring Trapping

The trapping feature enables a device to send trap information in response to four types of events:

- **addresslearned**: A port learns a new address.
- **dot1xlogfailure/dot1xlogon/dot1xlogoff**: A port learns 802.1x authentication failure/successful 802.1x authentication/802.1x user logoff.
- **ralmlogfailure**/**ralmlogoff**: A port learns MAC authentication failure/MAC authentication user logoff.
- **intrusion**: A port learns illegal frames.

Follow these steps to configure port security trapping:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable port security traps	**port-security trap** { **addresslearned** | **dot1xlogfailure** | **dot1xlogoff** | **dot1xlogon** | **intrusion** | **ralmlogfailure** | **ralmlogoff** | **ralmlogon** }	Required By default, no port security trap is enabled.

Configuring Secure MAC Addresses

Secure MAC addresses are special MAC addresses. They never age out or get lost if saved before the device restarts. One secure MAC address can be added to only one port in the same VLAN. Thus, you can bind a MAC address to one port in the same VLAN.

Secure MAC addresses can be:

- Learned by a port working in autoLearn mode.
- Manually configured through the command line interface (CLI) or management information base (MIB).

When the maximum number of secure MAC addresses is reached, no more can be added. The port allows only the packets with the source MAC address being the secure MAC address.

Configuration Prerequisites

- Enable port security
- Set the maximum number of secure MAC addresses allowed on the port
- Set the port security mode to autoLearn

Configuration Procedure

Follow these steps to configure a secure MAC address:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Configure a secure MAC address	In system view	**port-security mac-address security** *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id*	Required Use either approach No secure MAC address is configured by default.
	In interface view	**interface** *interface-type interface-number*	
		port-security mac-address security *mac-address* **vlan** *vlan-id*	

Note

The configured secure MAC addresses are saved in the configuration file and will not get lost when the port goes up or goes down. After you save the configuration file, the secure MAC address saved in the configuration file are maintained even after the device restarts.

Ignoring Authorization Information from the Server

After an 802.1X user or MAC authenticated user passes RADIUS authentication, the RADIUS server delivers the authorization information to the device. You can configure a port to ignore the authorization information from the RADIUS server.

Follow these steps to configure a port to ignore the authorization information from the RADIUS server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter Ethernet port view	**interface** *interface-type interface-number*	—
Ignore the authorization information from the RADIUS server	**port-security authorization ignore**	Required By default, a port uses the authorization information from the RADIUS server.

Displaying and Maintaining Port Security

To do…	Use the command…	Remarks
Display port security configuration information, operation information, and statistics about one or more ports or all ports	**display port-security** [**interface** *interface-list*]	Available in any view
Display information about secure MAC addresses	**display port-security mac-address security** [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]	Available in any view

To do…	Use the command…	Remarks
Display information about blocked MAC addresses	**display port-security mac-address block** [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]	Available in any view

Port Security Configuration Examples

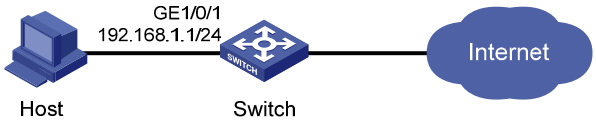
Configuring the autoLearn Mode

Network requirements

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow up to 64 users to access the port without authentication and permit the port to learn and add the MAC addresses of the users as secure MAC addresses.
- After the number of secure MAC addresses reaches 64, the port stops learning MAC addresses. If any frame with an unknown MAC address arrives, intrusion protection is triggered and the port is disabled and stays silence for 30 seconds.

Figure 1-1 Network diagram for configuring the autoLearn mode



Configuration procedure

1) Configure port security

Enable port security.

```
<Switch> system-view
[Switch] port-security enable
```

Enable intrusion protection trap.

```
[Switch] port-security trap intrusion
[Switch] interface gigabitethernet 1/0/1
```

Set the maximum number of secure MAC addresses allowed on the port to 64.

```
[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64
```

Set the port security mode to autoLearn.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
```

Configure the port to be silent for 30 seconds after the intrusion protection feature is triggered.

```
[Switch-GigabitEthernet1/0/1] port-security intrusion-mode disableport-temporarily
[Switch-GigabitEthernet1/0/1] quit
[Switch] port-security timer disableport 30
```

2) Verify the configuration

After completing the above configurations, you can use the following command to view the port security configuration information:

```
<Switch> display port-security interface gigabitethernet 1/0/1
```

```
 Equipment port-security is enabled
 Intrusion trap is enabled
 Disableport Timeout: 30s
 OUI value:

GigabitEthernet1/0/1 is link-up
   Port mode is autoLearn
   NeedToKnow mode is disabled
   Intrusion Protection mode is DisablePortTemporarily
   Max MAC address number is 64
   Stored MAC address number is 0
   Authorization is permitted
```

As shown in the output, the maximum number of secure MAC addresses allowed on the port is 64, the port security mode is autoLearn, the intrusion protection trap is enabled, and the intrusion protection action is to disable the port (DisablePortTemporarily) for 30 seconds.

You can also use the above command repeatedly to track the number of MAC addresses learned by the port, or use the **display this** command in interface view to display the secure MAC addresses learned, as shown below:

```
<Switch> system-view
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] display this
#
interface GigabitEthernet1/0/1
 port-security max-mac-count 64
 port-security port-mode autolearn
 port-security intrusion-mode disableport-temporarily
 port-security mac-address security 0002-0000-0015 vlan 1
 port-security mac-address security 0002-0000-0014 vlan 1
 port-security mac-address security 0002-0000-0013 vlan 1
 port-security mac-address security 0002-0000-0012 vlan 1
 port-security mac-address security 0002-0000-0011 vlan 1
#
```

Issuing the **display port-security interface** command after the number of MAC addresses learned by the port reaches 64, you will see that the port security mode has changed to secure. When any frame with a new MAC address arrives, intrusion protection is triggered and you will see trap messages as follows:

```
#May  2 03:15:55:871 2000 Switch PORTSEC/1/VIOLATION:Traph3cSecureViolation
 A intrusion occurs!
 IfIndex: 9437207
 Port: 9437207
 MAC Addr: 0.2.0.0.0.21
 VLAN ID: 1
 IfAdminStatus: 1
```

In addition, you will see that the port security feature has disabled the port if you issue the following command:

```
[Switch-GigabitEthernet1/0/1] display interface gigabitethernet 1/0/1
```

```
GigabitEthernet1/0/1 current state:  Port Security Disabled
IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 000f-cb00-5558
Description: GigabitEthernet1/0/1 Interface
......
```

The port should be re-enabled 30 seconds later.

```
[Switch-GigabitEthernet1/0/1] display interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 current state: UP
IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 000f-cb00-5558
Description: GigabitEthernet1/0/1 Interface
......
```

Now, if you manually delete several secure MAC addresses, the port security mode of the port will be restored to autoLearn, and the port will be able to learn MAC addresses again.

Configuring the userLoginWithOUI Mode

Network requirements

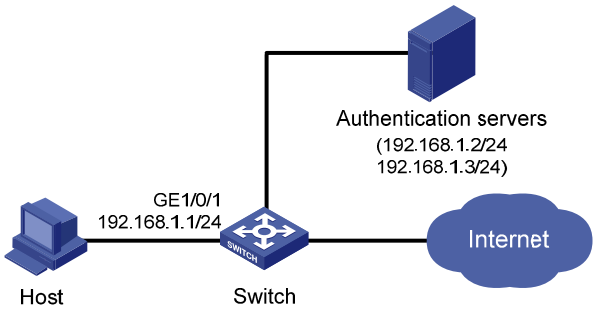
The client is connected to the switch through port GigabitEthernet 1/0/1. The switch authenticates the client by the RADIUS server. If the authentication succeeds, the client is authorized to access the Internet.

- RADIUS server 192.168.1.2 functions as the primary authentication server and the secondary accounting server, and RADIUS server 192.168.1.3 functions as the secondary authentication server and the primary accounting server. The shared key for authentication is name, and that for accounting is money.
- All users belong to default domain sun, which can accommodate up to 30 users.
- The RADIUS server response timeout time is five seconds and the maximum number of RADIUS packet retransmission attempts is five. The switch sends real-time accounting packets to the RADIUS server at an interval of 15 minutes, and sends user names without domain names to the RADIUS server.

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow only one 802.1X user to be authenticated.
- Allow up to 16 OUI values to be configured and allow one additional user whose MAC address has an OUI among the configured ones to access the port.

Figure 1-2 Network diagram for configuring the userLoginWithOUI mode



Configuration procedure

Note

- The following configuration steps cover some AAA/RADIUS configuration commands. For details about the commands, refer to *AAA Configuration* in the *Security Volume*.
- Configurations on the host and RADIUS servers are omitted.

1) Configure the RADIUS protocol

Configure a RADIUS scheme named **radsun**.

```
<Switch> system-view
[Switch] radius scheme radsun
[Switch-radius-radsun] primary authentication 192.168.1.2
[Switch-radius-radsun] primary accounting 192.168.1.3
[Switch-radius-radsun] secondary authentication 192.168.1.3
[Switch-radius-radsun] secondary accounting 192.168.1.2
[Switch-radius-radsun] key authentication name
[Switch-radius-radsun] key accounting money
[Switch-radius-radsun] timer response-timeout 5
[Switch-radius-radsun] retry 5
[Switch-radius-radsun] timer realtime-accounting 15
[Switch-radius-radsun] user-name-format without-domain
[Switch-radius-radsun] quit
```

Configure an ISP domain named **sun**.

```
[Switch] domain sun
[Switch-isp-sun] authentication default radius-scheme radsun
[Switch-isp-sun] authorization default radius-scheme radsun
[Switch-isp-sun] accounting default radius-scheme radsun
[Switch-isp-sun] access-limit enable 30
[Switch-isp-sun] quit
```

2) Configure port security

Enable port security.

```
[Switch] port-security enable
```

Add five OUI values.

```
[Switch] port-security oui 1234-0100-1111 index 1
[Switch] port-security oui 1234-0200-1111 index 2
[Switch] port-security oui 1234-0300-1111 index 3
[Switch] port-security oui 1234-0400-1111 index 4
[Switch] port-security oui 1234-0500-1111 index 5
[Switch] interface gigabitethernet 1/0/1
```

Set the port security mode to userLoginWithOUI.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode userlogin-withoui
```

3) Verify the configuration

After completing the above configurations, you can use the following command to view the configuration information of the RADIUS scheme named **radsun**:

```
<Switch> display radius scheme radsun
SchemeName  : radsun
  Index : 1                            Type : standard
  Primary Auth IP  : 192.168.1.2      Port : 1812   State : active
  Primary Acct IP  : 192.168.1.3      Port : 1813   State : active
  Second  Auth IP  : 192.168.1.3      Port : 1812   State : active
  Second  Acct IP  : 192.168.1.2      Port :1813    State : active
  Auth Server Encryption Key : name
  Acct Server Encryption Key : money
  Accounting-On packet disable, send times : 5 , interval : 3s
  Interval for timeout(second)                              : 5
  Retransmission times for timeout                          : 5
  Interval for realtime accounting(minute)                  : 15
  Retransmission times of realtime-accounting packet        : 5
  Retransmission times of stop-accounting packet            : 500
  Quiet-interval(min)                                       : 5
  Username format                                           : without-domain
  Data flow unit                                            : Byte
  Packet unit                                               : one
```

Use the following command to view the configuration information of the ISP domain named **sun**:

```
<Switch> display domain sun
   Domain = sun
   State = Active
   Access-limit = 30
   Accounting method = Required
   Default authentication scheme      : radius=radsun
   Default authorization scheme       : radius=radsun
   Default accounting scheme          : radius=radsun
   Domain User Template:
   Idle-cut = Disabled
   Self-service = Disabled
```

Use the following command to view the port security configuration information:

```
<Switch> display port-security interface gigabitethernet 1/0/1
 Equipment port-security is enabled
 Trap is disabled
 Disableport Timeout: 20s
 OUI value:
   Index is 1,  OUI value is 123401
   Index is 2,  OUI value is 123402
   Index is 3,  OUI value is 123403
   Index is 4,  OUI value is 123404
   Index is 5,  OUI value is 123405

 GigabitEthernet1/0/1 is link-up
   Port mode is userLoginWithOUI
```

```
  NeedToKnow mode is disabled
  Intrusion Protection mode is NoAction
  Max MAC address number is not configured
  Stored MAC address number is 0
  Authorization is permitted
```

After an 802.1X user gets online, you can see that the number of secure MAC addresses stored is 1. You can also use the following command to view information about 802.1X users:

```
<Switch> display dot1x interface gigabitethernet 1/0/1
 Equipment 802.1X protocol is enabled
 CHAP authentication is enabled
 EAD quick deploy is disabled

  Configuration: Transmit Period   30 s,  Handshake Period       15 s
                 Quiet Period      60 s,  Quiet Period Timer is disabled
                 Supp Timeout      30 s,  Server Timeout        100 s
                 The maximal retransmitting times    2
  EAD quick deploy configuration:
               EAD timeout:    30m

 The maximum 802.1X user resource number is 1024 per slot
 Total current used 802.1X resource number is 1

 GigabitEthernet1/0/1  is link-up
   802.1X protocol is enabled
   Handshake is enabled
   The port is an authenticator
   Authentication Mode is Auto
   Port Control Type is Mac-based
   Guest VLAN: 0
   Max number of on-line users is 256

   EAPOL Packet: Tx 16331, Rx 102
   Sent EAP Request/Identity Packets : 16316
        EAP Request/Challenge Packets: 6
        EAP Success Packets: 4, Fail Packets: 5
   Received EAPOL Start Packets : 6
            EAPOL LogOff Packets: 2
            EAP Response/Identity Packets : 80
            EAP Response/Challenge Packets: 6
            Error Packets: 0
 1. Authenticated user : MAC address: 0002-0000-0011

   Controlled User(s) amount to 1
```

In addition, the port allows an additional user whose MAC address has an OUI among the specified OUIs to access the port. You can use the following command to view the related information:

```
<Switch> display mac-address interface gigabitethernet 1/0/1
MAC ADDR        VLAN ID   STATE          PORT INDEX               AGING TIME(s)
```

```
1234-0300-0011  1         Learned        GigabitEthernet1/0/1    AGING

  ---  1 mac address(es) found  ---
```

Configuring the macAddressElseUserLoginSecure Mode

Network requirements

The client is connected to the switch through GigabitEthernet 1/0/1. The switch authenticates the client by the RADIUS server. If the authentication succeeds, the client is authorized to access the Internet.

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow more than one MAC authenticated user to log on.
- For 802.1X users, perform MAC authentication first and then, if MAC authentication fails, 802.1X authentication. Allow only one 802.1X user to log on.
- Set fixed username and password for MAC-based authentication. Set the total number of MAC authenticated users and 802.1X-authenticated users to 64.
- Enable NTK to prevent frames from being sent to unknown MAC addresses.

See Figure 1-2.

Configuration procedure

Note

- Configurations on the host and RADIUS servers are omitted.

---

1) Configure the RADIUS protocol

The required RADIUS authentication/accounting configurations are the same as those in Configuring the userLoginWithOUI Mode.

2) Configure port security

Enable port security.

```
<Switch> system-view
[Switch] port-security enable
```

Configure a MAC authentication user, setting the user name and password to aaa and 123456 respectively.

```
[Switch] mac-authentication user-name-format fixed account aaa password simple 123456
[Switch] interface gigabitethernet 1/0/1
```

Set the maximum number of secure MAC addresses allowed on the port to 64.

```
[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64
```

Set the port security mode to macAddressElseUserLoginSecure.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode mac-else-userlogin-secure
```

Set the NTK mode of the port to ntkonly.

```
[Switch-GigabitEthernet1/0/1] port-security ntk-mode ntkonly
```

3) Verify the configuration

After completing the above configurations, you can use the following command to view the port security configuration information:

```
<Switch> display port-security interface gigabitethernet 1/0/1
 Equipment port-security is enabled
 Trap is disabled
 Disableport Timeout: 20s
 OUI value:

 GigabitEthernet1/0/1 is link-up
   Port mode is macAddressElseUserLoginSecure
   NeedToKnow mode is NeedToKnowOnly
   Intrusion Protection mode is NoAction
   Max MAC address number is 64
   Stored MAC address number is 0
   Authorization is permitted
```

Use the following command to view MAC authentication information:

```
<Switch> display mac-authentication interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 is link-up
  MAC address authentication is enabled
  Authenticate success: 3, failed: 7
  Current online user number is 3
    MAC ADDR         Authenticate state           AuthIndex
    1234-0300-0011   MAC_AUTHENTICATOR_SUCCESS     13
    1234-0300-0012   MAC_AUTHENTICATOR_SUCCESS     14
    1234-0300-0013   MAC_AUTHENTICATOR_SUCCESS     15
```

Use the following command to view 802.1X authentication information:

```
<Switch> display dot1x interface gigabitethernet 1/0/1
 Equipment 802.1X protocol is enabled
 CHAP authentication is enabled
 EAD quick deploy is disabled

 Configuration: Transmit Period   30 s,  Handshake Period       15 s
                Quiet Period      60 s,  Quiet Period Timer is disabled
                Supp Timeout      30 s,  Server Timeout        100 s
                The maximal retransmitting times    2
 EAD quick deploy configuration:
                EAD timeout:    30m

 Total maximum 802.1X user resource number is 1024 per slot
 Total current used 802.1X resource number is 1

 GigabitEthernet1/0/1  is link-up
   802.1X protocol is enabled
   Handshake is enabled
   The port is an authenticator
```

```
  Authentication Mode is Auto
  Port Control Type is Mac-based
  Guest VLAN: 0
  Max number of on-line users is 256

  EAPOL Packet: Tx 16331, Rx 102
  Sent EAP Request/Identity Packets : 16316
       EAP Request/Challenge Packets: 6
       EAP Success Packets: 4, Fail Packets: 5
  Received EAPOL Start Packets : 6
           EAPOL LogOff Packets: 2
           EAP Response/Identity Packets : 80
           EAP Response/Challenge Packets: 6
           Error Packets: 0
1. Authenticated user : MAC address: 0002-0000-0011

  Controlled User(s) amount to 1
```

In addition, as NTK is enabled, frames with unknown destination MAC addresses, multicast addresses, and broadcast addresses should be discarded.

Troubleshooting Port Security

Cannot Set the Port Security Mode

Symptom

Cannot set the port security mode.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
 Error:When we change port-mode, we should first change it to noRestrictions, then change it to the other.
```

Analysis

For a port working in a port security mode other than noRestrictions, you cannot change the port security mode by using the **port-security port-mode** command directly.

Solution

Set the port security mode to noRestrictions first.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
```

Cannot Configure Secure MAC Addresses

Symptom

Cannot configure secure MAC addresses.

```
[Switch-GigabitEthernet1/0/1] port-security mac-address security 1-1-2 vlan 1
 Error:Can not operate security MAC address for current port mode is not autoLearn!
```

Analysis

No secure MAC address can be configured on a port operating in a port security mode other than autoLearn.

Solution

Set the port security mode to autoLearn.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
[Switch-GigabitEthernet1/0/1] port-security mac-address security 1-1-2 vlan 1
```

Cannot Change Port Security Mode When a User Is Online

Symptom

Port security mode cannot be changed when an 802.1X-authenticated or MAC authenticated user is online.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
 Error:Cannot configure port-security for there is 802.1X user(s) on line on port GigabitEthernet1/0/1.
```

Analysis

Changing port security mode is not allowed when an 802.1X-authenticated or MAC authenticated user is online.

Solution

Use the **cut** command to forcibly disconnect the user from the port before changing the port security mode.

```
[Switch-GigabitEthernet1/0/1] quit
[Switch] cut connection interface gigabitethernet 1/0/1
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
```

Table of Contents

1 IP Source Guard Configuration

When configuring IP Source Guard, go to these sections for information you are interested in:

IP Source Guard Overview

By filtering packets on a per-port basis, IP source guard prevents illegal packets from traveling through, thus improving the network security. After receiving a packet, the port looks up the key attributes (including IP address, MAC address and VLAN tag) of the packet in the binding entries of the IP source guard. If there is a match, the port forwards the packet. Otherwise, the port discards the packet.

IP source guard filters packets based on the following types of binding entries:

- IP-port binding entry
- MAC-port binding entry
- IP-MAC-port binding entry
- IP-VLAN-port binding entry
- MAC-VLAN-port binding entry
- IP-MAC-VLAN-port binding entry

You can manually set static binding entries, or use DHCP snooping or DHCP relay to provide dynamic binding entries. Binding is on a per-port basis. After a binding entry is configured on a port, it is effective only to the port.

Caution

Enabling IP source guard on a port is mutually exclusive with adding the port to an aggregation group and adding the port to a service loopback group.

Configuring a Static Binding Entry

Follow these steps to configure a static binding entry:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Configure a static binding entry	**user-bind** { **ip-address** *ip-address* | **ip-address** *ip-address* **mac-address** *mac-address* | **mac-address** *mac-address* } [**vlan** *vlan-id*]	Required No static binding entry exists by default.

Note

- The system does not support repeatedly binding a binding entry to one port.
- For products supporting multi-port binding, a binding entry can be configured to multiple ports; for products that do not support multi-port binding, a binding entry can be configured to only one port.
- Supported binding entry types vary by device.
- In a valid binding entry, the MAC address cannot be all 0s, all Fs (a broadcast address), or a multicast address, and the IP address can only be a Class A, Class B, or Class C address and can be neither 127.x.x.x nor 0.0.0.0.
- A static binding entry can be configured on only Layer-2 Ethernet ports.

Configuring Dynamic Binding Function

After the dynamic binding function is enabled on a port, IP source guard will receive and process corresponding DHCP snooping or DHCP relay entries, which contain such information as MAC address, IP address, VLAN tag, port information or entry type. It adds the obtained information to the dynamic binding entries to enable the port to filter packets according to the binding entries.

Follow these steps to configure port filtering:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure dynamic binding function	**ip check source** { **ip-address** | **ip-address mac-address** | **mac-address** }	Required Not configured by default

Note

- The dynamic binding function can be configured on Layer-2 Ethernet ports and VLAN interfaces.
- A port takes only the latest dynamic binding entries configured on it.

Displaying and Maintaining IP Source Guard

To do…	Use the command…	Remarks
Display information about static binding entries	**display user-bind** [**interface** *interface-type interface-number* | **ip-address** *ip-address* | **mac-address** *mac-address*]	Available in any view
Display information about dynamic binding entries	**display ip check source** [**interface** *interface-type interface-number* | **ip-address** *ip-address* | **mac-address** *mac-address*]	Available in any view

IP Source Guard Configuration Examples

Static Binding Entry Configuration Example

Network requirements

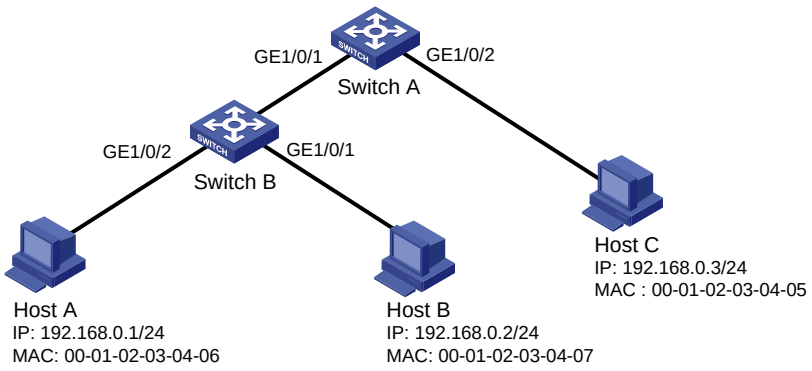
As shown in Figure 1-1, Host A and Host B are connected to ports GigabitEthernet 1/0/2 and GigabitEthernet 1/0/1 of Switch B respectively, Host C is connected to port GigabitEthernet 1/0/2 of Switch A, and Switch B is connected to port GigabitEthernet 1/0/1 of Switch A.

Configure static binding entries on Switch A and Switch B to meet the following requirements:

- On port GigabitEthernet 1/0/2 of Switch A, only IP packets from Host C can pass.
- On port GigabitEthernet 1/0/1 of Switch A, only IP packets from Host A can pass.
- On port GigabitEthernet 1/0/2 of Switch B, only IP packets from Host A can pass.
- On port GigabitEthernet 1/0/1 of Switch B, only IP packets from Host B can pass.

Network diagram

Figure 1-1 Network diagram for configuring static binding entries



Configuration procedure

1) Configure Switch A

Configure the IP addresses of various interfaces (omitted).

Configure port GigabitEthernet 1/0/2 of Switch A to allow only IP packets with the source MAC address of 00-01-02-03-04-05 and the source IP address of 192.168.0.3 to pass.

```
<SwitchA> system-view
[SwitchA] interface gigabitethernet 1/0/2
```

```
[SwitchA-GigabitEthernet1/0/2] user-bind ip-address 192.168.0.3 mac-address 0001-0203-0405
[SwitchA-GigabitEthernet1/0/2] quit
```

Configure port GigabitEthernet 1/0/1 of Switch A to allow only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 to pass.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] user-bind ip-address 192.168.0.1 mac-address 0001-0203-0406
```

2) Configure Switch B

Configure the IP addresses of various interfaces (omitted).

Configure port GigabitEthernet 1/0/2 of Switch B to allow only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 to pass.

```
<SwitchB> system-view
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] user-bind ip-address 192.168.0.1 mac-address 0001-0203-0406
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure port GigabitEthernet 1/0/1 of Switch B to allow only IP packets with the source MAC address of 00-01-02-03-04-07 and the source IP address of 192.168.0.2 to pass.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] user-bind ip-address 192.168.0.2 mac-address 0001-0203-0407
```

3) Verify the configuration

On Switch A, static binding entries are configured successfully.

```
<SwitchA> display user-bind
Total entries found: 2
 MAC               IP              Vlan   Port                         Status
 0001-0203-0405   192.168.0.3      N/A    GigabitEthernet1/0/2         Static
 0001-0203-0406   192.168.0.1      N/A    GigabitEthernet1/0/1         Static
```

On Switch B, static binding entries are configured successfully.

```
<SwitchB> display user-bind
Total entries found: 2
MAC               IP              Vlan   Port                         Status
 0001-0203-0406   192.168.0.1      N/A    GigabitEthernet1/0/2         Static
 0001-0203-0407   192.168.0.2      N/A    GigabitEthernet1/0/1         Static
```

Dynamic Binding Function Configuration Example

Network requirements

Switch A connects to Client A and the DHCP server through ports GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 respectively. DHCP snooping is enabled on Switch A.

Detailed requirements are as follows:

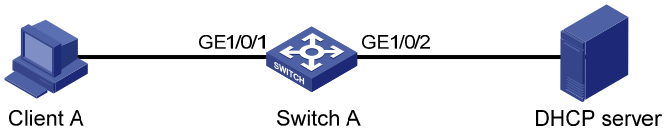
- Client A (with the MAC address of 00-01-02-03-04-06) obtains an IP address through the DHCP server.
- On Switch A, create a DHCP snooping entry for Client A.
- On port GigabitEthernet 1/0/1 of Switch A, enable dynamic binding function to prevent attackers from using forged IP addresses to attack the server.

Note

For detailed configuration of a DHCP server, refer to *DHCP Configuration* in the *IP Service Volume.*

Network diagram

Figure 1-2 Network diagram for configuring dynamic binding function



Configuration procedure

1) Configure Switch A

Configure dynamic binding function on port GigabitEthernet 1/0/1.

```
<SwitchA> system-view
[SwitchA] interface gigabitethernet1/0/1
[SwitchA-GigabitEthernet1/0/1] ip check source ip-address mac-address
[SwitchA-GigabitEthernet1/0/1] quit
```

Enable DHCP snooping.

```
[SwitchA] dhcp-snooping
```

Configure the port connecting to the DHCP server as a trusted port.

```
[SwitchA] interface gigabitethernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] dhcp-snooping trust
[SwitchA-GigabitEthernet1/0/2] quit
```

2) Verify the configuration

Display dynamic binding function is configured successfully on port GigabitEthernet 1/0/1.

```
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] display this
#
interface GigabitEthernet1/0/1
 ip check source ip-address mac-address
#
return
```

Display the dynamic binding entries that port GigabitEthernet 1/0/1 has obtained from DHCP snooping.

```
[SwitchA-GigabitEthernet1/0/1] display ip check source
Total entries found: 1
 MAC               IP              Vlan   Port                        Status
 0001-0203-0406    192.168.0.1     1      GigabitEthernet 1/0/1        DHCP-SNP
```

Display the dynamic entries of DHCP snooping and check it is identical with the dynamic entries that port GigabitEthernet 1/0/1 has obtained.

```
[SwitchA-GigabitEthernet1/0/1] display dhcp-snooping
DHCP Snooping is enabled.
The client binding table for all untrusted ports.
Type : D--Dynamic , S--Static
Type IP Address      MAC Address    Lease        VLAN Interface
==== =============== ============== ============ ==== =================
D    192.168.0.1     0001-0203-0406 86335        1    GigabitEthernet1/0/1
```

As you see, port GigabitEthernet 1/0/1 has obtained the dynamic entries generated by DHCP snooping after it is configured with dynamic binding function.

Troubleshooting IP Source Guard

Failed to Configure Static Binding Entries and Dynamic Binding Function

Symptom

Configuring static binding entries and dynamic binding function fails on a port.

Analysis

IP Source Guard is not supported on the port which has joined an aggregation group. Neither static binding entries nor dynamic binding function can be configured on the port which has joined an aggregation group.

Solution

Remove the port from the aggregation group.

Table of Contents

1 SSH2.0 Configuration

When configuring SSH2.0, go to these sections for information you are interested in:

SSH2.0 Overview

Introduction to SSH2.0

Secure Shell (SSH) offers an approach to securely logging into a remote device. By encryption and strong authentication, it protects devices against attacks such as IP spoofing and plain text password interception.

The device can not only work as an SSH server to support connections with SSH clients, but also work as an SSH client to allow users to establish SSH connections with a remote device acting as the SSH server.

Note

Currently, when acting as an SSH server, the device supports two SSH versions: SSH2.0 and SSH1. When acting as an SSH client, the device supports SSH2.0 only.

Operation of SSH

The session establishment and interaction between an SSH client and the SSH server involves the following five stages:

Table 1-1 Stages in session establishment and interaction between an SSH client and the server

Stages	Description
Version negotiation	SSH1 and SSH2.0 are supported. The two parties negotiate a version to use.
Key and algorithm negotiation	SSH supports multiple algorithms. The two parties negotiate an algorithm for communication.
Authentication	The SSH server authenticates the client in response to the client's authentication request.

Stages	Description
Session request	After passing authentication, the client sends a session request to the server.
Interaction	After the server grants the request, the client and server start to communicate with each other.

Version negotiation

1) The server opens port 22 to listen to connection requests from clients.
2) The client sends a TCP connection request to the server. After the TCP connection is established, the server sends the first packet to the client, which includes a version identification string in the format of “SSH-<primary protocol version number>.<secondary protocol version number>-<software version number>”. The primary and secondary protocol version numbers constitute the protocol version number, while the software version number is used for debugging.
3) The client receives and resolves the packet. If the protocol version of the server is lower but supportable, the client uses the protocol version of the server; otherwise, the client uses its own protocol version.
4) The client sends to the server a packet that contains the number of the protocol version it decides to use. The server compares the version carried in the packet with that of its own. If the server supports the version, the server and client will use the version. Otherwise, the negotiation fails.
5) If the negotiation is successful, the server and the client proceed with key and algorithm negotiation; otherwise, the server breaks the TCP connection.

Note

All the packets involved in the above steps are transferred in plain text.

Key and algorithm negotiation

- The server and the client send key algorithm negotiation packets to each other, which include the supported public key algorithm list, encryption algorithm list, Message Authentication Code (MAC) algorithm list, and compression algorithm list.
- Based on the received algorithm negotiation packets, the server and the client figure out the algorithms to be used. If the negotiation of any type of algorithm fails, the algorithm negotiation fails and the server tears down the connection with the client.
- The server and the client use the DH key exchange algorithm and parameters such as the host key pair to generate the session key and session ID and the client authenticates the identity of the server.

Through the above steps, the server and client get the same session key and session ID. The session key will be used to encrypt and decrypt data exchanged between the server and client later, and the session ID will be used to identify the session established between the server and client and will be used in the authentication stage.

Caution

Before the negotiation, the server must have already generated a DSA or RSA key pair, which is not only used for generating the session key, but also used by the client to authenticate the identity of the server. For details about DSA and RSA key pairs, refer to *Public Key Configuration* in the *Security Volume*.

Authentication

SSH provides two authentication methods: password authentication and publickey authentication.

- Password authentication: The server uses AAA for authentication of the client. During password authentication, the client encrypts its username and password, encapsulates them into a password authentication request, and sends the request to the server. Upon receiving the request, the server decrypts the username and password, checks the validity of the username and password locally or by a remote AAA server, and then informs the client of the authentication result**.**
- Publickey authentication: The server authenticates the client by the digital signature. During publickey authentication, the client sends to the server a publickey authentication request that contains its username, public key, and publickey algorithm information. The server checks whether the public key is valid. If the public key is invalid, the authentication fails; otherwise, the server authenticates the client by the digital signature. Finally, the server sends a message to the client to inform the success or failure of the authentication. Currently, the device supports two publickey algorithms for digital signature: RSA and DSA.

The following gives the steps of the authentication stage:

1) The client sends to the server an authentication request, which includes the username, authentication method (password authentication or publickey authentication), and information related to the authentication method (for example, the password in the case of password authentication).
2) The server authenticates the client. If the authentication fails, the server informs the client by sending a message, which includes a list of available methods for re-authentication.
3) The client selects a method from the list to initiate another authentication.
4) The above process repeats until the authentication succeeds or the failed authentication times exceed the maximum of authentication attempts and the session is torn down.

Note

Besides password authentication and publickey authentication, SSH2.0 provides another two authentication methods:

- **password-publickey**: Performs both password authentication and publickey authentication if the client is using SSH2.0 and performs either if the client is running SSH1.
- **any**: Performs either password authentication or publickey authentication.

Session request

After passing authentication, the client sends a session request to the server, while the server listens to and processes the request from the client. After successfully processing the request, the server sends back to the client an SSH_SMSG_SUCCESS packet and goes on to the interactive session stage with the client. Otherwise, the server sends back to the client an SSH_SMSG_FAILURE packet, indicating that the processing fails or it cannot resolve the request.

Interaction

In this stage, the server and the client exchanges data in the following way:

- The client encrypts and sends the command to be executed to the server.
- The server decrypts and executes the command, and then encrypts and sends the result to the client.
- The client decrypts and displays the result on the terminal.

Note

- In the interaction stage, you can execute commands from the client by pasting the commands in text format (the text must be within 2000 bytes). It is recommended that the commands are in the same view; otherwise, the server may not be able to perform the commands correctly.
- If the command text exceeds 2000 bytes, you can execute the commands by saving the text as a configuration file, uploading the configuration file to the server through SFTP, and then using the configuration file to restart the server.

Configuring the Device as an SSH Server

SSH Server Configuration Task List

Complete the following tasks to configure an SSH server:

Task	Remarks
Generating a DSA or RSA Key Pair	Required
Enabling SSH Server	Required
Configuring the User Interfaces for SSH Clients	Required
Configuring a Client Public Key	Required for publickey authentication users and optional for password authentication users
Configuring an SSH User	Optional
Setting the SSH Management Parameters	Optional

Generating a DSA or RSA Key Pair

The DSA or RSA key pair will be used to generate the session ID in the key and algorithm negotiation stage and used by the client to authenticate the server.

Follow these steps to generate a DSA or RSA key pair on the SSH server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Generate the local DSA or RSA key pair	**public-key local create** { **dsa** | **rsa** }	Required By default, there is neither DSA key pair nor RSA key pair.

Note

- For details about the **public-key local create** command, refer to *Public Key Commands* in the *Security Volume*.
- To ensure that all SSH clients can log into the SSH server successfully, you are recommended to generate both DSA and RSA key pairs on the SSH server. This is because different SSH clients may use different publickey algorithms, though a single client usually uses only one type of publickey algorithm.
- The **public-key local create rsa** command generates two RSA key pairs: a server key pair and a host key pair. Each of the key pairs consists of a public key and a private key. The public key in the server key pair of the SSH server is used in SSH1 to encrypt the session key for secure transmission of the key. As SSH2 uses the DH algorithm to generate the session key on the SSH server and client respectively, no session key transmission is required in SSH2 and the server key pair is not used.
- The length of the modulus of RSA server keys and host keys must be in the range 512 to 2048 bits. Some SSH2 clients require that the length of the key modulus be at least 768 bits on the SSH server side.
- The **public-key local create dsa** command generates only the host key pair. SSH1 does not support the DSA algorithm.
- The length of the modulus of DSA host keys must be in the range 512 to 2048 bits. Some SSH2 clients require that the length of the key modulus be at least 768 bits on the SSH server side.

Enabling SSH Server

Follow these steps to enable SSH server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the SSH server function	**ssh server enable**	Required Disabled by default

Configuring the User Interfaces for SSH Clients

An SSH client accesses the device through a VTY user interface. Therefore, you need to configure the user interfaces for SSH clients to allow SSH login. Note that the configuration takes effect only for clients logging in after the configuration.

Follow these steps to configure the protocols for the current user interface to support:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter user interface view of one or more user interfaces	**user-interface vty** *number* [*ending-number*]	—
Set the login authentication mode to **scheme**	**authentication-mode scheme** [**command-authorization**]	Required By default, the authentication mode is **password**.
Configure the user interface(s) to support SSH login	**protocol inbound** { **all** | **ssh** }	Optional All protocols are supported by default.

Caution

- For detailed information about the **authentication-mode** and **protocol inbound** commands, refer to *User Interface Commands* of the *System Volume.*
- If you configure a user interface to support SSH, be sure to configure the corresponding authentication method with the **authentication-mode scheme** command.
- For a user interface configured to support SSH, you cannot change the authentication mode. To change the authentication mode, undo the SSH support configuration first.

Configuring a Client Public Key

Note

This configuration task is only necessary for SSH users using publickey authentication.

For each SSH user that uses publickey authentication to login, you must configure the client's DSA or RSA host public key on the server, and configure the client to use the corresponding private key.

To configure the public key of an SSH client, you can:

- Configure it manually: You can input or copy the public key to the local host. The copied public key must have not been converted and be in the distinguished encoding rules (DER) encoding format.
- Import it from the public key file: During the import process, the system will automatically convert the public key to a string coded using the Public Key Cryptography Standards (PKCS). Before importing the public key, you must upload the public key file (in binary) to the local host through FTP or TFTP.

Caution

- You are recommended to configure a client public key by importing it from a public key file.
- You can configure at most 20 client pubic keys on an SSH server.

Configuring a client public key manually

Follow these steps to configure the client public key manually:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter public key view	**public-key peer** *keyname*	—
Enter public key code view	**public-key-code begin**	—
Configure a client public key	Enter the content of the public key	Required Spaces and carriage returns are allowed between characters.
Return from public key code view to public key view	**public-key-code end**	— When you exit public key code view, the system automatically saves the public key.
Return from public key view to system view	**peer-public-key end**	—

Importing a client public key from a public key file

Follow these steps to import a public key from a public key file:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Import the public key from a public key file	**public-key peer** *keyname* **import sshkey** *filename*	Required

Note

For information about client side public key configuration and the relevant commands, refer to *Public Key Configuration* in the *Security Volume*.

Configuring an SSH User

This configuration allows you to create an SSH user and specify the service type and authentication mode.

Follow these steps to configure an SSH user and specify the service type and authentication mode:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Create an SSH user, and specify the service type and authentication mode	For Stelnet users	**ssh user** *username* **service-type stelnet authentication-type** { **password** | { **any** | **password-publickey** | **publickey** } **assign publickey** *keyname* }	Required Use either command.
	For all users or SFTP users	**ssh user** *username* **service-type** { **all** | **sftp** } **authentication-type** { **password** | { **any** | **password-publickey** | **publickey** } **assign publickey** *keyname* **work-directory** *directory-name* }	

Caution

- A user without an SSH account can still pass password authentication and log into the server through Stelnet or SFTP, as long as the user can pass AAA authentication and the service type is SSH.
- An SSH server supports up to 1024 SSH users.
- The service type of an SSH user can be Stelnet (Secure Telnet) or SFTP (Secure FTP). For information about Stelnet, refer to SSH2.0 Overview. For information about SFTP, refer to SFTP Overview.
- For successful login through SFTP, you must set the user service type to **sftp** or **all**.
- As SSH1 does not support service type **sftp**, if the client uses SSH1 to log into the server, you must set the service type to **stelnet** or **all** on the server. Otherwise, the client will fail to log in.
- The working folder of an SFTP user is subject to the user authentication method. For a user using only password authentication, the working folder is the AAA authorized one. For a user using only publickey authentication or using both the publickey and password authentication methods, the working folder is the one set by using the **ssh user** command.
- The configured authentication method takes effect only for users logging in after the configuration.

Note

For users using publickey authentication:

- You must configure on the device the corresponding username and public keys.
- After login, the commands available for a user are determined by the user privilege level, which is configured with the **user privilege level** command on the user interface.

For users using password authentication:

- You can configure the accounting information either on the device or on the remote authentication server (such as RADIUS authentication server).
- After login, the commands available to a user are determined by AAA authorization.

Setting the SSH Management Parameters

SSH management includes:

- Enabling the SSH server to be compatible with SSH1 client
- Setting the server key pair update interval, applicable to users using SSH1 client
- Setting the SSH user authentication timeout period
- Setting the maximum number of SSH authentication attempts

Setting the above parameters can help avoid malicious guess at and cracking of the keys and usernames, securing your SSH connections.

Follow these steps to set the SSH management parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the SSH server to work with SSH1 clients	**ssh server compatible-ssh1x enable**	Optional By default, the SSH server can work with SSH1 clients.
Set the RSA server key pair update interval	**ssh server rekey-interval** *hours*	Optional 0 by default, that is, the RSA server key pair is not updated.
Set the SSH user authentication timeout period	**ssh server authentication-timeout** *time-out-value*	Optional 60 seconds by default
Set the maximum number of SSH authentication attempts	**ssh server authentication-retries** *times*	Optional 3 by default

Note

Authentication will fail if the number of authentication attempts (including both publickey and password authentication) exceeds that specified in the **ssh server authentication-retries** command.

Configuring the Device as an SSH Client

SSH Client Configuration Task List

Complete the following tasks to configure an SSH client:

Task	Remarks
Specifying a Source IP address/Interface for the SSH client	Optional
Configuring Whether First-time Authentication is Supported	Optional
Establishing a Connection Between the SSH Client and the Server	Required

Specifying a Source IP address/Interface for the SSH client

This configuration task allows you to specify a source IP address or interface for the client to access the SSH server, improving service manageability.

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Specify a source IP address or interface for the SSH client	Specify a source IPv4 address or interface for the SSH client	**ssh client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	Required By default, the address of the interface decided by the routing is used to access the SSH server
	Specify a source IPv6 address or interface for the SSH client	**ssh client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	

Configuring Whether First-time Authentication is Supported

When the device connects to the SSH server as an SSH client, you can configure whether the device supports first-time authentication.

- With first-time authentication, when an SSH client not configured with the server host public key accesses the server for the first time, the user can continue accessing the server, and save the host public key on the client. When accessing the server again, the client will use the saved server host public key to authenticate the server.
- Without first-time authentication, a client not configured with the server host public key will deny to access the server. To access the server, a user must configure in advance the server host public key locally and specify the public key name for authentication.

Enable the device to support first-time authentication

Follow these steps to enable the device to support first-time authentication:

To do...	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the device to support first-time authentication	**ssh client first-time enable**	Optional By default, first-time authentication is supported on a client.

Disable first-time authentication

For successful authentication of an SSH client not supporting first-time authentication, the server host public key must be configured on the client and the public key name must be specified.

Follow these steps to disable first-time authentication:

To do...	Use the command…	Remarks
Enter system view	**system-view**	—
Disable first-time authentication support	**undo ssh client first-time**	Optional By default, first-time authentication is supported on a client.

To do...	Use the command…	Remarks
Configure the server public key	Refer to Configuring a Client Public Key	Required The method of configuring server public key on the client is similar to that of configuring client public key on the server.
Specify the host public key name of the server	**ssh client authentication server** *server* **assign publickey** *keyname*	Required

Establishing a Connection Between the SSH Client and the Server

Follow these steps to establish the connection between the SSH client and the server:

To do...		Use the command…	Remarks
Establish a connection between the SSH client and server, and specify the public key algorithm, preferred encryption algorithms, preferred HMAC algorithms and preferred key exchange algorithm	For an IPv4 server	**ssh2** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Use either command in user view.
	For an IPv4 IPv6 server	**ssh2 ipv6** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	

Displaying and Maintaining SSH

To do…	Use the command…	Remarks
Display the source IP address or interface currently set for the SFTP client	**display sftp client source**	Available in any view
Display the source IP address or interface currently set for the SSH client	**display ssh client source**	Available in any view
Display SSH server status information or session information on an SSH server	**display ssh server** { **status** | **session** }	Available in any view
Display the mappings between SSH servers and their host public keys saved on an SSH client	**display ssh server-info**	Available in any view
Display information about a specified or all SSH users on the SSH server	**display ssh user-information** [*username*]	Available in any view

To do…	Use the command…	Remarks
Display the public keys of the local key pairs	**display public-key local** { **dsa** | **rsa** } **public**	Available in any view
Display the public keys of the SSH peers	**display public-key peer** [**brief** | **name** *publickey-name*]	Available in any view

Note

For information about the **display public-key local** and **display public-key peer** commands, refer to *Public Key Commands* in the *Security Volume*.

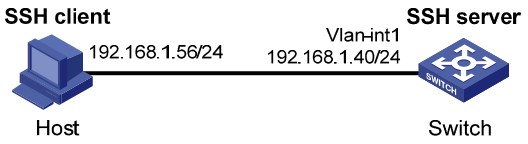
SSH Server Configuration Examples

When Switch Acts as Server for Password Authentication

Network requirements

- As shown in Figure 1-1, a local SSH connection is established between the host (the SSH client) and the switch (the SSH server) for secure data exchange.
- Password authentication is required. The username and password are saved on the switch.

Figure 1-1 Switch acts as server for password authentication



Configuration procedure

1) Configure the SSH server

Generate RSA and DSA key pairs and enable the SSH server.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
[Switch] ssh server enable
```

Configure an IP address for VLAN interface 1. This address will serve as the destination of the SSH connection.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface1] quit
```

Set the authentication mode for the user interfaces to AAA.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

Enable the user interfaces to support SSH.

```
[Switch-ui-vty0-4] protocol inbound ssh
[Switch-ui-vty0-4] quit
```

Create local user **client001**, and set the user command privilege level to 3

```
[Switch] local-user client001
[Switch-luser-client001] password simple aabbcc
[Switch-luser-client001] service-type ssh
[Router-luser-client001] authorization-attribute level 3
[Switch-luser-client001] quit
```

Specify the service type for user **client001** as **Stelnet**, and the authentication mode as password. This step is optional.

```
[Switch] ssh user client001 service-type stelnet authentication-type password
```

2) Configure the SSH client

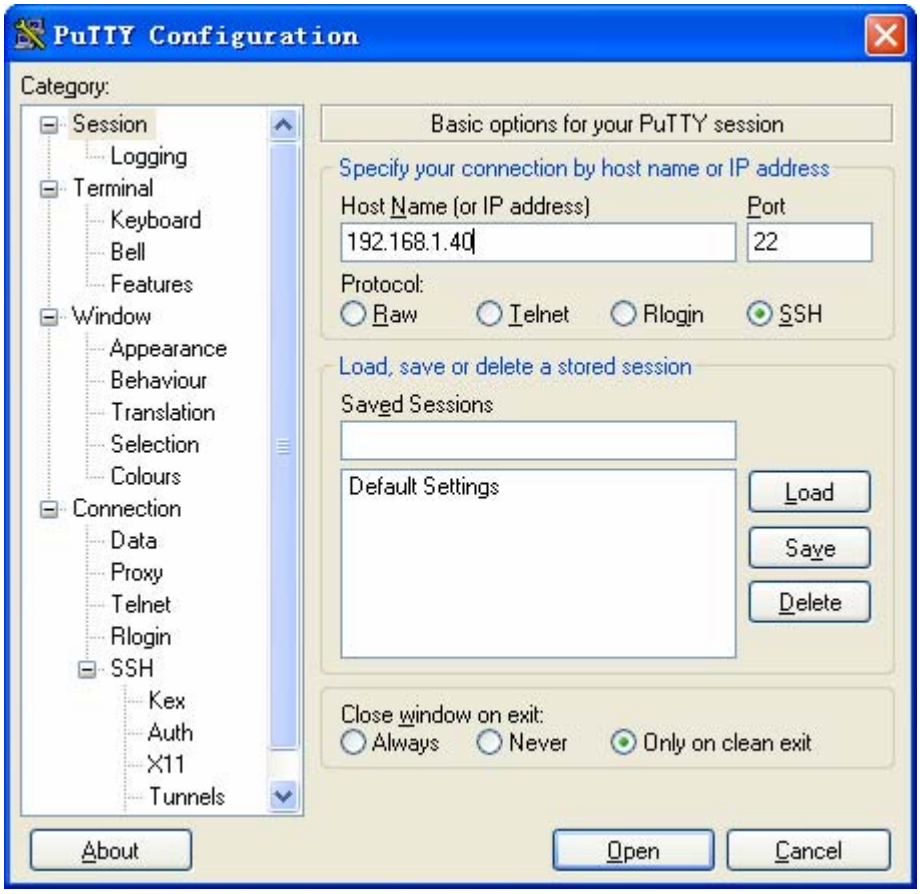
Note

There are many kinds of SSH client software, such as PuTTY, and OpenSSH. The following is an example of configuring SSH client using Putty Version 0.58.

Establish a connection with the SSH server

Launch PuTTY.exe to enter the following interface. In the **Host Name (or IP address)** text box, enter the IP address of the server (192.168.1.40).

Figure 1-2 SSH client configuration interface



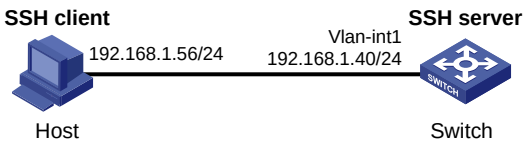
In the window shown in Figure 1-2, click **Open**. If the connection is normal, you will be prompted to enter the username and password. After entering the correct username (**client001**) and password (**aabbcc**), you can enter the configuration interface.

When Switch Acts as Server for Publickey Authentication

Network requirements

- As shown in Figure 1-3, a local SSH connection is established between the host (the SSH client) and the switch (the SSH server) for secure data exchange.
- Publickey authentication is used, the algorithm is RSA.

Figure 1-3 Switch acts as server for publickey authentication



Configuration procedure

1) Configure the SSH server

Generate RSA and DSA key pairs and enable SSH server.

```
<Switch> system-view
[Switch] public-key local create rsa
```

```
[Switch] public-key local create dsa
[Switch] ssh server enable
```

Configure an IP address for VLAN interface 1. This address will serve as the destination of the SSH connection.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface1] quit
```

Set the authentication mode for the user interfaces to AAA.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

Enable the user interfaces to support SSH.

```
[Switch-ui-vty0-4] protocol inbound ssh
```

Set the user command privilege level to 3.

```
[Switch-ui-vty0-4] user privilege level 3
[Switch-ui-vty0-4] quit
```

Note

Before performing the following tasks, you must use the client software to generate an RSA key pair on the client, save the public key in a file named **key.pub**, and then upload the file to the SSH server through FTP or TFTP. For details, refer to Configure the SSH client below.

Import the client's public key from file **key.pub** and name it **Switch001**.

```
[Switch] public-key peer Switch001 import sshkey key.pub
```

Specify the authentication type for user **client002** as publickey, and assign the public key **Switch001** to the user.

```
[Switch] ssh user client002 service-type stelnet authentication-type publickey assign publickey Switch001
```

2) Configure the SSH client

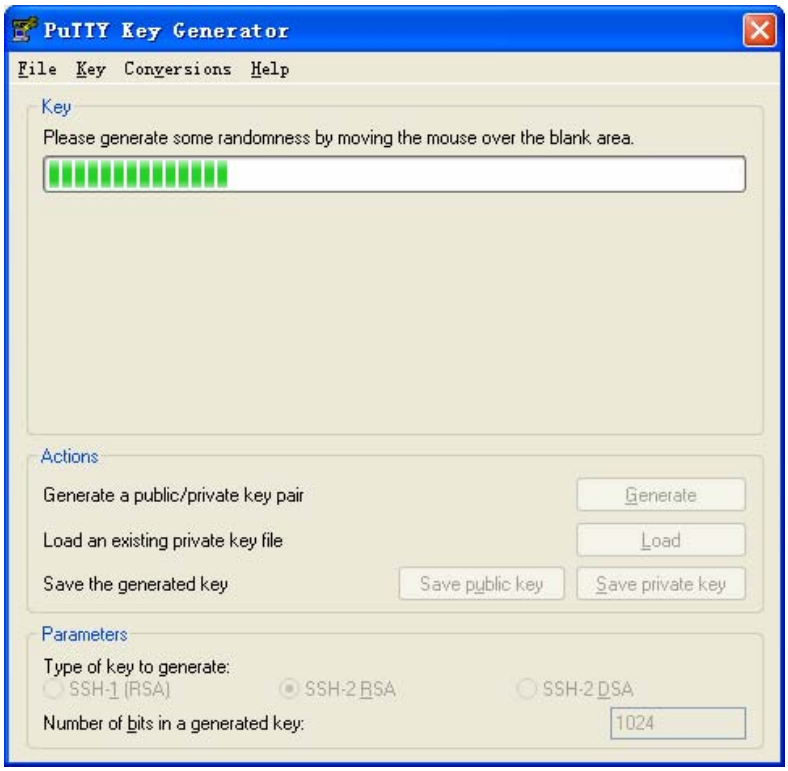
Generate an RSA key pair.

Run PuTTYGen.exe, select **SSH-2 RSA** and click **Generate**.

Figure 1-4 Generate a client key pair 1)

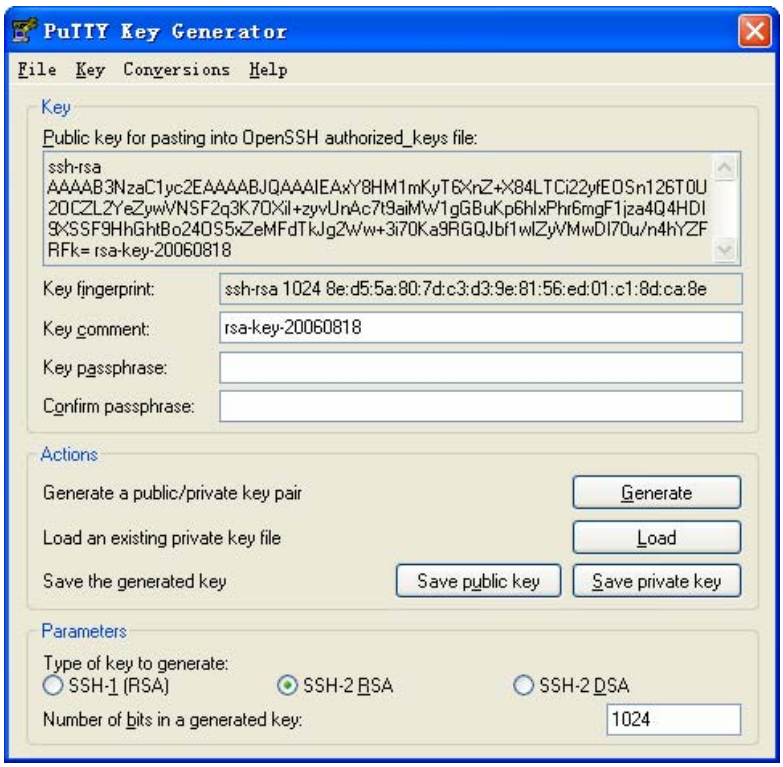
While generating the key pair, you must move the mouse continuously and keep the mouse off the green process bar shown in Figure 1-5. Otherwise, the process bar stops moving and the key pair generating process will be stopped.

Figure 1-5 Generate a client key pair 2)



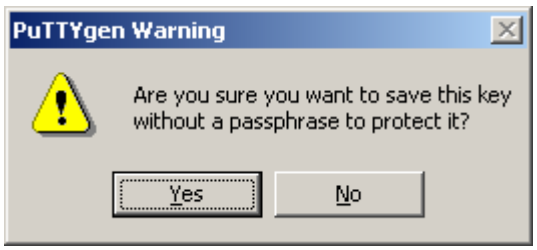
After the key pair is generated, click **Save public key** and specify the file name as **key.pub** to save the public key.

Figure 1-6 Generate a client key pair 3)



Likewise, to save the private key, click **Save private key**. A warning window pops up to prompt you whether to save the private key without any protection. Click **Yes** and enter the name of the file for saving the key (**private** in this case).

Figure 1-7 Generate a client key pair 4)



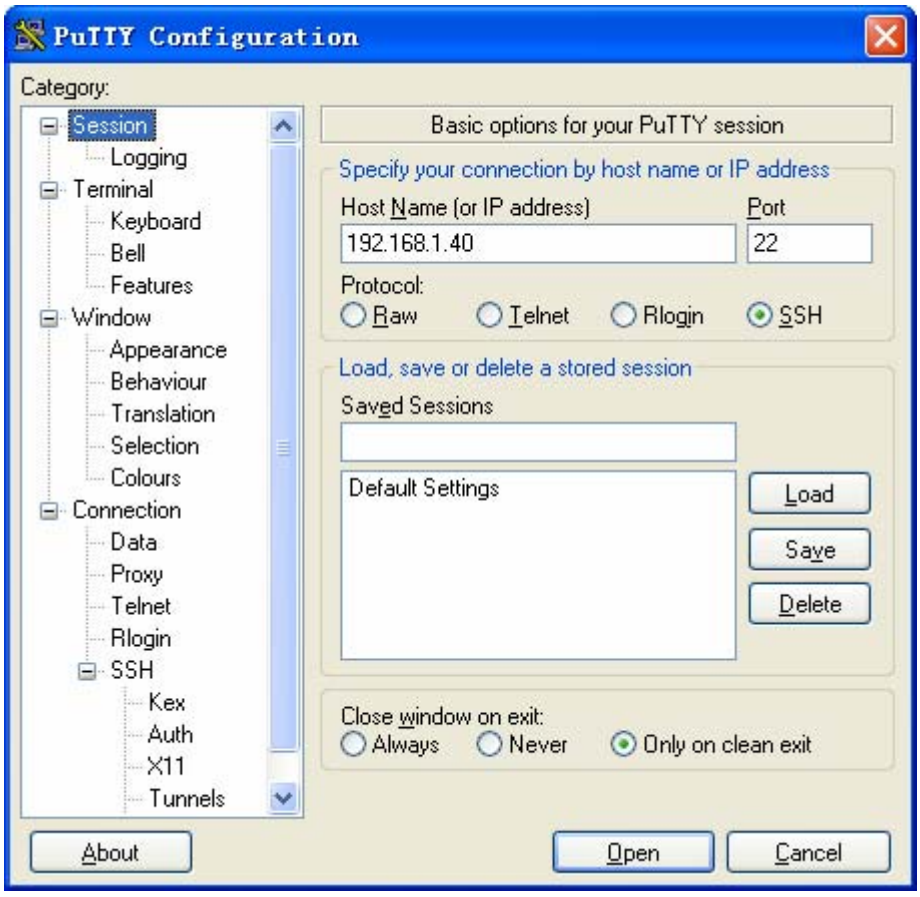
Note

After generating a key pair on a client, you need to transmit the saved public key file to the server through FTP or TFTP and have the configuration on the server done before continuing configuration of the client.

Specify the private key file and establish a connection with the SSH server

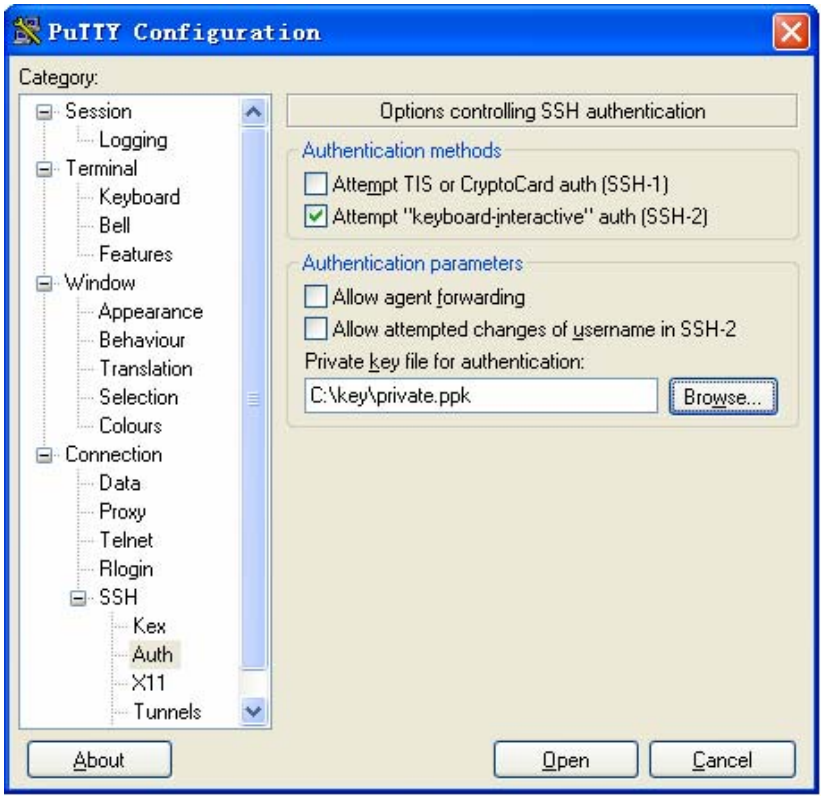
Launch PuTTY.exe to enter the following interface. In the **Host Name (or IP address)** text box, enter the IP address of the server (192.168.1.40).

Figure 1-8 SSH client configuration interface 1)



Select **Connection/SSH/Auth** from the navigation tree. The following window appears. Click **Browse…** to bring up the file selection window, navigate to the private key file and click **OK**.

Figure 1-9 SSH client configuration interface 2)



In the window shown in Figure 1-9, click **Open**. If the connection is normal, you will be prompted to enter the username. After entering the correct username (**client002**), you can enter the configuration interface.

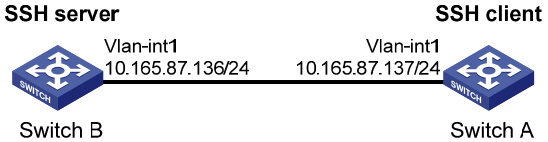
SSH Client Configuration Examples

When Switch Acts as Client for Password Authentication

Network requirements

- As shown in Figure 1-10, Switch A (the SSH client) needs to log into Switch B (the SSH server) through the SSH protocol.
- The username of the SSH client is **client001** and the password is **aabbcc**. Password authentication is required.

Figure 1-10 Switch acts as client for password authentication



Configuration procedure

1) Configure the SSH server

Create RSA and DSA key pairs and enable the SSH server.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable
```

Create an IP address for VLAN interface 1, which the SSH client will use as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 10.165.87.136 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Set the authentication mode for the user interfaces to AAA.

```
[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
```

Enable the user interfaces to support SSH.

```
[SwitchB-ui-vty0-4] protocol inbound ssh
[SwitchB-ui-vty0-4] quit
```

Create local user **client001**.

```
[SwitchB] local-user client001
[SwitchB-luser-client001] password simple aabbcc
[SwitchB-luser-client001] service-type ssh
[SwitchB-luser-client001] authorization-attribute level 3
[SwitchB-luser-client001] quit
```

Specify the service type for user **client001** as **Stelnet**, and the authentication type as **password**. This step is optional.

```
[SwitchB] ssh user client001 service-type stelnet authentication-type password
```

2) Configure the SSH client

Configure an IP address for VLAN interface 1.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.165.87.137 255.255.255.0
[SwitchA-Vlan-interface1] quit
[SwitchA] quit
```

- If the client support first-time authentication, you can directly establish a connection from the client to the server.

Establish an SSH connection to server 10.165.87.136.

```
<SwitchA> ssh2 10.165.87.136
Username: client001
Trying 10.165.87.136 ...
Press CTRL+K to abort
Connected to 10.165.87.136 ...

The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
Enter password:
```

After you enter the correct username, you can log into Switch B successfully.

- If the client does not support first-time authentication, you need to perform the following configurations.

Disable first-time authentication.

```
[SwitchA] undo ssh client first-time
```

Configure the host public key of the SSH server. You can get the server host public key by using the **display public-key local dsa public** command on the server.

```
[SwitchA] public-key peer key1
[SwitchA-pkey-public-key] public-key-code begin
[SwitchA-pkey-key-code]308201B73082012C06072A8648CE3804013082011F0281810
0D757262C4584C44C211F18BD96E5F0
[SwitchA-pkey-key-code]61C4F0A423F7FE6B6B85B34CEF72CE14A0D3A5222FE08CECE
65BE6C265854889DC1EDBD13EC8B274
[SwitchA-pkey-key-code]DA9F75BA26CCB987723602787E922BA84421F22C3C89CB9B0
6FD60FE01941DDD77FE6B12893DA76E
[SwitchA-pkey-key-code]EBC1D128D97F0678D7722B5341C8506F358214B16A2FAC4B3
68950387811C7DA33021500C773218C
[SwitchA-pkey-key-code]737EC8EE993B4F2DED30F48EDACE915F0281810082269009E
14EC474BAF2932E69D3B1F18517AD95
[SwitchA-pkey-key-code]94184CCDFCEAE96EC4D5EF93133E84B47093C52B20CD35D02
492B3959EC6499625BC4FA5082E22C5
[SwitchA-pkey-key-code]B374E16DD00132CE71B020217091AC717B612391C76C1FB2E
88317C1BD8171D41ECB83E210C03CC9
[SwitchA-pkey-key-code]B32E810561C21621C73D6DAAC028F4B1585DA7F42519718CC
9B09EEF0381840002818000AF995917
[SwitchA-pkey-key-code]E1E570A3F6B1C2411948B3B4FFA256699B3BF871221CC9C5D
F257523777D033BEE77FC378145F2AD
[SwitchA-pkey-key-code]D716D7DB9FCABB4ADBF6FB4FDB0CA25C761B308EF53009F71
01F7C62621216D5A572C379A32AC290
[SwitchA-pkey-key-code]E55B394A217DA38B65B77F0185C8DB8095522D1EF044B465E
8716261214A5A3B493E866991113B2D
[SwitchA-pkey-key-code]485348
[SwitchA-pkey-key-code] public-key-code end
[SwitchA-pkey-public-key] peer-public-key end
```

Specify the host public key for the SSH server (10.165.87.136) as **key1**.

```
[SwitchA] ssh client authentication server 10.165.87.136 assign publickey key1
[SwitchA] quit
```

Establish an SSH connection to server 10.165.87.136.

```
<SwitchA> ssh2 10.165.87.136
Username: client001
Trying 10.165.87.136
Press CTRL+K to abort
Connected to 10.165.87.136...
Enter password:
```

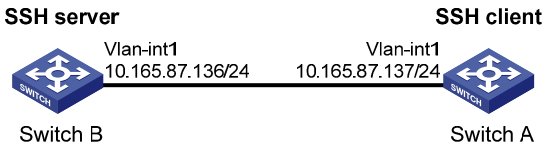
After you enter the correct username and password, you can log into Switch B successfully.

When Switch Acts as Client for Publickey Authentication

Network requirements

- As shown in Figure 1-11, Switch A (the SSH client) needs to log into Switch B (the SSH server) through the SSH protocol.
- Publickey authentication is used, and the public key algorithm is DSA.

Figure 1-11 Switch acts as client for publickey authentication



Configuration procedure

1) Configure the SSH server

Generate RSA and DSA key pairs and enable SSH server.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable
```

Configure an IP address for VLAN interface 1, which the SSH client will use as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 10.165.87.136 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Set the authentication mode for the user interfaces to AAA.

```
[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
```

Enable the user interfaces to support SSH.

```
[SwitchB-ui-vty0-4] protocol inbound ssh
```

Set the user command privilege level to 3.

```
[SwitchB-ui-vty0-4] user privilege level 3
[SwitchB-ui-vty0-4] quit
```

Note

Before performing the following tasks, you must use the client software to generate an RSA key pair on the client, save the public key in a file named **key.pub**, and then upload the file to the SSH server through FTP or TFTP. For details, refer to Configure the SSH client below.

Import the peer public key from the file **key.pub**.

```
[SwitchB] public-key peer Switch001 import sshkey key.pub
```

Specify the authentication type for user **client002** as publickey, and assign the public key **Switch001** to the user.

```
[SwitchB] ssh user client002 service-type stelnet authentication-type publickey assign publickey Switch001
```

2) Configure the SSH client

Configure an IP address for Vlan interface 1.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.165.87.137 255.255.255.0
[SwitchA-Vlan-interface1] quit
```

Generate a DSA key pair.

```
[SwitchA] public-key local create dsa
```

Export the DSA public key to the file **key.pub**.

```
[SwitchA] public-key local export dsa ssh2 key.pub
[SwitchA] quit
```

Note

After generating a key pair on a client, you need to transmit the saved public key file to the server through FTP or TFTP and have the configuration on the server done before continuing configuration of the client.

Establish an SSH connection to the server (10.165.87.136).

```
<SwitchA> ssh2 10.165.87.136
Username: client002
Trying 10.165.87.136 ...
Press CTRL+K to abort
Connected to 10.165.87.136 ...

The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
```

Later, you will find that you have logged into Switch B successfully.

2 SFTP Service

When configuring SFTP, go to these sections for information you are interested in:

SFTP Overview

The secure file transfer protocol (SFTP) is a new feature in SSH2.0.

SFTP uses the SSH connection to provide secure data transfer. The device can serve as the SFTP server, allowing a remote user to log into the SFTP server for secure file management and transfer. The device can also server as an SFTP client, enabling a user to login from the device to a remote device for secure file transfer.

Configuring an SFTP Server

Configuration Prerequisites

- You have configured the SSH server. For the detailed configuration procedure, refer to.Configuring the Device as an SSH Server
- You have used the ssh user service-type command to set the service type of SSH users to sftp or all. For configuration procedure, refer to Configuring an SSH User.

Enabling the SFTP Server

This configuration task is to enable the SFTP service so that a client can log into the SFTP server through SFTP.

Follow these steps to enable the SFTP server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the SFTP server	**sftp server enable**	Required Disabled by default

Note

When the device functions as the SFTP server, only one client can access the SFTP server at a time. If the SFTP client uses WinSCP, a file on the server cannot be modified directly; it can only be downloaded to a local place, modified, and then uploaded to the server.

Configuring the SFTP Connection Idle Timeout Period

Once the idle period of an SFTP connection exceeds the specified threshold, the system automatically tears the connection down, so that a user cannot occupy a connection for nothing.

Follow these steps to configure the SFTP connection idle timeout period:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the SFTP connection idle timeout period	**sftp server idle-timeout** *time-out-value*	Optional 10 minutes by default

Configuring an SFTP Client

Specifying a Source IP Address or Interface for the SFTP Client

You can configure a client to use only a specified source IP address or interface to access the SFTP server, thus enhancing the service manageability.

Follow these steps to specify a source IP address or interface for the SFTP client:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Specify a source IP address or interface for the SFTP client	Specify a source IPv4 address or interface for the SFTP client	**sftp client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	Required Use either command. By default, an SFTP client uses the interface address specified by the route of the device to access the SFTP server.
	Specify a source IPv6 address or interface for the SFTP client	**sftp client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	

Establishing a Connection to the SFTP Server

This configuration task is to enable the SFTP client to establish a connection with the remote SFTP server and enter SFTP client view.

Follow these steps to enable the SFTP client:

To do…		Use the command…	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	Establish a connection to the remote IPv4 SFTP server and enter SFTP client view	**sftp** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Use either command in user view.
	Establish a connection to the remote IPv6 SFTP server and enter SFTP client view	**sftp ipv6** *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	

Working with the SFTP Directories

SFTP directory operations include:

- Changing or displaying the current working directory
- Displaying files under a specified directory or the directory information
- Changing the name of a specified directory on the server
- Creating or deleting a directory

Follow these steps to work with the SFTP directories:

To do…	Use the command…	Remarks
Enter SFTP client view	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Execute the command in user view.
Change the working directory of the remote SFTP server	**cd** [*remote-path*]	Optional
Return to the upper-level directory	**cdup**	Optional
Display the current working directory of the remote SFTP server	**pwd**	Optional
Display files under a specified directory	**dir** [**-a** | **-l**] [*remote-path*]	Optional The **dir** command functions as the **ls** command.
	ls [**-a** | **-l**] [*remote-path*]	
Change the name of a specified directory on the SFTP server	**rename** *oldname newname*	Optional

To do…	Use the command…	Remarks
Create a new directory on the remote SFTP server	**mkdir** *remote-path*	Optional
Delete a directory from the SFTP server	**rmdir** *remote-path*&<1-10>	Optional

Working with SFTP Files

SFTP file operations include:

- Changing the name of a file
- Downloading a file
- Uploading a file
- Displaying a list of the files
- Deleting a file

Follow these steps to work with SFTP files:

To do…	Use the command…	Remarks
Enter SFTP client view	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Execute the command in user view.
Change the name of a specified file or directory on the SFTP server	**rename** *old-name new-name*	Optional
Download a file from the remote server and save it locally	**get** *remote-file* [*local-file*]	Optional
Upload a local file to the remote SFTP server	**put** *local-file* [*remote-file*]	Optional
Display the files under a specified directory	**dir** [**-a** | **-l**] [*remote-path*]	Optional The **dir** command functions as the **ls** command.
	ls [**-a** | **-l**] [*remote-path*]	
Delete a file from the SFTP server	**delete** *remote-file*&<1-10>	Optional The **delete** command functions as the **remove** command.
	remove *remote-file*&<1-10>	

Displaying Help Information

This configuration task is to display a list of all commands or the help information of an SFTP client command, such as the command format and parameters.

Follow these steps to display a list of all commands or the help information of an SFTP client command:

To do…	Use the command…	Remarks
Enter SFTP client view	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Execute the command in user view.
Display a list of all commands or the help information of an SFTP client command	**help** [**all** | *command-name*]	Required

Terminating the Connection to the Remote SFTP Server

Follow these steps to terminate the connection to the remote SFTP server:

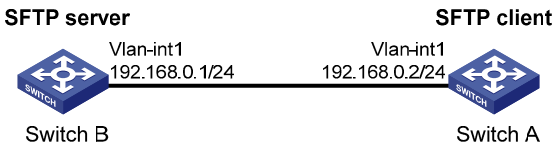
To do…	Use the command…	Remarks
Enter SFTP client view	**sftp** [**ipv6**] *server* [*port-number*] [**identity-key** { **dsa** | **rsa** } | **prefer-ctos-cipher** { **aes128** | **des** } | **prefer-ctos-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** } | **prefer-kex** { **dh-group-exchange** | **dh-group1** | **dh-group14** } | **prefer-stoc-cipher** { **aes128** | **des** } | **prefer-stoc-hmac** { **md5** | **md5-96** | **sha1** | **sha1-96** }] *	Required Execute the command in user view.
Terminate the connection to the remote SFTP server and return to user view	**bye**	Required. Use any of the commands. These three commands function in the same way.
	exit	
	quit	

SFTP Client Configuration Example

Network requirements

As shown in Figure 2-1, an SSH connection is established between Switch A and Switch B. Switch A, an SFTP client, logs in to Switch B for file management and file transfer. An SSH user uses publickey authentication with the public key algorithm being RSA.

Figure 2-1 Network diagram for SFTP client configuration



Configuration procedure

1) Configure the SFTP server (Switch B)

Generate RSA and DSA key pairs and enable the SSH server.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable
```

Enable the SFTP server.

```
[SwitchB] sftp server enable
```

Configure an IP address for VLAN interface 1, which the SSH client uses as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Set the authentication mode on the user interfaces to AAA.

```
[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
```

Set the protocol that a remote user uses to log in as **SSH**.

```
[SwitchB-ui-vty0-4] protocol inbound ssh
[SwitchB-ui-vty0-4] quit
```

 **Note**

Before performing the following tasks, you must generate use the client software to generate RSA key pairs on the client, save the host public key in a file named **pubkey**, and then upload the file to the SSH server through FTP or TFTP. For details, refer to Configure the SFTP client (Switch A) below.

Import the peer public key from the file **pubkey**.

```
[SwitchB] public-key peer Switch001 import sshkey pubkey
```

For user **client001**, set the service type as SFTP, authentication type as publickey, public key as **Switch001**, and working folder as **flash:/**

```
[SwitchB] ssh user client001 service-type sftp authentication-type publickey assign publickey Switch001 work-directory flash:/
```

2) Configure the SFTP client (Switch A)

Configure an IP address for VLAN interface 1.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 192.168.0.2 255.255.255.0
[SwitchA-Vlan-interface1] quit
```

Generate RSA key pairs.

```
[SwitchA] public-key local create rsa
```

Export the host public key to file **pubkey**.

```
[SwitchA] public-key local export rsa ssh2 pubkey
```

```
[SwitchA] quit
```

Note

After generating key pairs on a client, you need to transmit the saved public key file to the server through FTP or TFTP and have the configuration on the server done before continuing configuration of the client.

Establish a connection to the remote SFTP server and enter SFTP client view.

```
<SwitchA> sftp 192.168.0.1 identity-key rsa
Input Username: client001
Trying 192.168.0.1 ...
Press CTRL+K to abort
Connected to 192.168.0.1 ...

The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n

sftp-client>
```

Display files under the current directory of the server, delete the file named **z**, and check if the file has been deleted successfully.

```
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
-rwxrwxrwx   1 noone    nogroup         0 Sep 01 08:00 z
sftp-client> delete z
The following File will be deleted:
/z
Are you sure to delete it? [Y/N]:y
This operation may take a long time.Please wait...

File successfully Removed
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
```

Add a directory named **new1** and check if it has been created successfully.

```
sftp-client> mkdir new1
New directory created
```

```
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:30 new1
```

Rename directory **new1** to **new2** and check if the directory has been renamed successfully.

```
sftp-client> rename new1 new2
File successfully renamed
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
```

Download the file **pubkey2** from the server and change the name to **public**.

```
sftp-client> get pubkey2 public
Remote  file:/pubkey2 --->  Local file: public
Downloading file successfully ended
```

Upload the local file **pu** to the server, save it as **puk**, and check if the file has been uploaded successfully.

```
sftp-client> put pu puk
Local file:pu --->  Remote file: /puk
Uploading file successfully ended
sftp-client> dir
-rwxrwxrwx   1 noone    nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx   1 noone    nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx   1 noone    nogroup       283 Aug 24 07:39 pubkey
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
drwxrwxrwx   1 noone    nogroup         0 Sep 02 06:33 new2
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:35 pub
-rwxrwxrwx   1 noone    nogroup       283 Sep 02 06:36 puk
sftp-client>
```

Terminate the connection to the remote SFTP server.

```
sftp-client> quit
Bye
<SwitchA>
```

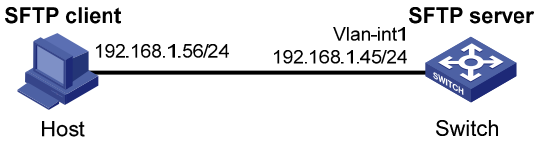
SFTP Server Configuration Example

Network requirements

As shown in Figure 2-2, an SSH connection is established between the host and the switch. The host, an SFTP client, logs into the switch for file management and file transfer. An SSH user uses password

authentication with the username being **client002** and the password being **aabbcc**. The username and password are saved on the switch.

Figure 2-2 Network diagram for SFTP server configuration



Configuration procedure

1) Configure the SFTP server

Generate RSA and DSA key pairs and enable the SSH server.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
[Switch] ssh server enable
```

Enable the SFTP server.

```
[Switch] sftp server enable
```

Configure an IP address for VLAN-interface 1, which the client will use as the destination for SSH connection.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.1.45 255.255.255.0
[Switch-Vlan-interface1] quit
```

Set the authentication mode of the user interfaces to AAA.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

Enable the user interfaces to support SSH.

```
[Switch-ui-vty0-4] protocol inbound ssh
[Switch-ui-vty0-4] quit
```

Configure a local user named **client002** with the password being **aabbcc** and the service type being SSH.

```
[Switch] local-user client002
[Switch-luser-client002] password simple aabbcc
[Switch-luser-client002] service-type ssh
[Switch-luser-client002] quit
```

Configure the user authentication type as password and service type as SFTP.

```
[Switch] ssh user client002 service-type sftp authentication-type password
```

2) Configure the SFTP client

Note

- There are many kinds of SSH client software. The following takes the PSFTP of Putty Version 0.58 as an example.
- The PSFTP supports only password authentication.

Establish a connection with the remote SFTP server.

Run the psftp.exe to launch the client interface as shown in Figure 2-3, and enter the following command:

```
open 192.168.1.45
```

Enter username **client002** and password **aabbcc** as prompted to log into the SFTP server.

Figure 2-3 SFTP client interface

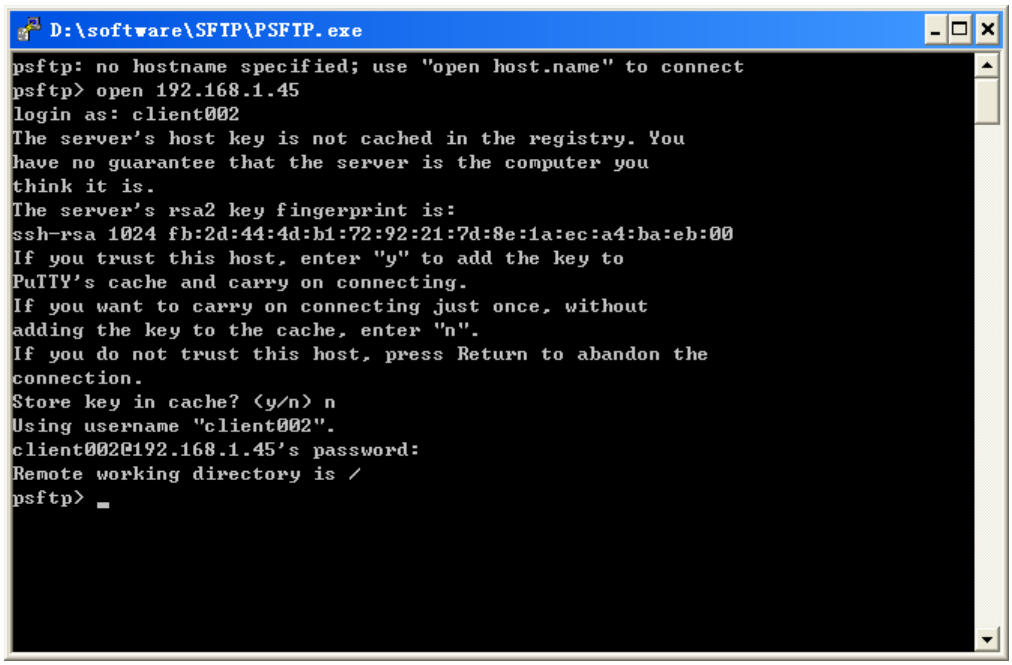


Table of Contents

1 PKI Configuration

When configuring PKI, go to these sections for information you are interested in:

Introduction to PKI

This section covers these topics:

PKI Overview

The Public Key Infrastructure (PKI) is a general security infrastructure for providing information security through public key technologies.

PKI, also called asymmetric key infrastructure, uses a key pair to encrypt and decrypt the data. The key pair consists of a private key and a public key. The private key must be kept secret while the public key needs to be distributed. Data encrypted by one of the two keys can only be decrypted by the other.

A key problem of PKI is how to manage the public keys. Currently, PKI employs the digital certificate mechanism to solve this problem. The digital certificate mechanism binds public keys to their owners, helping distribute public keys in large networks securely.

With digital certificates, the PKI system provides network communication and e-commerce with security services such as user authentication, data non-repudiation, data confidentiality, and data integrity.

PKI Terms

Digital certificate

A digital certificate is a file signed by a certificate authority (CA) for an entity. It includes mainly the identity information of the entity, the public key of the entity, the name and signature of the CA, and the validity period of the certificate, where the signature of the CA ensures the validity and authority of the certificate. A digital certificate must comply with the international standard of ITU-T X.509. Currently, the most common standard is X.509 v3.

This manual involves two types of certificates: local certificate and CA certificate. A local certificate is a digital certificate signed by a CA for an entity, while a CA certificate is the certificate of a CA. If multiple CAs are trusted by different users in a PKI system, the CAs will form a CA tree with the root CA at the top

level. The root CA has a CA certificate signed by itself while each lower level CA has a CA certificate signed by the CA at the next higher level.

CRL

An existing certificate may need to be revoked when, for example, the user name changes, the private key leaks, or the user stops the business. Revoking a certificate is to remove the binding of the public key with the user identity information. In PKI, the revocation is made through certificate revocation lists (CRLs). Whenever a certificate is revoked, the CA publishes one or more CRLs to show all certificates that have been revoked. The CRLs contain the serial numbers of all revoked certificates and provide an effective way for checking the validity of certificates.

A CA may publish multiple CRLs when the number of revoked certificates is so large that publishing them in a single CRL may degrade network performance, and it uses CRL distribution points to indicate the URLs of these CRLs.

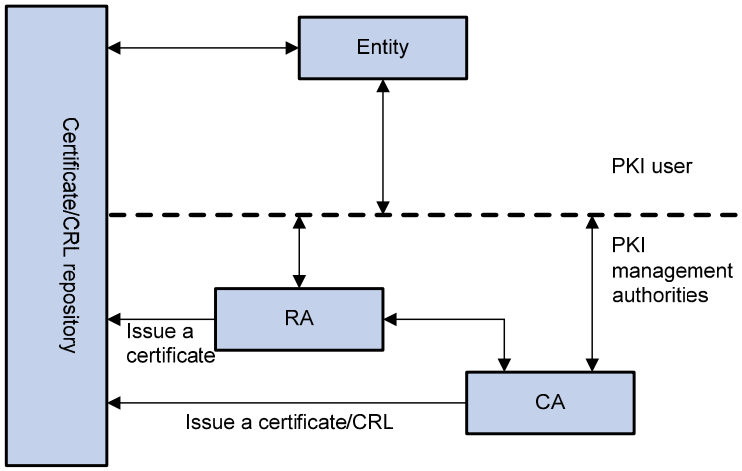
CA policy

A CA policy is a set of criteria that a CA follows in processing certificate requests, issuing and revoking certificates, and publishing CRLs. Usually, a CA advertises its policy in the form of certification practice statement (CPS). A CA policy can be acquired through out-of-band means such as phone, disk, and e-mail. As different CAs may use different methods to check the binding of a public key with an entity, make sure that you understand the CA policy before selecting a trusted CA for certificate request.

Architecture of PKI

A PKI system consists of entities, a CA, a registration authority (RA) and a PKI repository, as shown in Figure 1-1.

Figure 1-1 PKI architecture



Entity

An entity is an end user of PKI products or services, such as a person, an organization, a device like a router or a switch, or a process running on a computer.

CA

A CA is a trusted authority responsible for issuing and managing digital certificates. A CA issues certificates, specifies the validity periods of certificates, and revokes certificates as needed by publishing CRLs.

RA

A registration authority (RA) is an extended part of a CA or an independent authority. An RA can implement functions including identity authentication, CRL management, key pair generation and key pair backup. The PKI standard recommends that an independent RA be used for registration management to achieve higher security of application systems.

PKI repository

A PKI repository can be a Lightweight Directory Access Protocol (LDAP) server or a common database. It stores and manages information like certificate requests, certificates, keys, CRLs and logs while providing a simple query function.

LDAP is a protocol for accessing and managing PKI information. An LDAP server stores user information and digital certificates from the RA server and provides directory navigation service. From an LDAP server, an entity can retrieve local and CA certificates of its own as well as certificates of other entities.

Applications of PKI

The PKI technology can satisfy the security requirements of online transactions. As an infrastructure, PKI has a wide range of applications. Here are some application examples.

VPN

A virtual private network (VPN) is a private data communication network built on the public communication infrastructure. A VPN can leverage network layer security protocols (for instance, IPSec) in conjunction with PKI-based encryption and digital signature technologies for confidentiality.

Secure E-mail

E-mails require confidentiality, integrity, authentication, and non-repudiation. PKI can address these needs. The secure E-mail protocol that is currently developing rapidly is Secure/Multipurpose Internet Mail Extensions (S/MIME), which is based on PKI and allows for transfer of encrypted mails with signature.

Web security

For Web security, two peers can establish a Secure Sockets Layer (SSL) connection first for transparent and secure communications at the application layer. With PKI, SSL enables encrypted communications between a browser and a server. Both the communication parties can verify the identity of each other through digital certificates.

Operation of PKI

In a PKI-enabled network, an entity can request a local certificate from the CA and the device can check the validity of certificates. Here is how it works:

1) An entity submits a certificate request to the RA.

2) The RA reviews the identity of the entity and then sends the identity information and the public key with a digital signature to the CA.
3) The CA verifies the digital signature, approves the application, and issues a certificate.
4) The RA receives the certificate from the CA, sends it to the LDAP server to provide directory navigation service, and notifies the entity that the certificate is successfully issued.
5) The entity retrieves the certificate. With the certificate, the entity can communicate with other entities safely through encryption and digital signature.
6) The entity makes a request to the CA when it needs to revoke its certificate, while the CA approves the request, updates the CRLs and publishes the CRLs on the LDAP server.

PKI Configuration Task List

Complete the following tasks to configure PKI:

Task		Remarks
Configuring an Entity DN		Required
Configuring a PKI Domain		Required
Submitting a PKI Certificate Request	Submitting a Certificate Request in Auto Mode	Required Use either approach
	Submitting a Certificate Request in Manual Mode	
Retrieving a Certificate Manually		Optional
Configuring PKI Certificate		Optional
Destroying a Local RSA Key Pair		Optional
Deleting a Certificate		Optional
Configuring an Access Control Policy		Optional

Configuring an Entity DN

A certificate is the binding of a public key and the identity information of an entity, where the identity information is identified by an entity distinguished name (DN). A CA identifies a certificate applicant uniquely by entity DN.

An entity DN is defined by these parameters:

- Common name of the entity.
- Country code of the entity, a standard 2-character code. For example, CN represents China and US represents the United States of America.
- Fully qualified domain name (FQDN) of the entity, a unique identifier of an entity on the network. It consists of a host name and a domain name and can be resolved to an IP address. For example, www.whatever.com is an FQDN, where www is a host name and whatever.com a domain name.
- IP address of the entity.
- Locality where the entity resides.
- Organization to which the entity belongs.
- Unit of the entity in the organization.
- State where the entity resides.

Note

The configuration of an entity DN must comply with the CA certificate issue policy. You need to determine, for example, which entity DN parameters are mandatory and which are optional. Otherwise, certificate request may be rejected.

Follow these steps to configure an entity DN:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an entity and enter its view	**pki entity** *entity-name*	Required No entity exists by default.
Configure the common name for the entity	**common-name** *name*	Optional No common name is specified by default.
Configure the country code for the entity	**country** *country-code-str*	Optional No country code is specified by default.
Configure the FQDN for the entity	**fqdn** *name-str*	Optional No FQDN is specified by default.
Configure the IP address for the entity	**ip** *ip-address*	Optional No IP address is specified by default.
Configure the locality of the entity	**locality** *locality-name*	Optional No locality is specified by default.
Configure the organization name for the entity	**organization** *org-name*	Optional No organization is specified by default.
Configure the unit name for the entity	**organization-unit** *org-unit-name*	Optional No unit is specified by default.
Configure the state or province for the entity	**state** *state-name*	Optional No state or province is specified by default.

Note

- Currently, up to two entities can be created on a device.
- The Windows 2000 CA server has some restrictions on the data length of a certificate request. If the entity DN in a certificate request goes beyond a certain limit, the server will not respond to the certificate request.

Configuring a PKI Domain

Before requesting a PKI certificate, an entity needs to be configured with some enrollment information, which is referred to as a PKI domain. A PKI domain is intended only for convenience of reference by other applications like IKE and SSL, and has only local significance.

A PKI domain is defined by these parameters:

- Trusted CA

An entity requests a certificate from a trusted CA.

- Entity

A certificate applicant uses an entity to provide its identity information to a CA.

- RA

Generally, an independent RA is in charge of certificate request management. It receives the registration request from an entity, checks its qualification, and determines whether to ask the CA to sign a digital certificate. The RA only checks the application qualification of an entity; it does not issue any certificate. Sometimes, the registration management function is provided by the CA, in which case no independent RA is required. You are recommended to deploy an independent RA.

- URL of the registration server

An entity sends a certificate request to the registration server through Simple Certification Enrollment Protocol (SCEP), a dedicated protocol for an entity to communicate with a CA.

- Polling interval and count

After an applicant makes a certificate request, the CA may need a long period of time if it verifies the certificate request manually. During this period, the applicant needs to query the status of the request periodically to get the certificate as soon as possible after the certificate is signed. You can configure the polling interval and count to query the request status.

- IP address of the LDAP server

An LDAP server is usually deployed to store certificates and CRLs. If this is the case, you need to configure the IP address of the LDAP server.

- Fingerprint for root certificate verification

Upon receiving the root certificate of the CA, an entity needs to verify the fingerprint of the root certificate, namely, the hash value of the root certificate content. This hash value is unique to every certificate. If the fingerprint of the root certificate does not match the one configured for the PKI domain, the entity will reject the root certificate.

Follow these steps to configure a PKI domain:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a PKI domain and enter its view	**pki domain** *domain-name*	Required No PKI domain exists by default.
Specify the trusted CA	**ca identifier** *name*	Required No trusted CA is specified by default.

To do…	Use the command…	Remarks
Specify the entity for certificate request	**certificate request entity** *entity-name*	Required No entity is specified by default. The specified entity must exist.
Specify the authority for certificate request	**certificate request from** { **ca** | **ra** }	Required No authority is specified by default.
Configure the URL of the server for certificate request	**certificate request url** *url-string*	Required No URL is configured by default.
Configure the polling interval and attempt limit for querying the certificate request status	**certificate request polling** { **count** *count* | **interval** *minutes* }	Optional The polling is executed for up to 50 times at the interval of 20 minutes by default.
Specify the LDAP server	**ldap-server ip** *ip-address* [**port** *port-number*] [**version** *version-number*]	Optional No LDP server is specified by default.
Configure the fingerprint for root certificate verification	**root-certificate fingerprint** { **md5** | **sha1** } *string*	Required when the certificate request mode is auto and optional when the certificate request mode is manual. In the latter case, if you do not configure this command, the fingerprint of the root certificate must be verified manually. No fingerprint is configured by default.

Note

- Currently, up to two PKI domains can be created on a device.
- The CA name is required only when you retrieve a CA certificate. It is not used when in local certificate request.
- Currently, the URL of the server for certificate request does not support domain name resolving.

Submitting a PKI Certificate Request

When requesting a certificate, an entity introduces itself to the CA by providing its identity information and public key, which will be the major components of the certificate. A certificate request can be submitted to a CA in two ways: online and offline. In offline mode, a certificate request is submitted to a CA by an “out-of-band” means such as phone, disk, or e-mail.

Online certificate request falls into two categories: manual mode and auto mode.

Submitting a Certificate Request in Auto Mode

In auto mode, an entity automatically requests a certificate through the SCEP protocol when it has no local certificate or the present certificate is about to expire.

Follow these steps to configure an entity to submit a certificate request in auto mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PKI domain view	**pki domain** *domain-name*	—
Set the certificate request mode to auto	**certificate request mode auto** [**key-length** *key-length* | **password** { **cipher** | **simple** } *password*] *	Required Manual by default

Submitting a Certificate Request in Manual Mode

In manual mode, you need to retrieve a CA certificate, generate a local RSA key pair, and submit a local certificate request for an entity.

The goal of retrieving a CA certificate is to verify the authenticity and validity of a local certificate.

Generating an RSA key pair is an important step in certificate request. The key pair includes a public key and a private key. The private key is kept by the user, while the public key is transferred to the CA along with some other information. For detailed information about RSA key pair configuration, refer to *Public Key Configuration* in the *Security Volume*.

Follow these steps to submit a certificate request in manual mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PKI domain view	**pki domain** *domain-name*	—
Set the certificate request mode to manual	**certificate request mode manual**	Optional Manual by default
Return to system view	**quit**	—
Retrieve a CA certificate manually	Refer to Retrieving a Certificate Manually	Required
Generate a local RSA key pair	**public-key local create rsa**	Required No local RSA key pair exists by default.
Submit a local certificate request manually	**pki request-certificate domain** *domain-name* [*password*] [**pkcs10** [**filename** *filename*]]	Required

Note

- If a PKI domain already has a local certificate, creating an RSA key pair will result in inconsistency between the key pair and the certificate. To generate a new RSA key pair, delete the local certificate and then issue the **public-key local create** command. For information about the **public-key local create** command, refer to *Public Key Commands* in the *Security Volume*.
- A newly created key pair will overwrite the existing one. If you perform the **public-key local create** command in the presence of a local RSA key pair, the system will ask you whether you want to overwrite the existing one.
- If a PKI domain has already a local certificate, you cannot request another certificate for it. This is to avoid inconsistency between the certificate and the registration information resulting from configuration changes. To request a new certificate, use the **pki delete-certificate** command to delete the existing local certificate and the CA certificate stored locally.
- When it is impossible to request a certificate from the CA through SCEP, you can save the request information by using the **pki request-certificate domain** command with the **pkcs10** and **filename** keywords, and then send the file to the CA by an out-of-band means.
- Make sure the clocks of the entity and the CA are synchronous. Otherwise, the validity period of the certificate will be abnormal.
- The **pki request-certificate domain** configuration will not be saved in the configuration file.

Retrieving a Certificate Manually

You can download an existing CA certificate, local certificate, or peer entity certificate from the CA server and save it locally. To do so, you can use two ways: online and offline. In offline mode, you need to retrieve a certificate by an out-of-band means like FTP, disk, e-mail and then import it into the local PKI system.

Certificate retrieval serves two purposes:

- Locally store the certificates associated with the local security domain for improved query efficiency and reduced query count,
- Prepare for certificate verification.

Before retrieving a local certificate in online mode, be sure to complete LDAP server configuration.

Follow these steps to retrieve a certificate manually:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Retrieve a certificate manually	Online	**pki retrieval-certificate** { **ca** | **local** } **domain** *domain-name*	Required Use either command.
	Offline	**pki import-certificate** { **ca** | **local** } **domain** *domain-name* { **der** | **p12** | **pem** } [**filename** *filename*]	

Caution

- If a PKI domain already has a CA certificate, you cannot retrieve another CA certificate for it. This is in order to avoid inconsistency between the certificate and registration information due to related configuration changes. To retrieve a new CA certificate, use the **pki delete-certificate** command to delete the existing CA certificate and local certificate first.
- The **pki retrieval-certificate** configuration will not be saved in the configuration file.

Configuring PKI Certificate Verification

A certificate needs to be verified before being used. Verifying a certificate is to check that the certificate is signed by the CA and that the certificate has neither expired nor been revoked.

Before verifying a certificate, you need to retrieve the CA certificate.

You can specify whether CRL checking is required in certificate verification. If you enable CRL checking, CRLs will be used in verification of a certificate.

Configuring CRL-checking-enabled PKI certificate verification

Follow these steps to configure CRL-checking-enabled PKI certificate verification:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PKI domain view	**pki domain** *domain-name*	—
Specify the URL of the CRL distribution point	**crl url** *url-string*	Optional No CRL distribution point URL is specified by default.
Set the CRL update period	**crl update-period** *hours*	Optional By default, the CRL update period depends on the next update field in the CRL file.
Enable CRL checking	**crl check enable**	Optional Enabled by default
Return to system view	**quit**	—
Retrieve the CA certificate	Refer to Retrieving a Certificate Manually	Required
Retrieve CRLs	**pki retrieval-crl domain** *domain-name*	Required
Verify the validity of a certificate	**pki validate-certificate** { **ca** | **local** } **domain** *domain-name*	Required

Configuring CRL-checking-disabled PKI certificate verification

Follow these steps to configure CRL-checking-disabled PKI certificate verification:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PKI domain view	**pki domain** *domain-name*	—
Disable CRL checking	**crl check disable**	Required Enabled by default
Return to system view	**quit**	—
Retrieve the CA certificate	Refer to Retrieving a Certificate Manually	Required
Verify the validity of the certificate	**pki validate-certificate** { **ca** | **local** } **domain** *domain-name*	Required

Note

- The CRL update period refers to the interval at which the entity downloads CRLs from the CRL server. The CRL update period configured manually is prior to that specified in the CRLs.
- The **pki retrieval-crl domain** configuration will not be saved in the configuration file.
- Currently, the URL of the CRL distribution point does not support domain name resolving.

Destroying a Local RSA Key Pair

A certificate has a lifetime, which is determined by the CA. When the private key leaks or the certificate is about to expire, you can destroy the old RSA key pair and then create a pair to request a new certificate.

Follow these steps to destroy a local RSA key pair:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Destroy a local RSA key pair	**public-key local destroy rsa**	Required

Note

For details about the **public-key local destroy** command, refer to *Public Key Commands* in the *Security Volume*.

Deleting a Certificate

When a certificate requested manually is about to expire or you want to request a new certificate, you can delete the current local certificate or CA certificate.

Follow these steps to delete a certificate:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Delete certificates	**pki delete-certificate** { **ca** | **local** } **domain** *domain-name*	Required

Configuring an Access Control Policy

By configuring a certificate attribute-based access control policy, you can further control access to the server, providing additional security for the server.

Follow these steps to configure a certificate attribute-based access control policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a certificate attribute group and enter its view	**pki certificate attribute-group** *group-name*	Required No certificate attribute group exists by default.
Configure an attribute rule for the certificate issuer name, certificate subject name, or alternative subject name	**attribute** *id* { **alt-subject-name** { **fqdn** | **ip** } | { **issuer-name** | **subject-name** } { **dn** | **fqdn** | **ip** } } { **ctn** | **equ** | **nctn** | **nequ** } *attribute-value*	Optional There is no restriction on the issuer name, certificate subject name and alternative subject name by default.
Return to system view	**quit**	—
Create a certificate attribute-based access control policy and enter its view	**pki certificate access-control-policy** *policy-name*	Required No access control policy exists by default.
Configure a certificate attribute-based access control rule	**rule** [*id*] { **deny** | **permit** } *group-name*	Required No access control rule exists by default.

Caution

A certificate attribute group must exist to be associated with a rule.

Displaying and Maintaining PKI

To do…	Use the command…	Remarks
Display the contents or request status of a certificate	**display pki certificate** { { **ca** | **local** } **domain** *domain-name* | **request-status** }	Available in any view
Display CRLs	**display pki crl domain** *domain-name*	Available in any view
Display information about one or all certificate attribute groups	**display pki certificate attribute-group** { *group-name* | **all** }	Available in any view

To do…	Use the command…	Remarks
Display information about one or all certificate attribute-based access control policies	**display pki certificate access-control-policy** { *policy-name* | **all** }	Available in any view

PKI Configuration Examples

Caution

- The SCEP plug-in is required when you use the Windows Server as the CA. In this case, when configuring the PKI domain, you need to use the **certificate request from ra** command to specify that the entity requests a certificate from an RA.
- The SCEP plug-in is not required when RSA Keon is used. In this case, when configuring a PKI domain, you need to use the **certificate request from ca** command to specify that the entity requests a certificate from a CA.

Requesting a Certificate from a CA Running RSA Keon

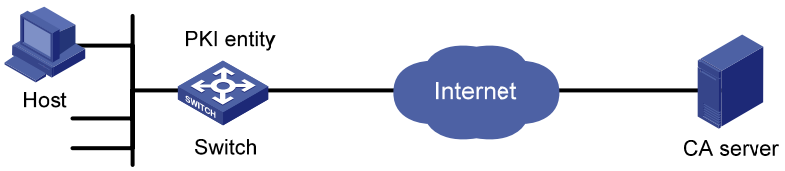
Note

The CA server runs RSA Keon in this configuration example.

Network requirements

- The device submits a local certificate request to the CA server.
- The device acquires the CRLs for certificate verification.

Figure 1-2 Request a certificate from a CA running RSA Keon



Configuration procedure

1) Configure the CA server

Create a CA server named **myca**.

In this example, you need to configure these basic attributes on the CA server at first:

- Nickname: Name of the trusted CA.

- Subject DN: DN information of the CA, including the Common Name (CN), Organization Unit (OU), Organization (O), and Country (C).

The other attributes may be left using the default values.

Configure extended attributes.

After configuring the basic attributes, you need to perform configuration on the jurisdiction configuration page of the CA server. This includes selecting the proper extension profiles, enabling the SCEP autovetting function, and adding the IP address list for SCEP autovetting.

Configure the CRL distribution behavior.

After completing the above configuration, you need to perform CRL related configurations. In this example, select the local CRL distribution mode of HTTP and set the HTTP URL to http://4.4.4.133:447/myca.crl.

After the above configuration, make sure that the system clock of the device is synchronous to that of the CA, so that the device can request certificates and retrieve CRLs properly.

2) Configure the switch

- Configure the entity DN

Configure the entity name as **aaa** and the common name as **switch**.

```
<Switch> system-view
[Switch] pki entity aaa
[Switch-pki-entity-aaa] common-name switch
[Switch-pki-entity-aaa] quit
```

- Configure the PKI domain

Create PKI domain **torsa** and enter its view.

```
[Switch] pki domain torsa
```

Configure the name of the trusted CA as **myca**.

```
[Switch-pki-domain-torsa] ca identifier myca
```

Configure the URL of the registration server in the format of http://host:port/Issuing Jurisdiction ID, where Issuing Jurisdiction ID is a hexadecimal string generated on the CA server.

```
[Switch-pki-domain-torsa] certificate request url
http://4.4.4.133:446/c95e970f632d27be5e8cbf80e971d9c4a9a93337
```

Set the registration authority to **CA**.

```
[Switch-pki-domain-torsa] certificate request from ca
```

Specify the entity for certificate request as **aaa**.

```
[Switch-pki-domain-torsa] certificate request entity aaa
```

Configure the URL for the CRL distribution point.

```
[Switch-pki-domain-torsa] crl url http://4.4.4.133:447/myca.crl
[Switch-pki-domain-torsa] quit
```

- Generate a local key pair using RSA

```
[Switch] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits in the modulus [default = 1024]:
```

```
Generating Keys...
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++
```

- Apply for certificates

Retrieve the CA certificate and save it locally.

```
[Switch] pki retrieval-certificate ca domain torsa
Retrieving CA/RA certificates. Please wait a while......
The trusted CA's finger print is:
    MD5  fingerprint:EDE9 0394 A273 B61A F1B3 0072 A0B1 F9AB
    SHA1 fingerprint: 77F9 A077 2FB8 088C 550B A33C 2410 D354 23B2 73A8

Is the finger print correct?(Y/N):y

Saving CA/RA certificates chain, please wait a moment......
CA certificates retrieval success.
```

Retrieve CRLs and save them locally.

```
[Switch] pki retrieval-crl domain torsa
Connecting to server for retrieving CRL. Please wait a while.....
CRL retrieval success!
```

Request a local certificate manually.

```
[Switch] pki request-certificate domain torsa challenge-word
Certificate is being requested, please wait......
[Switch]
Enrolling the local certificate,please wait a while......
Certificate request Successfully!
Saving the local certificate to device......
Done!
```

3) Verify your configuration

Use the following command to view information about the local certificate acquired.

```
<Switch> display pki certificate local domain torsa
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number:
            9A96A48F 9A509FD7 05FFF4DF 104AD094
        Signature Algorithm: sha1WithRSAEncryption
        Issuer:
            C=cn
            O=org
            OU=test
            CN=myca
        Validity
            Not Before: Jan  8 09:26:53 2007 GMT
```

```
            Not After : Jan  8 09:26:53 2008 GMT
        Subject:
            CN=switch
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00D67D50 41046F6A 43610335 CA6C4B11
                    F8F89138 E4E905BD 43953BA2 623A54C0
                    EA3CB6E0 B04649CE C9CDDD38 34015970
                    981E96D9 FF4F7B73 A5155649 E583AC61
                    D3A5C849 CBDE350D 2A1926B7 0AE5EF5E
                    D1D8B08A DBF16205 7C2A4011 05F11094
                    73EB0549 A65D9E74 0F2953F2 D4F0042F
                    19103439 3D4F9359 88FB59F3 8D4B2F6C
                    2B
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 CRL Distribution Points:
            URI:http://4.4.4.133:447/myca.crl

    Signature Algorithm: sha1WithRSAEncryption
        836213A4 F2F74C1A 50F4100D B764D6CE
        B30C0133 C4363F2F 73454D51 E9F95962
        EDE9E590 E7458FA6 765A0D3F C4047BC2
        9C391FF0 7383C4DF 9A0CCFA9 231428AF
        987B029C C857AD96 E4C92441 9382E798
        8FCC1E4A 3E598D81 96476875 E2F86C33
        75B51661 B6556C5E 8F546E97 5197734B
        C8C29AC7 E427C8E4 B9AAF5AA 80A75B3C
```

You can also use some other **display** commands to view detailed information about the CA certificate and CRLs. Refer to the parts related to **display pki certificate ca domain** and **display pki crl domain** commands in *PKI Commands* of the *Security Volume*.

Requesting a Certificate from a CA Running Windows 2003 Server

---

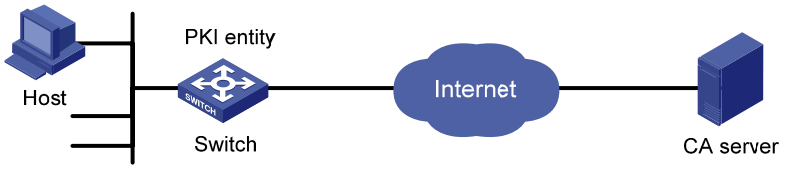
 **Note**

The CA server runs the Windows 2003 server in this configuration example.

Network requirements

Configure PKI entity Switch to request a local certificate from the CA server.

Figure 1-3 Request a certificate from a CA running Windows 2003 server



Configuration procedure

1) Configure the CA server

- Install the certificate server suites

From the start menu, select **Control Panel** > **Add or Remove Programs**, and then select **Add/Remove Windows Components** > **Certificate Services** and click **Next** to begin the installation.

- Install the SCEP plug-in

As a CA server running the Windows 2003 server does not support SCEP by default, you need to install the SCEP plug-in so that the switch can register and obtain its certificate automatically. After the SCEP plug-in installation completes, a URL is displayed, which you need to configure on the switch as the URL of the server for certificate registration.

- Modify the certificate service attributes

From the start menu, select **Control Panel** > **Administrative Tools** > **Certificate Authority**. If the CA server and SCEP plug-in have been installed successfully, there should be two certificates issued by the CA to the RA. Right-click on the CA server in the navigation tree and select **Properties** > **Policy Module**. Click **Properties** and then select **Follow the settings in the certificate template, if applicable. Otherwise, automatically issue the certificate**.

- Modify the Internet Information Services (IIS) attributes

From the start menu, select **Control Panel** > **Administrative Tools** > **Internet Information Services (IIS) Manager** and then select **Web Sites** from the navigation tree. Right-click on **Default Web Site** and select **Properties** > **Home Directory**. Specify the path for certificate service in the **Local path** text box. In addition, you are recommended to specify an available port number as the TCP port number of the default Web site to avoid conflict with existing services.

After completing the above configuration, check that the system clock of the switch is synchronous to that of the CA server, ensuring that the switch can request a certificate normally.

2) Configure the switch

- Configure the entity DN

Configure the entity name as **aaa** and the common name as **switch**.

```
<Switch> system-view
[Switch] pki entity aaa
[Switch-pki-entity-aaa] common-name switch
[Switch-pki-entity-aaa] quit
```

- Configure the PKI domain

Create PKI domain **torsa** and enter its view.

```
[Switch] pki domain torsa
```

Configure the name of the trusted CA as **myca**.

```
[Switch-pki-domain-torsa] ca identifier myca
```

Configure the URL of the registration server in the format of http://host:port/ certsrv/mscep/mscep.dll, where host:port indicates the IP address and port number of the CA server.

```
[Switch-pki-domain-torsa]              certificate              request              url
http://4.4.4.1:8080/certsrv/mscep/mscep.dll
```

Set the registration authority to **RA**.

```
[Switch-pki-domain-torsa] certificate request from ra
```

Specify the entity for certificate request as **aaa**.

```
[Switch-pki-domain-torsa] certificate request entity aaa
```

- Generate a local key pair using RSA

```
[Switch] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits in the modulus [default = 1024]:
Generating Keys...
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++++++++++++++
++++++++++++++++++++++
.
```

- Apply for certificates

Retrieve the CA certificate and save it locally.

```
[Switch] pki retrieval-certificate ca domain torsa
Retrieving CA/RA certificates. Please wait a while......
The trusted CA's finger print is:
    MD5  fingerprint:766C D2C8 9E46 845B 4DCE 439C 1C1F 83AB
    SHA1 fingerprint:97E5 DDED AB39 3141 75FB DB5C E7F8 D7D7 7C9B 97B4

Is the finger print correct?(Y/N):y

Saving CA/RA certificates chain, please wait a moment......
CA certificates retrieval success.
```

Request a local certificate manually.

```
[Switch] pki request-certificate domain torsa challenge-word
Certificate is being requested, please wait......
[Switch]
Enrolling the local certificate,please wait a while......
Certificate request Successfully!
Saving the local certificate to device......
Done!
```

3) Verify your configuration

Use the following command to view information about the local certificate acquired.

```
<Switch> display pki certificate local domain torsa
Certificate:
```

```
    Data:
        Version: 3 (0x2)
        Serial Number:
            48FA0FD9 00000000 000C
        Signature Algorithm: sha1WithRSAEncryption
        Issuer:
            CN=CA server
        Validity
            Not Before: Nov 21 12:32:16 2007 GMT
            Not After : Nov 21 12:42:16 2008 GMT
        Subject:
            CN=switch
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (1024 bit)
                Modulus (1024 bit):
                    00A6637A 8CDEA1AC B2E04A59 F7F6A9FE
                    5AEE52AE 14A392E4 E0E5D458 0D341113
                    0BF91E57 FA8C67AC 6CE8FEBB 5570178B
                    10242FDD D3947F5E 2DA70BD9 1FAF07E5
                    1D167CE1 FC20394F 476F5C08 C5067DF9
                    CB4D05E6 55DC11B6 9F4C014D EA600306
                    81D403CF 2D93BC5A 8AF3224D 1125E439
                    78ECEFE1 7FA9AE7B 877B50B8 3280509F
                    6B
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Subject Key Identifier:
            B68E4107 91D7C44C 7ABCE3BA 9BF385F8 A448F4E1
            X509v3 Authority Key Identifier:
            keyid:9D823258 EADFEFA2 4A663E75 F416B6F6 D41EE4FE

            X509v3 CRL Distribution Points:
            URI:http://l00192b/CertEnroll/CA%20server.crl
            URI:file://\\l00192b\CertEnroll\CA server.crl

            Authority Information Access:
            CA Issuers - URI:http://l00192b/CertEnroll/l00192b_CA%20server.crt
            CA Issuers - URI:file://\\l00192b\CertEnroll\l00192b_CA server.crt

            1.3.6.1.4.1.311.20.2:
                .0.I.P.S.E.C.I.n.t.e.r.m.e.d.i.a.t.e.O.f.f.l.i.n.e
    Signature Algorithm: sha1WithRSAEncryption
        81029589 7BFA1CBD 20023136 B068840B
(Omitted)
```

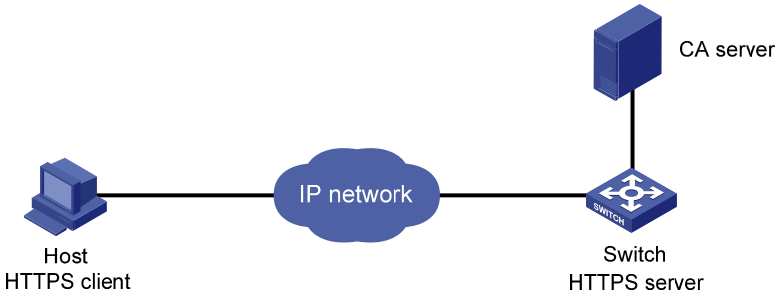
You can also use some other **display** commands to view detailed information about the CA certificate. Refer to the **display pki certificate ca domain** command in *PKI Commands* of the *Security Volume*.

Configuring a Certificate Attribute-Based Access Control Policy

Network requirements

- The client accesses the remote HTTP Security (HTTPS) server through the HTTPS protocol.
- SSL is configured to ensure that only legal clients log into the HTTPS server.
- Create a certificate attribute-based access control policy to control access to the HTTPS server.

Figure 1-4 Configure a certificate attribute-based access control policy



Configuration procedure

Note

- For detailed information about SSL configuration, refer to *SSL Configuration* in the *Security Volume*.
- For detailed information about HTTPS configuration, refer to *HTTP Server Configuration* in the *System Volume*.
- The PKI domain to be referenced by the SSL policy must be created in advance. For detailed configuration of the PKI domain, refer to Configure the PKI domain.

---

1) Configure the HTTPS server

Configure the SSL policy for the HTTPS server to use.

```
<Switch> system-view
[Switch] ssl server-policy myssl
[Switch-ssl-server-policy-myssl] pki-domain 1
[Switch-ssl-server-policy-myssl] client-verify enable
[Switch-ssl-server-policy-myssl] quit
```

2) Configure the certificate attribute group

Create certificate attribute group **mygroup1** and add two attribute rules. The first rule defines that the DN of the subject name includes the string **aabbcc**, and the second rule defines that the IP address of the certificate issuer is 10.0.0.1.

```
[Switch] pki certificate attribute-group mygroup1
[Switch-pki-cert-attribute-group-mygroup1] attribute 1 subject-name dn ctn aabbcc
[Switch-pki-cert-attribute-group-mygroup1] attribute 2 issuer-name ip equ 10.0.0.1
[Switch-pki-cert-attribute-group-mygroup1] quit
```

Create certificate attribute group **mygroup2** and add two attribute rules. The first rule defines that the FQDN of the alternative subject name does not include the string of **apple**, and the second rule defines that the DN of the certificate issuer name includes the string **aabbcc**.

```
[Switch] pki certificate attribute-group mygroup2
[Switch-pki-cert-attribute-group-mygroup2] attribute 1 alt-subject-name fqdn nctn apple
[Switch-pki-cert-attribute-group-mygroup2] attribute 2 issuer-name dn ctn aabbcc
[Switch-pki-cert-attribute-group-mygroup2] quit
```

3) Configure the certificate attribute-based access control policy

Create the certificate attribute-based access control policy of **myacp** and add two access control rules.

```
[Switch] pki certificate access-control-policy myacp
[Switch-pki-cert-acp-myacp] rule 1 deny mygroup1
[Switch-pki-cert-acp-myacp] rule 2 permit mygroup2
[Switch-pki-cert-acp-myacp] quit
```

4) Apply the SSL server policy and certificate attribute-based access control policy to HTTPS service and enable HTTPS service.

Apply SSL server policy **myssl** to HTTPS service.

```
[Switch] ip https ssl-server-policy myssl
```

Apply the certificate attribute-based access control policy of **myacp** to HTTPS service.

```
[Switch] ip https certificate access-control-policy myacp
```

Enable HTTPS service.

```
[Switch] ip https enable
```

Troubleshooting PKI

Failed to Retrieve a CA Certificate

Symptom

Failed to retrieve a CA certificate.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No trusted CA is specified.
- The URL of the registration server for certificate request is not correct or not configured.
- No authority is specified for certificate request.
- The system clock of the device is not synchronized with that of the CA.

Solution

- Make sure that the network connection is physically proper.
- Check that the required commands are configured properly.
- Use the **ping** command to check that the RA server is reachable.
- Specify the authority for certificate request.
- Synchronize the system clock of the device with that of the CA.

Failed to Request a Local Certificate

Symptom

Failed to request a local certificate.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No CA certificate has been retrieved.
- The current key pair has been bound to a certificate.
- No trusted CA is specified.
- The URL of the registration server for certificate request is not correct or not configured.
- No authority is specified for certificate request.
- Some required parameters of the entity DN are not configured.

Solution

- Make sure that the network connection is physically proper.
- Retrieve a CA certificate.
- Regenerate a key pair.
- Specify a trusted CA.
- Use the **ping** command to check that the RA server is reachable.
- Specify the authority for certificate request.
- Configure the required entity DN parameters.

Failed to Retrieve CRLs

Symptom

Failed to retrieve CRLs.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No CA certificate has been retrieved before you try to retrieve CRLs.
- The IP address of LDAP server is not configured.
- The CRL distribution URL is not configured.
- The LDAP server version is wrong.

Solution

- Make sure that the network connection is physically proper.
- Retrieve a CA certificate.
- Specify the IP address of the LDAP server.
- Specify the CRL distribution URL.
- Re-configure the LDAP version.

Table of Contents

1 SSL Configuration

When configuring SSL, go to these sections for information you are interested in:

- SSL Overview
- SSL Configuration Task List
- Displaying and Maintaining SSL
- Troubleshooting SSL

SSL Overview

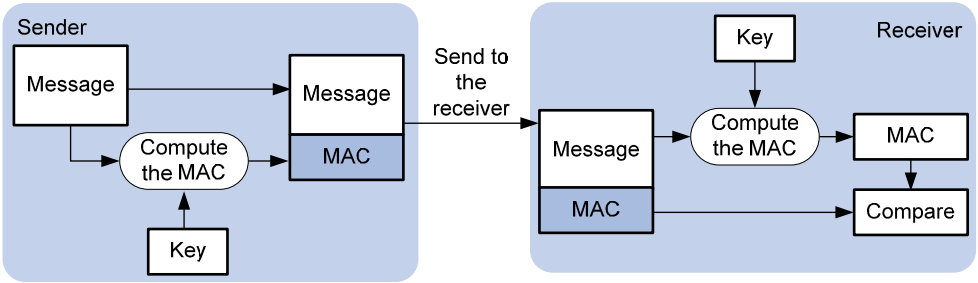
Secure Sockets Layer (SSL) is a security protocol providing secure connection service for TCP-based application layer protocols, for example, HTTP protocol. It is widely used in E-business and online bank fields to provide secure data transmission over the Internet.

SSL Security Mechanism

SSL provides these security services:

- Confidentiality: SSL uses a symmetric encryption algorithm to encrypt data and uses the asymmetric key algorithm of Rivest, Shamir, and Adelman (RSA) to encrypt the key to be used by the symmetric encryption algorithm.
- Authentication: SSL supports certificate-based identity authentication of the server and client by using the digital signatures, with the authentication of the client being optional. The SSL server and client obtain certificates from a certificate authority (CA) through the Public Key Infrastructure (PKI).
- Reliability: SSL uses the key-based message authentication code (MAC) to verify message integrity. A MAC algorithm transforms a message of any length to a fixed-length message. Figure 1-1 illustrates how SSL uses a MAC algorithm to verify message integrity. With the key, the sender uses the MAC algorithm to compute the MAC value of a message. Then, the sender suffixes the MAC value to the message and sends the result to the receiver. The receiver uses the same key and MAC algorithm to compute the MAC value of the received message, and compares the locally computed MAC value with that received. If the two matches, the receiver considers the message intact; otherwise, the receiver considers that the message has been tampered with in transit and discards the message.

Figure 1-1 Message integrity verification by a MAC algorithm



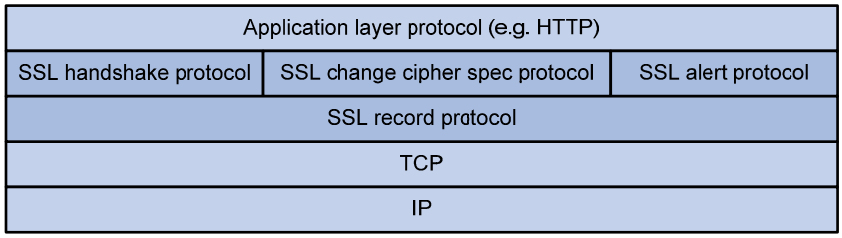
Note

- For details about symmetric key algorithms, asymmetric key algorithm RSA and digital signature, refer to *Public Key Configuration* in the *Security Volume*.
- For details about PKI, certificate, and CA, refer to *PKI Configuration* in the *Security Volume*.

SSL Protocol Stack

As shown in Figure 1-2, the SSL protocol consists of two layers of protocols: the SSL record protocol at the lower layer and the SSL handshake protocol, change cipher spec protocol, and alert protocol at the upper layer.

Figure 1-2 SSL protocol stack



- SSL handshake protocol: As a very important part of the SSL protocol stack, it is responsible for negotiating the cipher suite to be used during communication (including the symmetric encryption algorithm, key exchange algorithm, and MAC algorithm), exchanging the key between the server and client, and implementing identity authentication of the server and client. Through the SSL handshake protocol, a session is established between a client and the server. A session consists of a set of parameters, including the session ID, peer certificate, cipher suite, and master secret.
- SSL change cipher spec protocol: Used for notification between a client and the server that the subsequent packets are to be protected and transmitted based on the newly negotiated cipher suite and key.
- SSL alert protocol: Allowing a client and the server to send alert messages to each other. An alert message contains the alert severity level and a description.
- SSL record protocol: Fragmenting and compressing data to be transmitted, calculating and adding MAC to the data, and encrypting the data before transmitting it to the peer end.

SSL Configuration Task List

Different parameters are required on the SSL server and the SSL client.

Complete the following tasks to configure SSL:

Task	Remarks
Configuring an SSL Server Policy	Required
Configuring an SSL Client Policy	Optional

Configuring an SSL Server Policy

An SSL server policy is a set of SSL parameters for a server to use when booting up. An SSL server policy takes effect only after it is associated with an application layer protocol, HTTP protocol, for example.

Configuration Prerequisites

When configuring an SSL server policy, you need to specify the PKI domain to be used for obtaining the server side certificate. Therefore, before configuring an SSL server policy, you must configure a PKI domain. For details about PKI domain configuration, refer to *PKI Configuration* in the *Security Volume*.

Configuration Procedure

Follow these steps to configure an SSL server policy:

To do...	Use the command...	Remarks
Enter system view	**system-view**	—
Create an SSL server policy and enter its view	**ssl server-policy** *policy-name*	Required
Specify a PKI domain for the SSL server policy	**pki-domain** *domain-name*	Required By default, no PKI domain is specified for an SSL server policy.
Specify the cipher suite(s) for the SSL server policy to support	**ciphersuite** [**rsa_aes_128_cbc_sha** | **rsa_des_cbc_sha** | **rsa_rc4_128_md5** | **rsa_rc4_128_sha**] *	Optional By default, an SSL server policy supports all cipher suites.
Set the handshake timeout time for the SSL server	**handshake timeout** *time*	Optional 3,600 seconds by default
Configure the SSL connection close mode	**close-mode wait**	Optional Not wait by default
Set the maximum number of cached sessions and the caching timeout time	**session** { **cachesize** *size* | **timeout** *time* } *	Optional The defaults are as follows: 500 for the maximum number of cached sessions, 3600 seconds for the caching timeout time.
Enable certificate-based SSL client authentication	**client-verify enable**	Optional Not enabled by default

Note

- If you enable client authentication here, you must request a local certificate for the client.
- Currently, SSL mainly comes in these versions: SSL 2.0, SSL 3.0, and TLS 1.0, where TLS 1.0 corresponds to SSL 3.1. When the device acts as an SSL server, it can communicate with clients running SSL 3.0 or TLS 1.0, and can identify Hello packets from clients running SSL 2.0. If a client running SSL 2.0 also supports SSL 3.0 or TLS 1.0 (information about supported versions is carried in the packet that the client sends to the server), the server will notify the client to use SSL 3.0 or TLS 1.0 to communicate with the server.

SSL Server Policy Configuration Example

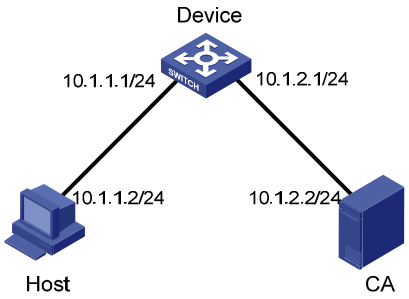
Network requirements

- Device works as the HTTPS server.
- A host works as the client and accesses the HTTPS server through HTTP secured with SSL.
- A certificate authority (CA) issues a certificate to Device.

Caution

In this instance, Windows Server works as the CA and the Simple Certificate Enrollment Protocol (SCEP) plug-in is installed on the CA.

Figure 1-3 Network diagram for SSL server policy configuration



Configuration procedure

1) Request a certificate for Device

Create a PKI entity named **en** and configure it.

```
<Device> system-view
[Device] pki entity en
[Device-pki-entity-en] common-name http-server1
[Device-pki-entity-en] fqdn ssl.security.com
[Device-pki-entity-en] quit
```

Create a PKI domain and configure it.

```
[Device] pki domain 1
[Device-pki-domain-1] ca identifier ca1
[Device-pki-domain-1] certificate request url http://10.1.2.2/certsrv/mscep/mscep.dll
[Device-pki-domain-1] certificate request from ra
[Device-pki-domain-1] certificate request entity en
[Device-pki-domain-1] quit
```

Create the local RSA key pairs.

```
[Device] public-key local create rsa
```

Retrieve the CA certificate.

```
[Device] pki retrieval-certificate ca domain 1
```

Request a local certificate.

```
[Device] pki request-certificate domain 1
```

2) Configure an SSL server policy

Create an SSL server policy named **myssl**.

```
[Device] ssl server-policy myssl
```

Specify the PKI domain for the SSL server policy as 1.

```
[Device-ssl-server-policy-myssl] pki-domain 1
```

Enable client authentication.

```
[Device-ssl-server-policy-myssl] client-verify enable
[Device-ssl-server-policy-myssl] quit
```

3) Associate HTTPS service with the SSL server policy and enable HTTPS service

Configure HTTPS service to use SSL server policy myssl.

```
[Device] ip https ssl-server-policy myssl
```

Enable HTTPS service.

```
[Device] ip https enable
```

4) Verify your configuration

Launch IE on the host and enter https://10.1.1.1 in the address bar. You should be able to log in to Device and manage it.

 **Note**

- For details about PKI configuration commands, refer to *PKI Commands* in the *Security Volume*.
- For details about the **public-key local create rsa** command, refer to *Public Key Commands* in the *Security Volume*.
- For details about HTTPS, refer to *HTTP Configuration* in the *System Volume*.

Configuring an SSL Client Policy

An SSL client policy is a set of SSL parameters for a client to use when connecting to the server. An SSL client policy takes effect only after it is associated with an application layer protocol.

Configuration Prerequisites

If the SSL server is configured to authenticate the SSL client, when configuring the SSL client policy, you need to specify the PKI domain to be used for obtaining the certificate of the client. Therefore, before configuring an SSL client policy, you must configure a PKI domain. For details about PKI domain configuration, refer to *PKI Configuration* in the *Security Volume*.

Configuration Procedure

Follow these steps to configure an SSL client policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an SSL client policy and enter its view	**ssl client-policy** *policy-name*	Required
Specify a PKI domain for the SSL client policy	**pki-domain** *domain-name*	Required No PKI domain is configured by default.
Specify the preferred cipher suite for the SSL client policy	**prefer-cipher** { **rsa_aes_128_cbc_sha** | **rsa_des_cbc_sha** | **rsa_rc4_128_md5** | **rsa_rc4_128_sha** }	Optional **rsa_rc4_128_md5** by default
Specify the SSL protocol version for the SSL client policy	**version** { **ssl3.0** | **tls1.0** }	Optional TLS 1.0 by default

Note

If you enable client authentication on the server, you must request a local certificate for the client.

Displaying and Maintaining SSL

To do…	Use the command…	Remarks
Display SSL server policy information	**display ssl server-policy** { *policy-name* | **all** }	Available in any view
Display SSL client policy information	**display ssl client-policy** { *policy-name* | **all** }	

Troubleshooting SSL

SSL Handshake Failure

Symptom

As the SSL server, the device fails to handshake with the SSL client.

Analysis

SSL handshake failure may result from the following causes:

- No SSL server certificate exists, or the certificate is not trusted.
- The server is expected to authenticate the client, but the SSL client has no certificate or the certificate is not trusted.
- The cipher suites used by the server and the client do not match.

Solution

1) You can issue the **debugging ssl** command and view the debugging information to locate the problem:

- If the SSL server has no certificate, request one for it.
- If the server certificate cannot be trusted, install on the SSL client the root certificate of the CA that issues the local certificate to the SSL server, or let the server requests a certificate from the CA that the SSL client trusts.
- If the SSL server is configured to authenticate the client, but the certificate of the SSL client does not exist or cannot be trusted, request and install a certificate for the client.

2) You can use the **display ssl server-policy** command to view the cipher suite used by the SSL server policy. If the cipher suite used by the SSL server does not match that used by the client, use the **ciphersuite** command to modify the cipher suite of the SSL server.

Table of Contents

1 Public Key Configuration

When configuring public keys, go to these sections for information you are interested in:

- Public Key Algorithm Overview
- Configuring the Local Asymmetric Key Pair
- Configuring the Public Key of a Peer
- Displaying and Maintaining Public Keys
- Public Key Configuration Examples

Public Key Algorithm Overview

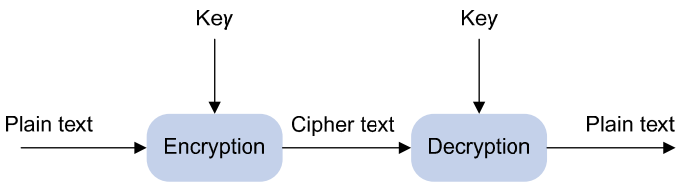
Basic Concepts

- Algorithm: A set of transformation rules for encryption and decryption.
- Plain text: Information without being encrypted.
- Cipher text: Encrypted information.
- Key: A string of characters that controls the transformation between plain text and cipher text. It participates in both the encryption and decryption.

Key Algorithm Types

As shown in Figure 1-1, the information is encrypted before being sent for confidentiality. The cipher text is transmitted in the network, and then is decrypted by the receiver to obtain the original pain text.

Figure 1-1 Encryption and decryption



There are two types of key algorithms, based on whether the keys for encryption and decryption are the same:

- Symmetric key algorithm: The same key is used for both encryption and decryption. Commonly used symmetric key algorithms include AES and DES.
- Asymmetric key algorithm: Also called public key algorithm. Both ends have their own key pair, consisting of a private key and a public key. The private key is kept secret while the public key may be distributed widely. The private key cannot be practically derived from the public key. The information encrypted with the public key/private key can be decrypted only with the corresponding private key/public key.

Asymmetric Key Algorithm Applications

Asymmetric key algorithms can be used for encryption and digital signature:

- Encryption: The information encrypted with a receiver's public key can be decrypted by the receiver possessing the corresponding private key. This is used to ensure confidentiality.
- Digital signature: The information encrypted with a sender's private key can be decrypted by anyone who has access to the sender's public key, thereby proving that the information is from the sender and has not been tampered with. For example, user 1 adds a signature to the data using the private key, and then sends the data to user 2. User 2 verifies the signature using the public key of user 1. If the signature is correct, the data is considered from user 1.

Revest-Shamir-Adleman Algorithm (RSA), and Digital Signature Algorithm (DSA) are all asymmetric key algorithms. RSA can be used for data encryption and signature, whereas DSA are used for signature only.

Note

Asymmetric key algorithms are usually used in digital signature applications for peer identity authentication because they involve complex calculations and are time-consuming; symmetric key algorithms are often used to encrypt data for security.

Configuring the Local Asymmetric Key Pair

You can create and destroy a local asymmetric key pair, and export the host public key of a local asymmetric key pair.

Creating an Asymmetric Key Pair

Follow these steps to create an asymmetric key pair:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a local DSA , or RSA key pairs	**public-key local create** { **dsa** | **rsa** }	Required By default, there is no such key pair.

Note

- Configuration of the **public-key local create** command can survive a reboot.
- The **public-key local create rsa** command generates two key pairs: one server key pair and one host key pair. Each key pair consists of a public key and a private key.
- The length of an RSA key modulus is in the range 512 to 2048 bits. After entering the **public-key local create rsa** command, you will be required to specify the modulus length. For security, a modulus of at least 768 bits is recommended.
- The **public-key local create dsa** command generates only one key pair, that is, the host key pair.
- The length of a DSA key modulus is in the range 512 to 2048 bits. After entering the **public-key local create dsa** command, you will be required to specify the modulus length. For security, a modulus of at least 768 bits is recommended.

Displaying or Exporting the Local RSA or DSA Host Public Key

You can display the local RSA or DSA host public key on the screen or export it to a specified file, so as to configure the local RSA or DSA host public key on the remote end.

Follow these steps to display or export the local RSA or DSA host public key:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Display the local RSA host public key on the screen in a specified format, or export it to a specified file	**public-key local export rsa** { **openssh** | **ssh1** | **ssh2** } [*filename*]	Select a command according to the type of the key to be exported.
Display the local DSA host public key on the screen in a specified format, or export it to a specified file	**public-key local export dsa** { **openssh** | **ssh2** } [*filename*]	

Destroying an Asymmetric Key Pair

An asymmetric key pair may expire or leak. In this case, you need to destroy it and generate a new pair.

Follow these steps to destroy an asymmetric key pair:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Destroy an asymmetric key pair	**public-key local destroy** { **dsa** | **rsa** }	Required

Configuring the Public Key of a Peer

To authenticate the remote host, you need to configure the RSA or DSA public key of that peer on the local host.

To configure the public key of the peer, you can:

- Configure it manually: You can input on or copy the public key of the peer to the local host. The copied public key must have not been converted and be in the distinguished encoding rules (DER) encoding format.
- Import it from the public key file: The system automatically converts the public key to a string coded using the PKCS (Public Key Cryptography Standards). Before importing the public key, you must upload the peer's public key file (in binary) to the local host through FTP or TFTP.

Caution

- You are recommended to configure the public key of the peer by importing it from a public key file.
- The device supports up to 20 host pubic keys of peers.

Follow these steps to configure the public key of a peer manually:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter public key view	**public-key peer** *keyname*	—
Enter public key code view	**public-key-code begin**	—
Configure a public key of the peer	Enter the key	Required Spaces and carriage returns are allowed between characters.
Return to public key view	**public-key-code end**	— When you exit public key code view, the system automatically saves the public key.
Return to system view	**peer-public-key end**	—

Follow these steps to import the host public key of a peer from the public key file:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Import the host public key of a peer from the public key file	**public-key peer** *keyname* **import sshkey** *filename*	Required

Displaying and Maintaining Public Keys

To do…	Use the command…	Remarks
Display the public keys of the local key pairs	**display public-key local** { **dsa** | **rsa** } **public**	Available in any view
Display the public keys of the peers	**display public-key peer** [**brief** | **name** *publickey-name*]	

Public Key Configuration Examples

Configuring the Public Key of a Peer Manually

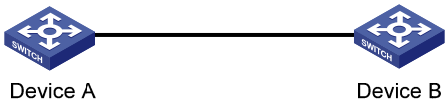
Network requirements

Device A is authenticated by Device B when accessing Device B, so the public key of Device A should be configured on Device B in advance.

In this example:

- RSA is used.
- The host public key of Device A is configured manually on Device B.

Figure 1-2 Network diagram for manually configuring the public key of a peer



Configuration procedure

1) Configure Device A

Create RSA key pairs on Device A.

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits of the modulus[default = 1024]:
Generating Keys...
++++++
++++++
++++++++
++++++++
```

Display the public keys of the created RSA key pairs.

```
[DeviceA] display public-key local rsa public

=====================================================
Time of Key pair created: 09:50:06  2007/08/07
Key name: HOST_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
```

```
=====================================================
Time of Key pair created: 09:50:07  2007/08/07
Key name: SERVER_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
307C300D06092A864886F70D0101010500036B0030680261009990 89E7AEE9802002D9EB2D0433B87BB6158E
35000AFB3FF310E42F109829D65BF70F7712507BE1A3E0BC5C2C03FAAF00DFDDC63D004B4490DACBA3CFA9E8
4B9151BDC7EECE1C8770D961557D192DE2B36CAF9974B7B293363BB372771C2C1F0203010001
```

2) Configure Device B

Configure the host public key of Device A on Device B. In public key code view, input the host public key of Device A. The host public key is the content of HOST_KEY displayed on Device A using the **display public-key local dsa public** command.

```
<DeviceB> system-view
[DeviceB] public-key peer devicea
Public key view: return to System View with "peer-public-key end".
[DeviceB-pkey-public-key] public-key-code begin
Public key code view: return to last view with "public-key-code end".
[DeviceB-pkey-key-code]30819F300D06092A864886F70D010101050003818D0030818902818100D90003F
A95F5A44A2A2CD3F814F9854C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A
9AB16C9E766BD995C669A784AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326
470034DC078B2BAA3BC3BCA80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
[DeviceB-pkey-key-code] public-key-code end
[DeviceB-pkey-public-key] peer-public-key end
```

Display the host public key of Device A saved on Device B.

```
[DeviceB] display public-key peer name devicea


=====================================
  Key Name  : devicea
  Key Type  : RSA
  Key Module: 1024
=====================================
Key Code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
```

Importing the Public Key of a Peer from a Public Key File

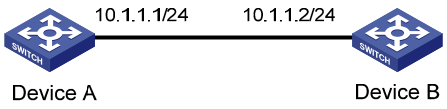
Network requirements

Device A is authenticated when accessing Device B, so the public host public key of Device A should be configured on Device B in advance.

In this example:

- RSA is used.
- The host public key of Device A is imported from the public key file to Device B.

Figure 1-3 Network diagram for importing the public key of a peer from a public key file



Configurtion procedure

1) Create key pairs on Device A and export the host public key

Create RSA key pairs on Device A.

```
<DeviceA> system-view
[DeviceA] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
It will take a few minutes.
Press CTRL+C to abort.
Input the bits of the modulus[default = 1024]:
Generating Keys...
++++++
++++++
++++++++
++++++++
```

Display the public keys of the created RSA key pairs.

```
[DeviceA] display public-key local rsa public

=====================================================
Time of Key pair created: 09:50:06  2007/08/07
Key name: HOST_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001

=====================================================
Time of Key pair created: 09:50:07  2007/08/07
Key name: SERVER_KEY
Key type: RSA Encryption Key
=====================================================
Key code:
307C300D06092A864886F70D0101010500036B0030680261009990 89E7AEE9802002D9EB2D0433B87BB6158E
35000AFB3FF310E42F109829D65BF70F7712507BE1A3E0BC5C2C03FAAF00DFDDC63D004B4490DACBA3CFA9E8
4B9151BDC7EECE1C8770D961557D192DE2B36CAF9974B7B293363BB372771C2C1F0203010001
```

Export the RSA host public key to a file named **devicea.pub**.

```
[DeviceA] public-key local export rsa ssh2 devicea.pub
```

```
[DeviceA] quit
```

2) Enable the FTP server function on Device B

Enable the FTP server function, create an FTP user with the username **ftp** and password **123**.

```
<DeviceB> system-view
[DeviceB] ftp server enable
[DeviceB] local-user ftp
[DeviceB-luser-ftp] password simple 123
[DeviceB-luser-ftp] service-type ftp
[DeviceB-luser-ftp] authorization-attribute level 3
[DeviceB-luser-ftp] quit
```

3) Upload the public key file of Device A to Device B

FTP the public key file **devicea.pub** to Device B.

```
<DeviceA> ftp 10.1.1.2
Trying 10.1.1.2 ...
Press CTRL+K to abort
Connected to 10.1.1.2.
220 FTP service ready.
User(10.1.1.2:(none)):ftp
331 Password required for ftp.
Password:
230 User logged in.
[ftp] put devicea.pub
227 Entering Passive Mode (10,1,1,2,5,148).
125 ASCII mode data connection already open, transfer starting for /devicea.pub.
226 Transfer complete.
FTP: 299 byte(s) sent in 0.189 second(s), 1.00Kbyte(s)/sec.
```

4) Import the host public key of Device A to Device B

Import the host public key of Device A from the key file **devicea.pub** to Device B.

```
[DeviceB] public-key peer devicea import sshkey devicea.pub
```

Display the host public key of Device A saved on Device B.

```
[DeviceB] display public-key peer name devicea

=====================================
  Key Name  : devicea
  Key Type  : RSA
  Key Module: 1024
=====================================
Key Code:
30819F300D06092A864886F70D010101050003818D0030818902818100D90003FA95F5A44A2A2CD3F814F985
4C4421B57CAC64CFFE4782A87B0360B600497D87162D1F398E6E5E51E5E353B3A9AB16C9E766BD995C669A78
4AD597D0FB3AA9F7202C507072B19C3C50A0D7AD3994E14ABC62DB125035EA326470034DC078B2BAA3BC3BCA
80AAB5EE01986BD1EF64B42F17CCAE4A77F1EF999B2BF9C4A10203010001
```

Table of Contents

1 ACL Overview

In order to filter traffic, network devices use sets of rules, called access control lists (ACLs), to identify and handle packets.

When configuring ACLs, go to these chapters for information you are interested in

- ACL Overview
- IPv4 ACL Configuration
- IPv6 ACL Configuration

Note

Unless otherwise stated, ACLs refer to both IPv4 ACLs and IPv6 ACLs throughout this document.

Introduction to ACL

Introduction

As network scale and network traffic are increasingly growing, network security and bandwidth allocation become more and more critical to network management. Packet filtering can be used to efficiently prevent illegal users from accessing networks and to control network traffic and save network resources. Access control lists (ACL) are often used to filter packets with configured matching rules.

ACLs are sets of rules (or sets of permit or deny statements) that decide what packets can pass and what should be rejected based on matching criteria such as source MAC address, destination MAC address, source IP address, destination IP address, and port number.

Application of ACLs on the Switch

The switch supports two ACL application modes:

- Hardware-based application: An ACL is assigned to a piece of hardware. For example, an ACL can be referenced by QoS for traffic classification. Note that when an ACL is referenced to implement QoS, the actions defined in the ACL rules, deny or permit, do not take effect; actions to be taken on packets matching the ACL depend on the traffic behavior definition in QoS. For details about traffic behavior, refer to the QoS part in this manual.
- Software-based application: An ACL is referenced by a piece of upper layer software. For example, an ACL can be referenced to configure login user control behavior, thus controlling Telnet, SNMP and Web users. Note that when an ACL is reference by the upper layer software, actions to be taken on packets matching the ACL depend on those defined by the ACL rules. For details about login user control, refer to the part about login configuration in this manual.

Note

- When an ACL is assigned to a piece of hardware and referenced by a QoS policy for traffic classification, the switch does not take action according to the traffic behavior definition on a packet that does not match the ACL.
- When an ACL is referenced by a piece of software to control Telnet, SNMP, and Web login users, the switch denies all packets that do not match the ACL.

Introduction to IPv4 ACL

This section covers these topics:

IPv4 ACL Classification

IPv4 ACLs, identified by ACL numbers, fall into three categories, as shown in Table 1-1.

Table 1-1 IPv4 ACL categories

Category	ACL number	Matching criteria
Basic IPv4 ACL	2000 to 2999	Source IP address
Advanced IPv4 ACL	3000 to 3999	Source IP address, destination IP address, protocol carried over IP, and other Layer 3 or Layer 4 protocol header information
Ethernet frame header ACL	4000 to 4999	Layer 2 protocol header fields such as source MAC address, destination MAC address, 802.1p priority, and link layer protocol type

IPv4 ACL Naming

When creating an IPv4 ACL, you can specify a unique name for it. Afterwards, you can identify the ACL by its name.

An IPv4 ACL can have only one name. Whether to specify a name for an ACL is up to you. After creating an ACL, you cannot specify a name for it, nor can you change or remove its name.

Note

The name of an IPv4 ACL must be unique among IPv4 ACLs. However, an IPv4 ACL and an IPv6 ACL can share the same name.

IPv4 ACL Match Order

An ACL may consist of multiple rules, which specify different matching criteria. These criteria may have overlapping or conflicting parts. The match order is for determining how packets should be matched against the rules.

Two match orders are available for IPv4 ACLs:

- **config**: Packets are compared against ACL rules in the order the rules are configured.
- **auto**: Packets are compared against ACL rules in the depth-first match order.

The term depth-first match has different meanings for different types of ACLs:

Depth-first match for a basic IPv4 ACL

The following shows how your device performs depth-first match in a basic IPv4 ACL:

1) Sort rules by VPN instance first and compare packets against the rule configured with a VPN instance.
2) In case of a tie, sort rules by source IP address wildcard and compare packets against the rule configured with more zeros in the source IP address wildcard.
3) If two rules are present with the same number of zeros in their source IP address wildcards, compare packets against the rule configured first.

Depth-first match for an advanced IPv4 ACL

The following shows how your device performs depth-first match in an advanced IPv4 ACL:

1) Sort rules by VPN instance first and compare packets against the rule configured with a VPN instance.
2) In case of a tie, look at the protocol carried over IP. A rule with no limit to the protocol type (that is, configured with the **ip** keyword) has the lowest precedence. Rules each of which has a single specified protocol type are of the same precedence level.
3) If the protocol types have the same precedence, look at the source IP address wildcards. Then, compare packets against the rule configured with more zeros in the source IP address wildcard.
4) If the numbers of zeros in the source IP address wildcards are the same, look at the destination IP address wildcards. Then, compare packets against the rule configured with more zeros in the destination IP address wildcard.
5) If the numbers of zeros in the destination IP address wildcards are the same, look at the Layer 4 port number ranges, namely the TCP/UDP port number ranges. Then compare packets against the rule configured with the smaller port number range.
6) If the port number ranges are the same, compare packets against the rule configured first.

Depth-first match for an Ethernet frame header ACL

The following shows how your device performs depth-first match in an Ethernet frame header ACL:

1) Sort rules by source MAC address mask first and compare packets against the rule configured with more ones in the source MAC address mask.
2) If two rules are present with the same number of ones in their source MAC address masks, look at the destination MAC address masks. Then, compare packets against the rule configured with more ones in the destination MAC address mask.
3) If the numbers of ones in the destination MAC address masks are the same, compare packets against the one configured first.

The comparison of a packet against ACL rules stops immediately after a match is found. The packet is then processed as per the rule.

IPv4 ACL Step

Meaning of the step

The step defines the difference between two neighboring numbers that are automatically assigned to ACL rules by the device. For example, with a step of 5, rules are automatically numbered 0, 5, 10, 15, and so on. By default, the step is 5.

Whenever the step changes, the rules are renumbered, starting from 0. For example, if four rules are numbered 5, 10, 15, and 20 respectively, changing the step from 5 to 2 will cause the rules to be renumbered 0, 2, 4, and 6.

Benefits of using the step

With the step and rule numbering/renumbering mechanism, you do not need to assign numbers to rules when defining them. The system will assign a newly defined rule a number that is the smallest multiple of the step bigger than the current biggest number. For example, with a step of five, if the biggest number is currently 28, the newly defined rule will get a number of 30. If the ACL has no rule defined already, the first defined rule will get a number of 0.

Another benefit of using the step is that it allows you to insert new rules between existing ones as needed. For example, after creating four rules numbered 0, 5, 10, and 15 in an ACL with a step of five, you can insert a rule numbered 1.

Effective Period of an IPv4 ACL

You can control when a rule can take effect by referencing a time range in the rule.

A referenced time range can be one that has not been created yet. The rule, however, can take effect only after the time range is defined and becomes active.

IP Fragments Filtering with IPv4 ACL

Traditional packet filtering performs match operation on, rather than all IP fragments, the first ones only. All subsequent non-first fragments are handled in the way the first fragments are handled. This causes security risk as attackers may fabricate non-first fragments to attack your network.

As for the configuration of a rule of an IPv4 ACL, the **fragment** keyword specifies that the rule applies to non-first fragment packets only, and does not apply to non-fragment packets or the first fragment packets. ACL rules that do not contain this keyword is applicable to both non-fragment packets and fragment packets.

Introduction to IPv6 ACL

This section covers these topics:

IPv6 ACL Classification

IPv6 ACLs, identified by ACL numbers, fall into three categories, as shown in Table 1-2.

Table 1-2 IPv6 ACL categories

Category	ACL number	Matching criteria
Basic IPv6 ACL	2000 to 2999	Source IPv6 address
Advanced IPv6 ACL	3000 to 3999	Source IPv6 address, destination IPv6 address, protocol carried over IPv6, and other Layer 3 or Layer 4 protocol header information

IPv6 ACL Naming

When creating an IPv6 ACL, you can specify a unique name for it. Afterwards, you can identify the IPv6 ACL by its name.

An IPv6 ACL can have only one name. Whether to specify a name for an ACL is up to you. After creating an ACL, you cannot specify a name for it, nor can you change or remove its name.

Note

The name of an IPv6 ACL must be unique among IPv6 ACLs. However, an IPv6 ACL and an IPv4 ACL can share the same name.

IPv6 ACL Match Order

Similar to IPv4 ACLs, an IPv6 ACL consists of multiple rules, each of which specifies different matching criteria. These criteria may have overlapping or conflicting parts. The match order is for determining how a packet should be matched against the rules.

Two match orders are available for IPv6 ACLs:

- **config**: Packets are compared against ACL rules in the order the rules are configured.
- **auto**: Packets are compared against ACL rules in the depth-first match order.

The term depth-first match has different meanings for different types of IPv6 ACLs:

Depth-first match for a basic IPv6 ACL

The following shows how your device performs depth-first match in a basic IPv6 ACL:

1) Sort rules by source IPv6 address prefix first and compare packets against the rule configured with a longer prefix for the source IPv6 address.
2) In case of a tie, compare packets against the rule configured first.

Depth-first match for an advanced IPv6 ACL

The following shows how your device performs depth-first match in an advanced IPv6 ACL:

1) Look at the protocol type field in the rules first. A rule with no limit to the protocol type (that is, configured with the **ipv6** keyword) has the lowest precedence. Rules each of which has a single specified protocol type are of the same precedence level. Compare packets against the rule with the highest precedence.
2) In case of a tie, look at the source IPv6 address prefixes. Then, compare packets against the rule configured with a longer prefix for the source IPv6 address.
3) If the prefix lengths for the source IPv6 addresses are the same, look at the destination IPv6 address prefixes. Then, compare packets against the rule configured with a longer prefix for the destination IPv6 address.
4) If the prefix lengths for the destination IPv6 addresses are the same, look at the Layer 4 port number ranges, namely the TCP/UDP port number ranges. Then compare packets against the rule configured with the smaller port number range.
5) If the port number ranges are the same, compare packets against the rule configured first.

The comparison of a packet against an ACL stops immediately after a match is found. The packet is then processed as per the rule.

IPv6 ACL Step

Refer to IPv4 ACL Step.

Effective Period of an IPv6 ACL

Refer to Effective Period of an IPv4 ACL.

2 IPv4 ACL Configuration

When configuring an IPv4 ACL, go to these sections for information you are interested in:

Creating a Time Range

Two types of time ranges are available:

- Periodic time range, which recurs periodically on the day or days of the week.
- Absolute time range, which takes effect only in a period of time and does not recur.

Configuration Procedure

Follow these steps to create a time range:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a time range	**time-range** *time-range-name* { *start-time* **to** *end-time days* [**from** *time1 date1*] [**to** *time2 date2*] | **from** *time1 date1* [**to** *time2 date2*] | **to** *time2 date2* }	Required
Display the configuration and status of one or all time ranges	**display time-range** { *time-range-name* | **all** }	Optional Available in any view

You may create a maximum of 256 time ranges.

A time range can be one of the following:

- Periodic time range created using the **time-range** *time-range-name start-time* **to** *end-time days* command. A time range thus created recurs periodically on the day or days of the week. A periodic time range is active only when the system time falls within it.
- Absolute time range created using the **time-range** *time-range-name* { **from** *time1 date1* [**to** *time2 date2*] \| **to** *time2 date2* } command. Unlike a periodic time range, a time range thus created does not recur. For example, to create an absolute time range that is active between January 1, 2004 00:00 and December 31, 2004 23:59, you may use the **time-range test from 00:00 01/01/2004 to 23:59 12/31/2004** command.
- Compound time range created using the **time-range** *time-range-name start-time* **to** *end-time days* { **from** *time1 date1* [**to** *time2 date2*] \| **to** *time2 date2* } command. A time range thus created recurs

on the day or days of the week only within the specified period. For example, to create a time range that is active from 12:00 to 14:00 on Wednesdays between January 1, 2004 00:00 and December 31, 2004 23:59, you may use the **time-range test 12:00 to 14:00 wednesday from 00:00 01/01/2004 to 23:59 12/31/2004** command.

- You may create individual time ranges identified with the same name. They are regarded as one time range whose active period is the result of ORing periodic ones, ORing absolute ones, and ANDing periodic and absolute ones.
- If you do not specify the start time and date, the time range starts from the earliest time that the system supports, namely 00:00 01/01/1970. If you do not specify the end time and date, the time range ends at the latest time that the system supports, namely 24:00 12/31/2100.

Configuration Example

Create a time range that is active from 8:00 to 18:00 every working day.

```
<Sysname> system-view
[Sysname] time-range test 8:00 to 18:00 working-day
```

Verify the configuration.

```
[Sysname] display time-range test
Current time is 22:17:42 1/5/2006 Thursday

Time-range : test ( Inactive )
 08:00 to 18:00 working-day
```

Create an absolute time range from 15:00, Jan 28, 2006 to 15:00, Jan 28, 2008.

```
<Sysname> system-view
[Sysname] time-range test from 15:00 1/28/2006 to 15:00 1/28/2008
[Sysname] display time-range test
Current time is 22:20:18 1/5/2006 Thursday

Time-range : test ( Inactive )
 from 15:00 1/28/2006 to 15:00 1/28/2008
```

Configuring a Basic IPv4 ACL

Basic IPv4 ACLs match packets based on only source IP address. They are numbered from 2000 to 2999.

Configuration Prerequisites

If you want to reference a time range in a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure a basic IPv4 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a basic IPv4 ACL and enter its view	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]	Required The default match order is **config**. If you specify a name for an IPv4 ACL when creating the ACL, you can use the **acl name** *acl-name* command to enter the view of the ACL later.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } [**fragment** | **logging** | **source** { *sour-addr sour-wildcard* | **any** } | **time-range** *time-range-name* | **vpn-instance** *vpn-instance-name*] *	Required To create or modify multiple rules, repeat this step. Note that the **logging** keyword is not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set the rule numbering step	**step** *step-value*	Optional 5 by default
Configure a description for the basic IPv4 ACL	**description** *text*	Optional By default, a basic IPv4 ACL has no ACL description.
Configure a rule description	**rule** *rule-id* **comment** *text*	Optional By default, an IPv4 ACL rule has no rule description.

Note that:

- You can only modify the existing rules of an ACL that uses the match order of **config**. When modifying a rule of such an ACL, you may choose to change just some of the settings, in which case the other settings remain the same.
- You cannot create a rule with, or modify a rule to have, the same permit/deny statement as an existing rule in the ACL.
- When the ACL match order is **auto**, a newly created rule will be inserted among the existing rules in the depth-first match order. Note that the IDs of the rules still remain the same.

Caution

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command, but only when the ACL does not contain any rules.
- The rule specified in the **rule comment** command must already exist.

Configuration Example

Configure IPv4 ACL 2000 to deny packets with source address 1.1.1.1.

```
<Sysname> system-view
[Sysname] acl number 2000
[Sysname-acl-basic-2000] rule deny source 1.1.1.1 0
```

Verify the configuration.

```
[Sysname-acl-basic-2000] display acl 2000
Basic ACL  2000, named -none-, 1 rule,
ACL's step is 5
 rule 0 deny source 1.1.1.1 0 (5 times matched)
```

Configuring an Advanced IPv4 ACL

Advanced IPv4 ACLs match packets based on source IP address, destination IP address, protocol carried over IP, and other protocol header fields, such as the TCP/UDP source port number, TCP/UDP destination port number, TCP flag, ICMP message type, and ICMP message code.

In addition, advanced IPv4 ACLs allow you to filter packets based on three priority criteria: type of service (ToS), IP precedence, and differentiated services codepoint (DSCP) priority.

Advanced IPv4 ACLs are numbered in the range 3000 to 3999. Compared with basic IPv4 ACLs, they allow of more flexible and accurate filtering.

Configuration Prerequisites

If you want to reference a time range in a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an advanced IPv4 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an advanced IPv4 ACL and enter its view	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]	Required The default match order is **config**. If you specify a name for an IPv4 ACL when creating the ACL, you can use the **acl name** *acl-name* command to enter the view of the ACL later.

To do…	Use the command…	Remarks
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [{ **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **destination** { *dest-addr dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **fragment** | **icmp-type** { *icmp-type icmp-code* | *icmp-message* } | **logging** | **precedence** *precedence* | **reflective** | **source** { *sour-addr sour-wildcard* | **any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name* | **tos** *tos* | **vpn-instance** *vpn-instance-name*] *	Required To create or modify multiple rules, repeat this step. Note that if the ACL is to be referenced by a QoS policy for traffic classification, the **logging** and **reflective** keywords are not supported and the *operator* argument cannot be: • **neq**, if the policy is for the inbound traffic, • **gt**, **lt**, **neq** or **range**, if the policy is for the outbound traffic.
Set the rule numbering step	**step** *step-value*	Optional 5 by default
Configure a description for the advanced IPv4 ACL	**description** *text*	Optional By default, an advanced IPv4 ACL has no ACL description.
Configure a rule description	**rule** *rule-id* **comment** *text*	Optional By default, an IPv4 ACL rule has no rule description.

Note that:

- You can only modify the existing rules of an ACL that uses the match order of **config**. When modifying a rule of such an ACL, you may choose to change just some of the settings, in which case the other settings remain the same.
- You cannot create a rule with, or modify a rule to have, the same permit/deny statement as an existing rule in the ACL.
- When the ACL match order is **auto**, a newly created rule will be inserted among the existing rules in the depth-first match order. Note that the IDs of the rules still remain the same.

Caution

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command, but only when the ACL does not contain any rules.
- The rule specified in the **rule comment** command must already exist.

Configuration Example

Configure IPv4 ACL 3000 to permit TCP packets with the destination port number of 80 from 129.9.0.0 to 202.38.160.0.

```
<Sysname> system-view
```

```
[Sysname] acl number 3000
[Sysname-acl-adv-3000] rule permit tcp source 129.9.0.0 0.0.255.255 destination 202.38.160.0
0.0.0.255 destination-port eq 80
```

Verify the configuration.

```
[Sysname-acl-adv-3000] display acl 3000
Advanced ACL  3000, named -none-, 1 rule,
ACL's step is 5
 rule  0  permit  tcp  source  129.9.0.0  0.0.255.255  destination  202.38.160.0  0.0.0.255
destination-port eq www (5 times matched)
```

Configuring an Ethernet Frame Header ACL

Ethernet frame header ACLs match packets based on Layer 2 protocol header fields such as source MAC address, destination MAC address, 802.1p priority (VLAN priority), and link layer protocol type. They are numbered in the range 4000 to 4999.

Configuration Prerequisites

If you want to reference a time range in a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an Ethernet frame header ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an Ethernet frame header ACL and enter its view	**acl number** *acl-number* [**name** *acl-name*] [**match-order** { **auto** | **config** }]	Required The default match order is **config**. If you specify a name for an IPv4 ACL when creating the ACL, you can use the **acl name** *acl-name* command to enter the view of the ACL later.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } [**cos** *vlan-pri* | **dest-mac** *dest-addr dest-mask* | **lsap** *lsap-code lsap-wildcard* | **source-mac** *sour-addr source-mask* | **time-range** *time-range-name* | **type** *type-code type-wildcard*] *	Required To create or modify multiple rules, repeat this step. Note that the **lsap** keyword is not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set the rule numbering step	**step** *step-value*	Optional 5 by default
Configure a description for the Ethernet frame header ACL	**description** *text*	Optional By default, an Ethernet frame header ACL has no ACL description.
Configure a rule description	**rule** *rule-id* **comment** *text*	Optional By default, an Ethernet frame header ACL rule has no rule description.

Note that:

- You can only modify the existing rules of an ACL that uses the match order of **config**. When modifying a rule of such an ACL, you may choose to change just some of the settings, in which case the other settings remain the same.
- You cannot create a rule with, or modify a rule to have, the same permit/deny statement as an existing rule in the ACL.
- When the ACL match order is **auto**, a newly created rule will be inserted among the existing rules in the depth-first match order. Note that the IDs of the rules still remain the same.

 **Caution**

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command, but only when the ACL does not contain any rules.
- The rule specified in the **rule comment** command must already exist.

Configuration Example

Configure ACL 4000 to deny frames with the 802.1p priority of 3.

```
<Sysname> system-view
[Sysname] acl number 4000
[Sysname-acl-ethernetframe-4000] rule deny cos 3
```

Verify the configuration.

```
[Sysname-acl-ethernetframe-4000] display acl 4000
Ethernet frame ACL  4000, named -none-, 1 rule,
ACL's step is 5
rule 0 deny cos excellent-effort(5 times matched)
```

Copying an IPv4 ACL

This feature allows you to copy an existing IPv4 ACL to generate a new one, which is of the same type and has the same match order, rules, rule numbering step and descriptions as the source IPv4 ACL.

Configuration Prerequisites

Make sure that the source IPv4 ACL exists while the destination IPv4 ACL does not.

Configuration Procedure

Follow these steps to copy an IPv4 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Copy an existing IPv4 ACL to generate a new one of the same type	**acl copy** { *source-acl-number* | **name** *source-acl-name* } **to** { *dest-acl-number* | **name** *dest-acl-name* }	Required

Caution

- The source IPv4 ACL and the destination IPv4 ACL must be of the same type.
- The destination ACL does not take the name of the source IPv4 ACL.

Displaying and Maintaining IPv4 ACLs

To do...	Use the command…	Remarks
Display information about one or all IPv4 ACLs	**display acl** { *acl-number* | **all** | **name** *acl-name* }	Available in any view
Display information about ACL uses of a switch	**display acl resource**	Available in any view
Display the configuration and state of a specified or all time ranges	**display time-range** { *time-range-name* | **all** }	Available in any view
Clear statistics about a specified or all IPv4 ACLs that are referenced by upper layer software	**reset acl counter** { *acl-number* | **all** | **name** *acl-name* }	Available in user view

IPv4 ACL Configuration Example

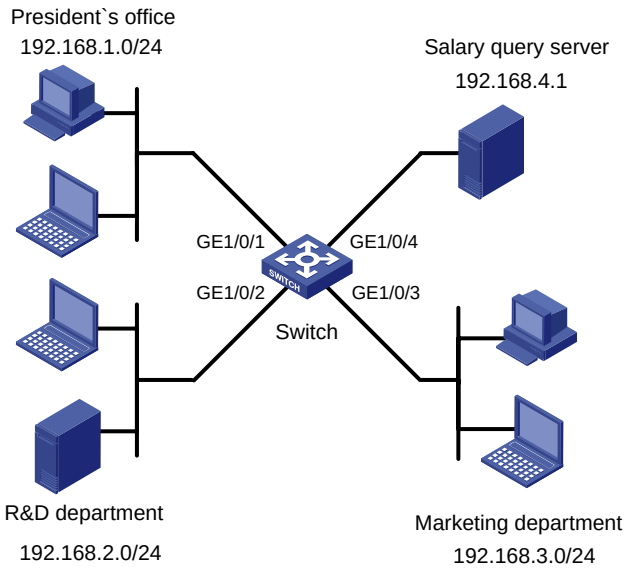
Network Requirements

As shown in Figure 2-1, a company interconnects its departments through the switch.

Configure an ACL to deny access of all departments but the President's office to the salary query server during office hours (from 8:00 to 18:00) in working days.

Network Diagram

Figure 2-1 Network diagram for IPv4 ACL configuration



Configuration Procedure

1) Create a time range for office hours

Create a periodic time range spanning 8:00 to 18:00 in working days.

```
<Switch> system-view
[Switch] time-range trname 8:00 to 18:00 working-day
```

2) Define an ACL to control access to the salary query server

Configure a rule to control access of the R&D Department to the salary query server.

```
[Switch] acl number 3000
[Switch-acl-adv-3000] rule deny ip source 192.168.2.0 0.0.0.255 destination 192.168.4.1
0.0.0.0 time-range trname
[Switch-acl-adv-3000] quit
```

Configure a rule to control access of the Marketing Department to the salary query server.

```
[Switch] acl number 3001
[Switch-acl-adv-3001] rule deny ip source 192.168.3.0 0.0.0.255 destination 192.168.4.1
0.0.0.0 time-range trname
[Switch-acl-adv-3001] quit
```

3) Apply the IPv4 ACL

Configure class c_rd for packets matching IPv4 ACL 3000.

```
[Switch] traffic classifier c_rd
[Switch-classifier-c_rd] if-match acl 3000
[Switch-classifier-c_rd] quit
```

Configure traffic behavior b_rd to deny matching packets.

```
[Switch] traffic behavior b_rd
[Switch-behavior-b_rd] filter deny
[Switch-behavior-b_rd] quit
```

Configure class c_market for packets matching IPv4 ACL 3001.

```
[Switch] traffic classifier c_market
[Switch-classifier-c_market] if-match acl 3001
[Switch-classifier-c_market] quit
```

Configure traffic behavior b_ market to deny matching packets.

```
[Switch] traffic behavior b_market
[Switch-behavior-b_market] filter deny
[Switch-behavior-b_market] quit
```

Configure QoS policy p_rd to use traffic behavior b_rd for class c_rd.

```
[Switch] qos policy p_rd
[Switch-qospolicy-p_rd] classifier c_rd behavior b_rd
[Switch-qospolicy-p_rd] quit
```

Configure QoS policy p_market to use traffic behavior b_market for class c_market.

```
[Switch] qos policy p_market
[Switch-qospolicy-p_market] classifier c_market behavior b_market
[Switch-qospolicy-p_market] quit
```

Apply QoS policy p_rd to interface GigabitEthernet 1/0/2.

```
[Switch] interface GigabitEthernet 1/0/2
[Switch-GigabitEthernet1/0/2] qos apply policy p_rd inbound
[Switch-GigabitEthernet1/0/2] quit
```

Apply QoS policy p_market to interface GigabitEthernet 1/0/3.

```
[Switch] interface GigabitEthernet 1/0/3
[Switch-GigabitEthernet1/0/3] qos apply policy p_market inbound
```

3 IPv6 ACL Configuration

When configuring IPv6 ACLs, go to these sections for information you are interested in:

Creating a Time Range

Refer to Creating a Time Range.

Configuring a Basic IPv6 ACL

Basic IPv6 ACLs match packets based on only source IPv6 address. They are numbered in the range 2000 to 2999.

Configuration Prerequisites

If you want to reference a time range in a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an IPv6 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a basic IPv6 ACL view and enter its view	**acl ipv6 number** *acl6-number* [**name** *acl6-name*] [**match-order** { **auto** | **config** }]	Required The default match order is **config**. If you specify a name for an IPv6 ACL when creating the ACL, you can use the **acl ipv6 name** *acl6-name* command to enter the view of the ACL later.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } [**fragment** | **logging** | **source** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* | **any** } | **time-range** *time-range-name*] *	Required To create or modify multiple rules, repeat this step. Note that the **logging** and **fragment** keywords are not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set the rule numbering step	**step** *step-value*	Optional 5 by default

To do…	Use the command…	Remarks
Configure a description for the basic IPv6 ACL	**description** *text*	Optional By default, a basic IPv6 ACL has no ACL description.
Configure a rule description	**rule** *rule-id* **comment** *text*	Optional By default, an IPv6 ACL rule has no rule description.

Note that:

- You can only modify the existing rules of an ACL that uses the match order of **config**. When modifying a rule of such an ACL, you may choose to change just some of the settings, in which case the other settings remain the same.
- You cannot create a rule with, or modify a rule to have, the same permit/deny statement as an existing rule in the ACL.
- When the ACL match order is **auto**, a newly created rule will be inserted among the existing rules in the depth-first match order. Note that the IDs of the rules still remain the same.

Caution

- You can modify the match order of an IPv6 ACL with the **acl ipv6 number** *acl6-number* [**name** *acl6-name*] **match-order** { **auto** | **config** } command, but only when the ACL does not contain any rules.
- The rule specified in the **rule comment** command must already exist.

Configuration Example

Configure IPv6 ACL 2000 to permit IPv6 packets with the source address of 2030:5060::9050/64 and deny IPv6 packets with the source address of fe80:5060::8050/96.

```
<Sysname> system-view
[Sysname] acl ipv6 number 2000
[Sysname-acl6-basic-2000] rule permit source 2030:5060::9050/64
[Sysname-acl6-basic-2000] rule deny source fe80:5060::8050/96
```

Verify the configuration.

```
[Sysname-acl6-basic-2000] display acl ipv6 2000
 Basic IPv6 ACL  2000, named -none-, 2 rules,
 ACL's step is 5
 rule 0 permit source 2030:5060::9050/64 (4 times matched)
 rule 5 deny source FE80:5060::8050/96 (5 times matched)
```

Configuring an Advanced IPv6 ACL

Advanced IPv6 ACLs match packets based on the source IPv6 address, destination IPv6 address, protocol carried over IPv6, and other protocol header fields such as the TCP/UDP source port number, TCP/UDP destination port number, ICMP message type, and ICMP message code.

Advanced IPv6 ACLs are numbered in the range 3000 to 3999. Compared with basic IPv6 ACLs, they allow of more flexible and accurate filtering.

Configuration Prerequisites

If you want to reference a time range in a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an advanced IPv6 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an advanced IPv6 ACL and enter its view	**acl ipv6 number** *acl6-number* [**name** *acl6-name*] [**match-order** { **auto** | **config** }]	Required The default match order is **config**. If you specify a name for an IPv6 ACL when creating the ACL, you can use the **acl ipv6 name** *acl6-name* command to enter the view of the ACL later.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [{ **ack** *ack-value* | **fin** *fin-value* | **psh** *psh-value* | **rst** *rst-value* | **syn** *syn-value* | **urg** *urg-value* } * | **destination** { *dest dest-prefix* | *dest/dest-prefix* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **fragment** | **icmpv6-type** { *icmpv6-type icmpv6-code* | *icmpv6-message* } | **logging** | **source** { *source source-prefix* | *source/source-prefix* **| any** } | **source-port** *operator port1* [*port2*] | **time-range** *time-range-name*] *	Required To create or modify multiple rules, repeat this step. Note that if the ACL is to be referenced by a QoS policy for traffic classification, the **logging** and **fragment** keywords are not supported and the *operator* argument cannot be: • **neq**, if the policy is for the inbound traffic, • **gt**, **lt**, **neq** or **range**, if the policy is for the outbound traffic.
Set the rule numbering step	**step** *step-value*	Optional 5 by default
Configure a description for the advanced IPv6 ACL	**description** *text*	Optional By default, an advanced IPv6 ACL has no ACL description.
Configure a rule description	**rule** *rule-id* **comment** *text*	Optional By default, an IPv6 ACL rule has no rule description.

Note that:

- You can only modify the existing rules of an ACL that uses the match order of **config**. When modifying a rule of such an ACL, you may choose to change just some of the settings, in which case the other settings remain the same.
- You cannot create a rule with, or modify a rule to have, the same permit/deny statement as an existing rule in the ACL.

- When the ACL match order is **auto**, a newly created rule will be inserted among the existing rules in the depth-first match order. Note that the IDs of the rules still remain the same.

Caution

- You can modify the match order of an IPv6 ACL with the **acl ipv6 number** *acl6-number* [**name** *acl6-name*] **match-order** { **auto** | **config** } command, but only when the ACL does not contain any rules.
- The rule specified in the **rule comment** command must already exist.

Configuration Example

Configure IPv6 ACL 3000 to permit TCP packets with the source address of 2030:5060::9050/64.

```
<Sysname> system-view
[Sysname] acl ipv6 number 3000
[Sysname-acl6-adv-3000] rule permit tcp source 2030:5060::9050/64
```

Verify the configuration.

```
[Sysname-acl6-adv-3000] display acl ipv6 3000
 Advanced IPv6 ACL  3000, named -none-, 1 rule,
 ACL's step is 5
 rule 0 permit tcp source 2030:5060::9050/64 (5 times matched)
```

Copying an IPv6 ACL

This feature allows you to copy an existing IPv6 ACL to generate a new one, which is of the same type and has the same match order, rules, rule numbering step, and descriptions as the source IPv6 ACL.

Configuration Prerequisites

Make sure that the source IPv6 ACL exists while the destination IPv6 ACL does not.

Configuration Procedure

Follow these steps to copy an IPv6 ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Copy an existing IPv6 ACL to generate a new one of the same type	**acl ipv6 copy** { *source-acl6-number* | **name** *source-acl6-name* } **to** { *dest-acl6-number* | **name** *dest-acl6-name* }	Required

Caution

- The source IPv6 ACL and the destination IPv6 ACL must be of the same type.
- The destination ACL does not take the name of the source IPv6 ACL.

Displaying and Maintaining IPv6 ACLs

To do…	Use the command…	Remarks
Display information about one or all IPv6 ACLs	**display acl ipv6** { *acl6-number* | **all** | **name** *acl6-name* }	Available in any view
Display information about ACL uses of a switch	**display acl resource**	Available in any view
Display the configuration and status on time range	**display time-range** { *time-range-name* | **all** }	Available in any view
Clear statistics about a specified or all IPv6 ACLs that are referenced by upper layer software	**reset acl ipv6 counter** { *acl6-number* | **all** | **name** *acl6-name* }	Available in user view

IPv6 ACL Configuration Example

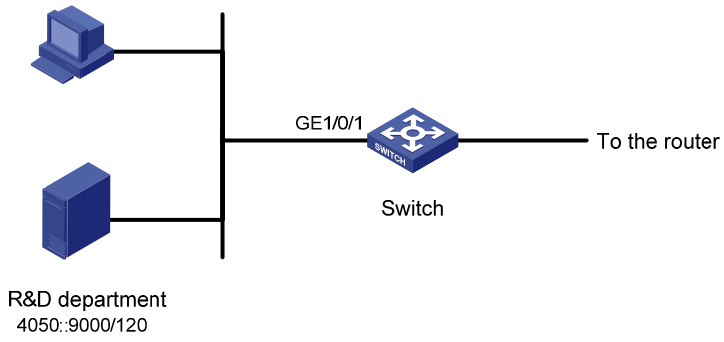
Network Requirements

As shown in Figure 3-1, a company interconnects its departments through the switch.

Configure an ACL to deny access of the R&D department to external networks.

Network Diagram

Figure 3-1 Network diagram for IPv6 ACL configuration



Configuration Procedure

Create an IPv6 ACL 2000.

```
<Switch> system-view
[Switch] acl ipv6 number 2000
[Switch-acl6-basic-2000] rule deny source 4050::9000/120
```

```
[Switch-acl6-basic-2000] quit
```

Configure class c_rd for packets matching IPv6 ACL 2000.

```
[Switch] traffic classifier c_rd
[Switch-classifier-c_rd] if-match acl ipv6 2000
[Switch-classifier-c_rd] quit
```

Configure traffic behavior b_rd to deny matching packets.

```
[Switch] traffic behavior b_rd
[Switch-behavior-b_rd] filter deny
[Switch-behavior-b_rd] quit
```

Configure QoS policy p_rd to use traffic behavior b_rd for class c_rd.

```
[Switch] qos policy p_rd
[Switch-qospolicy-p_rd] classifier c_rd behavior b_rd
[Switch-qospolicy-p_rd] quit
```

Apply QoS policy p_rd to interface GigabitEthernet 1/0/1.

```
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] qos apply policy p_rd inbound
```

System Volume Organization

Manual Version

6W100-20090120

Product Version

Release 2202

Organization

The System Volume is organized as follows:

Features	Description
Login	Upon logging into a device, you can configure user interface properties and manage the system conveniently. This document describes: • How to log in to your Ethernet switch • Introduction to the user interface and common configurations • Logging In Through the Console Port • Logging In Through Telnet • Logging in Through Web-based Network Management System • Logging In Through NMS • Specifying Source IP address/Interface for Telnet Packets • Controlling Login Users
Basic System Configuration	Basic system configuration involves the configuration of device name, system clock, welcome message, user privilege levels and so on. This document describes: • Configuration display • Basic configurations • CLI features

Features	Description
Device Management	Through the device management function, you can view the current condition of your device and configure running parameters. This document describes: • Device management overview • Rebooting a device • Configuring the scheduled automatic execution function • Specifying a file for the next device boot • Upgrading Boot ROM • Configuring a detection interval • Configuring temperature alarm thresholds for a board • Clearing the 16-bit interface indexes not used in the current system • Configuring the system load sharing function • Configuring the traffic forwarding mode of SRPUs • Configuring the working mode of EA LPUs • Enabling the port down function globally • Enabling expansion memory data recovery function on a board • Identifying and diagnosing pluggable transceivers
File System Management	A major function of the file system is to manage storage devices, mainly including creating the file system, creating, deleting, modifying and renaming a file or a directory and opening a file. This document describes: • File system management • Configuration File Management • FTP configuration • TFTP configuration
HTTP	Hypertext Transfer Protocol (HTTP) is used for transferring web page information across the Internet. This document describes: • HTTP Configuration • HTTPS Configuration
SNMP	Simple network management protocol (SNMP) offers a framework to monitor network devices through TCP/IP protocol suite. This document describes: • SNMP overview • Basic SNMP function configuration • SNMP log configuration • Trap configuration • MIB style configuration
RMON	RMON provides an efficient means of monitoring subnets and allows SNMP to monitor remote network devices in a more proactive and effective way. This document describes: • RMON overview • RMON configuration

Features	Description
MAC Address Table Management	A switch maintains a MAC address table for fast forwarding packets. This document describes: • MAC address table overview • Configuring MAC Address Entries • Disabling MAC Address Learning on a VLAN • Configuring MAC Address Aging Timer • Configuring the MAC Learning Limit • Configuring MAC Information
System Maintenance and Debugging	For the majority of protocols and features supported, the system provides corresponding debugging information to help users diagnose errors. This document describes: • Maintenance and debugging overview • Maintenance and debugging configuration
Information Center	As the system information hub, Information Center classifies and manages all types of system information. This document describes: • Information Center Overview • Setting to Output System Information to the Console • Setting to Output System Information to a Monitor Terminal • Setting to Output System Information to a Log Host • Setting to Output System Information to the Trap Buffer • Setting to Output System Information to the Log Buffer • Setting to Output System Information to the SNMP Module • Configuring Synchronous Information Output • Disabling a Port from Generating Link Up/Down Logging Information
PoE	The Power over Ethernet (PoE) feature enables the power sourcing equipment (PSE) to feed powered devices (PDs) from Ethernet ports through twisted pair cables. This document describes: • PoE overview • Configuring the PoE Interface • Configuring PoE power management • Configuring the PoE monitoring function • Online upgrading the PSE processing software • Configuring a PD Disconnection Detection Mode • Enabling the PSE to detect nonstandard PDs
Track	The track module is used to implement collaboration between different modules through established collaboration objects. The detection modules trigger the application modules to perform certain operations through the track module. This document describes: • Track Overview • Configuring Collaboration Between the Track Module and the Detection Modules • Configuring Collaboration Between the Track Module and the Application Modules

Features	Description
NQA	NQA analyzes network performance, services and service quality by sending test packets to provide you with network performance and service quality parameters. This document describes: • NQA Overview • Configuring the NQA Server • Enabling the NQA Client • Creating an NQA Test Group • Configuring an NQA Test Group • Configuring the Collaboration Function • Configuring Trap Delivery • Configuring the NQA Statistics Function • Configuring Optional Parameters Common to an NQA Test Group • Scheduling an NQA Test Group
NTP	Network Time Protocol (NTP) is the TCP/IP that advertises the accurate time throughout the network. This document describes: • NTP overview • Configuring the Operation Modes of NTP • Configuring Optional Parameters of NTP • Configuring Access-Control Rights • Configuring NTP Authentication
VRRP	Virtual Router Redundancy Protocol (VRRP) combines a group of switches (including a master and multiple backups) on a LAN into a virtual router called VRRP group. VRRP streamlines host configuration while providing high reliability. This document describes: • VRRP overview • IPv4-Based VRRP configuration • IPv6-Based VRRP configuration
Hotfix	Hotfix is a fast, cost-effective method to fix software defects of the device without interrupting the running services. This document describes: • Hotfix Overview • One-Step Patch Installation • Step-by-Step Patch Installation • Step-by-Step Patch Uninstallation • One-Step Patch Uninstallation
Cluster Management	A cluster is a group of network devices. Cluster management is to implement management of large numbers of distributed network devices. This document describes: • Cluster Management Overview • Configuring the Management Device • Configuring the Member Devices • Configuring Access Between the Management Device and Its Member Devices • Adding a Candidate Device to a Cluster • Configuring Advanced Cluster Functions

Features	Description
IRF Stack	Intelligent Resilient Framework (IRF) allows you to build an IRF stack, namely a united device, by interconnecting multiple devices through stack ports. You can manage all the devices in the IRF stack by managing the united device. This document describes: • IRF Stack Overview • IRF Stack Working Process • Configuring IRF Stack • Logging In to an IRF Stack
GR Overview	Graceful Restart ensures the continuity of packet forwarding when a protocol restarts. This document describes: • Introduction to Graceful Restart • Basic Concepts in Graceful Restart • Graceful Restart Communication Procedure • Graceful Restart Mechanism for Several Commonly Used Protocols
Automatic Configuration	Automatic configuration enables a device to automatically obtain and execute the configuration file when it starts up without loading the configuration file. This document describes: • Introduction to Automatic Configuration • Typical Networking of Automatic Configuration • How Automatic Configuration Works
IPC	Inter-Process Communication (IPC) is a reliable communication mechanism among different nodes. This document introduces the commands for Enabling IPC Performance Statistics.

Table of Contents

1 Logging In to an Ethernet Switch

When logging in to an Ethernet switch, go to these sections for information you are interested in:

Logging In to an Ethernet Switch

You can log in to an 3Com Switch 4800G in one of the following ways:

- Logging in locally through the Console port
- Telnetting locally or remotely to an Ethernet port
- Logging in through NMS (network management station)

Introduction to User Interface

Supported User Interfaces

3Com Switch 4800G supports two types of user interfaces: AUX and VTY.

Table 1-1 Description on user interface

User interface	Applicable user	Port used	Description
AUX	Users logging in through the Console port	Console port	Each switch can accommodate one AUX user.
VTY	Telnet users and SSH users	Ethernet port	Each switch can accommodate up to five VTY users.

Note

As the AUX port and the Console port of a 3Com Switch 4800G are the same one, you will be in the AUX user interface if you log in through this port.

User Interface Number

Two kinds of user interface index exist: absolute user interface index and relative user interface index.

1) The absolute user interface indexes are as follows:

- AUX user interface: 0
- VTY user interfaces: Numbered after AUX user interfaces and increases in the step of 1

2) A relative user interface index can be obtained by appending a number to the identifier of a user interface type. It is generated by user interface type. The relative user interface indexes are as follows:

- AUX user interface: AUX 0

- VTY user interfaces: VTY 0, VTY 1, VTY 2, and so on.

Common User Interface Configuration

Follow these steps to perform common user interface configuration:

To do…	Use the command…	Remarks
Lock the current user interface	**lock**	Optional Execute this command in user view. A user interface is not locked by default.
Specify to send messages to all user interfaces/a specified user interface	**send** { **all** | *number* | *type number* }	Optional Execute this command in user view.
Disconnect a specified user interface	**free user-interface** [*type*] *number*	Optional Execute this command in user view.
Enter system view	**system-view**	—
Set the banner	**header** { **incoming** | **legal** | **login** | **shell** | **motd** } *text*	Optional
Set a system name for the switch	**sysname** *string*	Optional Default is 4800G
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—
Define a shortcut key for aborting tasks	**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Set the history command buffer size	**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface	**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

To do…	Use the command…	Remarks
Set the maximum number of lines the screen can contain	**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Make terminal services available	**shell**	Optional By default, terminal services are available in all user interfaces.
Set the display type of a terminal	**terminal type** { **ansi** | **vt100** }	Optional By default, the terminal display type is ANSI. The device must use the same type of display as the terminal. If the terminal uses VT 100, the device should also use VT 100.
Display the information about the current user interface/all user interfaces	**display users** [**all**]	You can execute this command in any view.
Display the physical attributes and configuration of the current/a specified user interface	**display user-interface** [*type number* | *number*] [**summary**]	You can execute this command in any view.

2 Logging In Through the Console Port

When logging in through the Console port, go to these sections for information you are interested in:

Introduction

To log in through the Console port is the most common way to log in to a switch. It is also the prerequisite to configure other login methods. By default, you can log in to an 3Com Switch 4800G through its Console port only.

To log in to an Ethernet switch through its Console port, the related configuration of the user terminal must be in accordance with that of the Console port.

Table 2-1 lists the default settings of a Console port.

Table 2-1 The default settings of a Console port

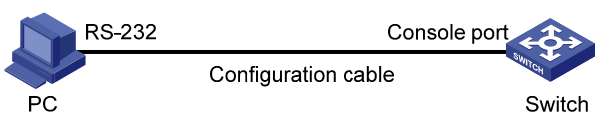
Setting	Default
Baud rate	19,200 bps
Flow control	Off
Check mode	No check bit
Stop bits	1
Data bits	8

After logging in to a switch, you can perform configuration for AUX users. Refer to Console Port Login Configuration for details.

Setting Up the Connection to the Console Port

- Connect the serial port of your PC/terminal to the Console port of the switch, as shown in Figure 2-1.

Figure 2-1 Diagram for setting the connection to the Console port



- If you use a PC to connect to the Console port, launch a terminal emulation utility (such as Terminal in Windows 3.X or HyperTerminal in Windows 9X/Windows 2000/Windows XP) and perform the configuration shown in Figure 2-2 through Figure 2-4 for the connection to be created. Normally, the parameters of a terminal are configured as those listed in Table 2-1.

Figure 2-2 Create a connection

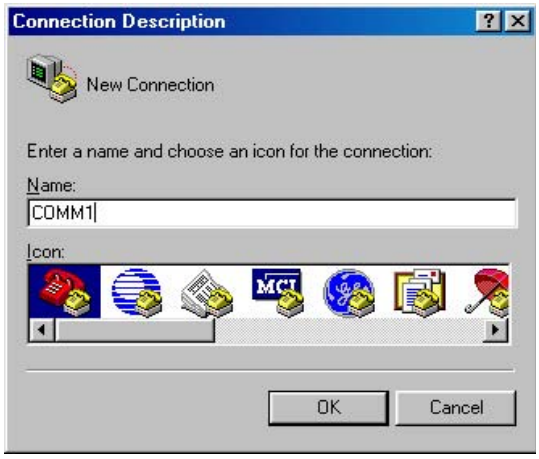
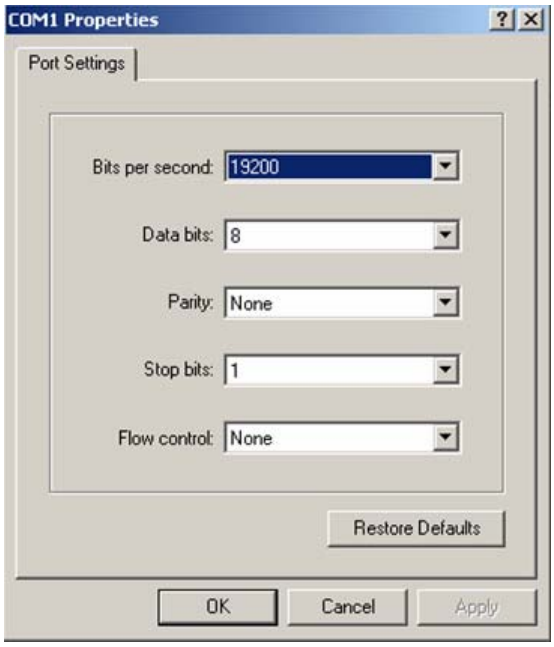


Figure 2-3 Specify the port used to establish the connection

Figure 2-4 Set port parameters terminal window



- Turn on the switch. The user will be prompted to press the Enter key if the switch successfully completes POST (power-on self test). The prompt (such as username) appears and you need input the username and password after the user presses the Enter key.

Note

The username is " admin " and none of the password.

- You can then configure the switch or check the information about the switch by executing commands. You can also acquire help by type the ? character. Refer to the following chapters for information about the commands.

Console Port Login Configuration

Common Configuration

Table 2-2 lists the common configuration of Console port login.

Table 2-2 Common configuration of Console port login

Configuration		Description
Console port configuration	Baud rate	Optional The default baud rate is 19,200 bps.
	Check mode	Optional By default, the check mode of the Console port is set to "none", which means no check bit.
	Stop bits	Optional The default stop bits of a Console port is 1.

Configuration		Description
	Data bits	Optional The default data bits of a Console port is 8.
	Flow control	Optional The default is **none**, which disables flow control.
AUX user interface configuration	Configure the command level available to the users logging in to the AUX user interface	Optional By default, commands of level 3 are available to the users logging in to the AUX user interface.
Terminal configuration	Define a shortcut key for aborting tasks	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
	Define a shortcut key for starting terminal sessions	Optional By default, pressing Enter key starts the terminal session.
	Make terminal services available	Optional By default, terminal services are available in all user interfaces
	Set the maximum number of lines the screen can contain	Optional By default, the screen can contain up to 24 lines.
	Set history command buffer size	Optional By default, the history command buffer can contain up to 10 commands.
	Set the timeout time of a user interface	Optional The default timeout time is 10 minutes.

Caution

Changing of Console port configuration terminates the connection to the Console port. To establish the connection again, you need to modify the configuration of the termination emulation utility running on your PC accordingly. Refer to Setting Up the Connection to the Console Port for details.

Console Port Login Configurations for Different Authentication Modes

Table 2-3 lists Console port login configurations for different authentication modes.

Table 2-3 Console port login configurations for different authentication modes

Authentication mode	Console port login configuration		Description
None	Perform common configuration	Perform common configuration for Console port login	Optional Refer to Common Configuration for details.

Authentication mode	Console port login configuration		Description
Password	Configure the password	Configure the password for local authentication	Required
	Perform common configuration	Perform common configuration for Console port login	Optional Refer to Common Configuration for details.
Scheme	Specify to perform local authentication or RADIUS authentication	AAA configuration specifies whether to perform local authentication or RADIUS authentication	Optional Local authentication is performed by default. Refer to the *AAA Configuration* in the *Security Volume* for details.
	Configure user name and password	Configure user names and passwords for local/remote users	Required • The user name and password of a local user are configured on the switch. • The user name and password of a remote user are configured on the RADIUS server. Refer to user manual of RADIUS server for details.
	Manage AUX users	Set service type for AUX users	Required
	Perform common configuration	Perform common configuration for Console port login	Optional Refer to Common Configuration for details.

Note

Changes of the authentication mode of Console port login will not take effect unless you exit and enter again the CLI.

Console Port Login Configuration with Authentication Mode Being None

Configuration Procedure

Follow these steps to perform Console port login configuration (with authentication mode being **none**):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter AUX user interface view	**user-interface aux** 0	—
Configure not to authenticate users	**authentication-mode none**	Required By default, users logging in through the Console port is scheme authenticated.

To do…		Use the command…	Remarks
Configure the Console port	Set the baud rate	**speed** speed-value	Optional The default baud rate of an AUX port (also the Console port) is 19,200 bps.
	Set the check mode	**parity** { **even** | **mark** | **none** | **odd** | **space** }	Optional By default, the check mode of a Console port is set to **none**, that is, no check bit.
	Set the stop bits	**stopbits** { **1** | **1.5** | **2** }	Optional The stop bits of a Console port is 1.
	Set the data bits	**databits** { **5** | **6** | **7** | **8** }	Optional The default data bits of a Console port is 8.
Configure the command level available to users logging in to the user interface		**user privilege level** *level*	Optional By default, commands of level 3 are available to users logging in to the AUX user interface.
Define a shortcut key for starting terminal sessions		**activation-key** *character*	Optional By default, pressing Enter key starts the terminal session.
Define a shortcut key for aborting tasks		**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available		**shell**	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain		**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Set the history command buffer size		**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface		**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that if you configure not to authenticate the users, the command level available to users logging in to a switch depends on both the **authentication-mode none** command and the **user privilege level** *level* command, as listed in the following table.

Table 2-4 Determine the command level (A)

Scenario			Command level
Authentication mode	**User type**	**Command**	
None (**authentication-mode none**)	Users logging in through Console ports	The **user privilege level** *level* command not executed	Level 3
		The **user privilege level** *level* command already executed	Determined by the *level* argument

Configuration Example

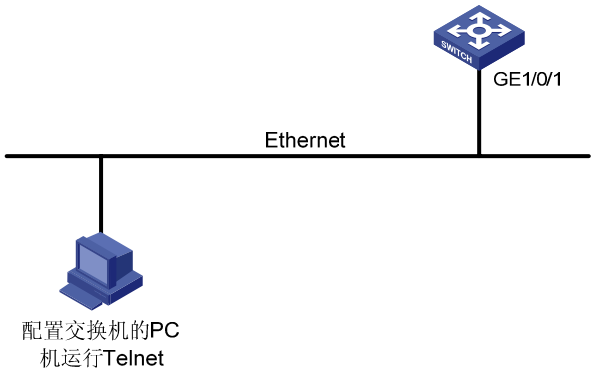
Network requirements

Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the console user at the following aspects.

- The user is not authenticated when logging in through the Console port.
- Commands of level 2 are available to user logging in to the AUX user interface.
- The baud rate of the Console port is 19200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 2-5 Network diagram for AUX user interface configuration (with the authentication mode being **none**)



Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter AUX user interface view.

```
[Sysname] user-interface aux 0
```

Specify not to authenticate the user logging in through the Console port.

```
[Sysname-ui-aux0] authentication-mode none
```

Specify commands of level 2 are available to the user logging in to the AUX user interface.

```
[Sysname-ui-aux0] user privilege level 2
```

Set the baud rate of the Console port to 19200 bps.

```
[Sysname-ui-aux0] speed 19200
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-aux0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-aux0] history-command max-size 20
```

Set the timeout time of the AUX user interface to 6 minutes.

```
[Sysname-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to Setting Up the Connection to the Console Port for details.

Console Port Login Configuration with Authentication Mode Being Password

Configuration Procedure

Follow these steps to perform Console port login configuration (with authentication mode being **password**):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter AUX user interface view	**user-interface aux** 0	—
Configure to authenticate users using the local password	**authentication-mode password**	Required By default, users logging in through the Console port and Telnet need to pass the scheme authentication.
Set the local password	**set authentication password** { **cipher** | **simple** } *password*	Required

To do…		Use the command…	Remarks
Configure the Console port	Set the baud rate	**speed** speed-value	Optional The default baud rate of an AUX port (also the Console port) is 19,200 bps.
	Set the check mode	**parity** { **even** | **mark** | **none** | **odd** | **space** }	Optional By default, the check mode of a Console port is set to **none**, that is, no check bit.
	Set the stop bits	**stopbits** { **1** | **1.5** | **2** }	Optional The default stop bits of a Console port is 1.
	Set the data bits	**databits** { **5** | **6** | **7** | **8** }	Optional The default data bits of a Console port is 8.
Configure the command level available to users logging in to the user interface		**user privilege level** *level*	Optional By default, commands of level 3 are available to users logging in to the AUX user interface.
Define a shortcut key for starting terminal sessions		**activation-key** *character*	Optional By default, pressing Enter key starts the terminal session.
Define a shortcut key for aborting tasks		**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available to the user interface		**shell**	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain		**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Set history command buffer size		**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface		**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that if you configure to authenticate the users in the password mode, the command level available to users logging in to a switch depends on both the **authentication-mode password** and the **user privilege level** *level* command, as listed in the following table.

Table 2-5 Determine the command level (B)

Scenario			Command level
Authentication mode	**User type**	**Command**	
Local authentication (**authentication-mode password**)	Users logging in to the AUX user interface	The **user privilege level** *level* command not executed	Level 3
		The **user privilege level** *level* command already executed	Determined by the *level* argument

Configuration Example

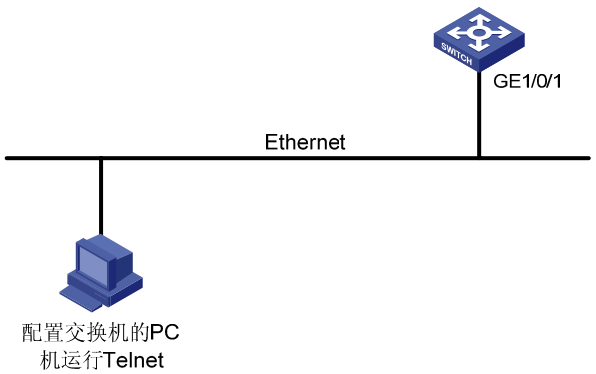
Network requirements

Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the Console user at the following aspects.

- The user is authenticated against the local password when logging in through the Console port.
- The local password is set to 123456 (in plain text).
- The commands of level 2 are available to users logging in to the AUX user interface.
- The baud rate of the Console port is 19,200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 2-6 Network diagram for AUX user interface configuration (with the authentication mode being **password**)



Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter AUX user interface view.

```
[Sysname] user-interface aux 0
```

Specify to authenticate the user logging in through the Console port using the local password.

```
[Sysname-ui-aux0] authentication-mode password
```

Set the local password to **123456** (in plain text).

```
[Sysname-ui-aux0] set authentication password simple 123456
```

Specify commands of level 2 are available to the user logging in to the AUX user interface.

```
[Sysname-ui-aux0] user privilege level 2
```

Set the baud rate of the Console port to 19200 bps.

```
[Sysname-ui-aux0] speed 19200
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-aux0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-aux0] history-command max-size 20
```

Set the timeout time of the AUX user interface to 6 minutes.

```
[Sysname-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to Setting Up the Connection to the Console Port for details.

Console Port Login Configuration with Authentication Mode Being Scheme

Configuration Procedure

Follow these steps to perform Console port login configuration (with authentication mode being **scheme**):

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Configure the authentication mode	Enter the default ISP domain view	**domain** *domain name*	Optional By default, the local AAA scheme is applied. If you specify to apply the local AAA scheme, you need to perform the configuration concerning local user as well. If you specify to apply an existing scheme by providing the *radius-scheme-name* argument, you need to perform the following configuration as well: • Perform AAA-RADIUS configuration on the switch. (Refer to *AAA Configuration* in the *Security Volume* for details.) • Configure the user name and password accordingly on the AAA server. (Refer to the user manual of AAA server.)
	Specify the AAA scheme to be applied to the domain	**authentication default** { **hwtacacs- scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	
	Quit to system view	**quit**	
Create a local user (Enter local user view.)		**local-user** *user-name*	Required local user is admin by default.
Set the authentication password for the local user		**password** { **simple** | **cipher** } *password*	Required
Specify the service type for AUX users		**service-type terminal**	Required
Quit to system view		**quit**	—
Enter AUX user interface view		**user-interface aux** 0	—
Configure to authenticate users locally or remotely		**authentication-mode scheme** [**command-authorization**]	Required The specified AAA scheme determines whether to authenticate users locally or remotely. Users are authenticated locally by default.
Configure the Console port	Set the baud rate	**speed** speed-value	Optional The default baud rate of the AUX port (also the Console port) is 19,200 bps.
	Set the check mode	**parity** { **even** | **mark** | **none** | **odd** | **space** }	Optional By default, the check mode of a Console port is set to **none**, that is, no check bit.
	Set the stop bits	**stopbits** { **1** | **1.5** | **2** }	Optional The default stop bits of a Console port is 1.
	Set the data bits	**databits** { **5** | **6** | **7** | **8** }	Optional The default data bits of a Console port is 8.
Configure the command level available to users logging in to the user interface		**user privilege level** *level*	Optional By default, commands of level 3 are available to users logging in to the AUX user interface.

To do…	Use the command…	Remarks
Define a shortcut key for starting terminal sessions	**activation-key** *character*	Optional By default, pressing Enter key starts the terminal session.
Define a shortcut key for aborting tasks	**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available to the user interface	**shell**	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Set history command buffer size	**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface	**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that the level the commands of which are available to users logging in to a switch depends on the **authentication-mode scheme** [**command-authorization**] command, and the **user privilege level** *level* command.

Configuration Example

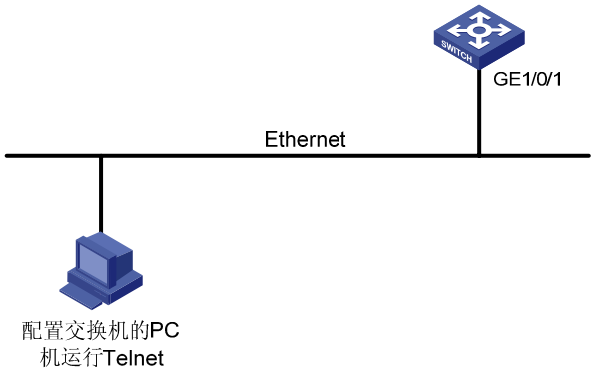
Network requirements

Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the console user at the following aspects.

- Configure the name of the local user to be “guest”.
- Set the authentication password of the local user to 123456 (in plain text).
- Set the service type of the local user to Terminal.
- Configure to authenticate the user logging in through the Console port in the scheme mode.
- The baud rate of the Console port is 19,200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 2-7 Network diagram for AUX user interface configuration (with the authentication mode being **scheme**)



Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create a local user named guest and enter local user view.

```
[Sysname] local-user guest
```

Set the authentication password to **123456** (in plain text).

```
[Sysname-luser-guest] password simple 123456
```

Set the service type to Terminal.

```
[Sysname-luser-guest] service-type terminal
[Sysname-luser-guest] quit
```

Enter AUX user interface view.

```
[Sysname] user-interface aux 0
```

Configure to authenticate the user logging in through the Console port in the scheme mode.

```
[Sysname-ui-aux0] authentication-mode scheme
```

Set the baud rate of the Console port to 19200 bps.

```
[Sysname-ui-aux0] speed 19200
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-aux0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-aux0] history-command max-size 20
```

Set the timeout time of the AUX user interface to 6 minutes.

```
[Sysname-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to Setting Up the Connection to the Console Port for details.

3 Logging In Through Telnet

When logging in through Telnet, go to these sections for information you are interested in:

Introduction

You can telnet to a remote switch to manage and maintain the switch. To achieve this, you need to configure both the switch and the Telnet terminal properly.

Table 3-1 Requirements for Telnet to a switch

Item	Requirement
Switch	Start the Telnet Server
	The IP address of the VLAN of the switch is configured and the route between the switch and the Telnet terminal is available.
	The authentication mode and other settings are configured. Refer to Table 3-2 and Table 3-3.
Telnet terminal	Telnet is running.
	The IP address of the management VLAN of the switch is available.

Note

- After you log in to the switch through Telnet, you can issue commands to the switch by way of pasting session text, which cannot exceed 2000 bytes, and the pasted commands must be in the same view; otherwise, the switch may not execute the commands correctly.
- If the session text exceeds 2000 bytes, you can save it in a configuration file, upload the configuration file to the switch and reboot the switch with this configuration file. For details, refer to *File System Management* in the *System Volume*.
- To log in on the switch using Telnet based on IPv6 is same as that based on IPv4, and you can refer to *IPv6 Configuration* for details.

Common Configuration

Table 3-2 lists the common Telnet configuration.

Table 3-2 Common Telnet configuration

Configuration		Remarks
VTY user interface configuration	Configure the command level available to users logging in to the VTY user interface	Optional By default, commands of level 0 are available to users logging in to a VTY user interface.
	Configure the protocols the user interface supports	Optional By default, Telnet and SSH protocol are supported.
	Set the command that is automatically executed when a user logs into the user interface	Optional By default, no command is automatically executed when a user logs into a user interface.
VTY terminal configuration	Define a shortcut key for aborting tasks	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
	Make terminal services available	Optional By default, terminal services are available in all user interfaces
	Set the maximum number of lines the screen can contain	Optional By default, the screen can contain up to 24 lines.
	Set history command buffer size	Optional By default, the history command buffer can contain up to 10 commands.
	Set the timeout time of a user interface	Optional The default timeout time is 10 minutes.

Caution

- The **auto-execute command** command may cause you unable to perform common configuration in the user interface, so use it with caution.
- Before executing the **auto-execute command** command and save your configuration, make sure you can log in to the switch in other modes and cancel the configuration.

Telnet Configurations for Different Authentication Modes

Table 3-3 lists Telnet configurations for different authentication modes.

Table 3-3 Telnet configurations for different authentication modes

Authentication mode	Telnet configuration		Remarks
None	Perform common configuration	Perform common Telnet configuration	Optional Refer to Table 3-2.

Authentication mode	Telnet configuration		Remarks
Password	Configure the password	Configure the password for local authentication	Required
	Perform common configuration	Perform common Telnet configuration	Optional Refer to Table 3-2.
Scheme	Specify to perform local authentication or RADIUS authentication	AAA configuration specifies whether to perform local authentication or RADIUS authentication	Optional Local authentication is performed by default. Refer to *AAA Configuration* in the *Security Volume* for details.
	Configure user name and password	Configure user names and passwords for local/remote users	Required • The user name and password of a local user are configured on the switch. • The user name and password of a remote user are configured on the RADIUS server. Refer to user manual of RADIUS server for details.
	Manage VTY users	Set service type for VTY users	Required
	Perform common configuration	Perform common Telnet configuration	Optional Refer to Table 3-2.

Telnet Configuration with Authentication Mode Being None

Configuration Procedure

Follow these steps to perform Telnet configuration (with authentication mode being **none**):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter one or more VTY user interface views	**user-interface vty** *first-number* [*last-number*]	—
Configure not to authenticate users logging in to VTY user interfaces	**authentication-mode none**	Required By default, VTY users are authenticated after logging in.
Configure the command level available to users logging in to VTY user interface	**user privilege level** *level*	Optional By default, commands of level 0 are available to users logging in to VTY user interfaces.
Configure the protocols to be supported by the VTY user interface	**protocol inbound** { **all** | **ssh** | **telnet** }	Optional By default, both Telnet protocol and SSH protocol are supported.

To do…	Use the command…	Remarks
Set the command that is automatically executed when a user logs into the user interface	**auto-execute command** *text*	Optional By default, no command is automatically executed when a user logs into a user interface.
Define a shortcut key for aborting tasks	**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available	**shell**	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Set the history command buffer size	**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time of the VTY user interface	**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that if you configure not to authenticate the users, the command level available to users logging in to a switch depends on both the **authentication-mode none** command and the **user privilege level** *level* command, as listed in Table 3-4.

Table 3-4 Determine the command level when users logging in to switches are not authenticated

Scenario			Command level
Authentication mode	**User type**	**Command**	
None (**authentication-mode none**)	VTY users	The **user privilege level** *level* command not executed	Level 0
		The **user privilege level** *level* command already executed	Determined by the *level* argument

Configuration Example

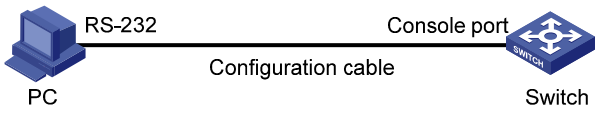
Network requirements

Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Do not authenticate users logging in to VTY 0.
- Commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 3-1 Network diagram for Telnet configuration (with the authentication mode being **none**)



Configuration procedure

Enter system view, and enable the Telnet service.

```
<Sysname> system-view
[Sysname] telnet server enable
```

Enter VTY 0 user interface view.

```
[Sysname] user-interface vty 0
```

Configure not to authenticate Telnet users logging in to VTY 0.

```
[Sysname-ui-vty0] authentication-mode none
```

Specify commands of level 2 are available to users logging in to VTY 0.

```
[Sysname-ui-vty0] user privilege level 2
```

Configure Telnet protocol is supported.

```
[Sysname-ui-vty0] protocol inbound telnet
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-vty0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-vty0] history-command max-size 20
```

Set the timeout time to 6 minutes.

```
[Sysname-ui-vty0] idle-timeout 6
```

Telnet Configuration with Authentication Mode Being Password

Configuration Procedure

Follow these steps to perform Telnet configuration (with authentication mode being **password**):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter one or more VTY user interface views	**user-interface vty** *first-number* [*last-number*]	—
Configure to authenticate users logging in to VTY user interfaces using the local password	**authentication-mode password**	Required
Set the local password	**set authentication password** { **cipher** | **simple** } *password*	Required
Configure the command level available to users logging in to the user interface	**user privilege level** *level*	Optional By default, commands of level 0 are available to users logging in to VTY user interface.
Configure the protocol to be supported by the user interface	**protocol inbound** { **all** | **ssh** | **telnet** }	Optional By default, both Telnet protocol and SSH protocol are supported.
Set the command that is automatically executed when a user logs into the user interface	**auto-execute command** *text*	Optional By default, no command is automatically executed when a user logs into a user interface.
Define a shortcut key for aborting tasks	**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available	**shell**	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.

To do…	Use the command…	Remarks
Set the history command buffer size	**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time of the user interface	**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that if you configure to authenticate the users in the password mode, the command level available to users logging in to a switch depends on both the **authentication-mode password** command and the **user privilege level** *level* command, as listed in Table 3-5.

Table 3-5 Determine the command level when users logging in to switches are authenticated in the password mode

Scenario			Command level
Authentication mode	**User type**	**Command**	
Password (authentication-mode password)	VTY users	The user privilege level *level* command not executed	Level 0
		The user privilege level *level* command already executed	Determined by the *level* argument

Configuration Example

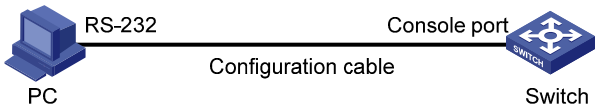
Network requirements

Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Authenticate users logging in to VTY 0 using the local password.
- Set the local password to 123456 (in plain text).
- Commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 3-2 Network diagram for Telnet configuration (with the authentication mode being **password**)



Configuration procedure

Enter system view, and enable the Telnet service.

```
<Sysname> system-view
[Sysname] telnet server enable
```

Enter VTY 0 user interface view.

```
[Sysname] user-interface vty 0
```

Configure to authenticate users logging in to VTY 0 using the local password.

```
[Sysname-ui-vty0] authentication-mode password
```

Set the local password to 123456 (in plain text).

```
[Sysname-ui-vty0] set authentication password simple 123456
```

Specify commands of level 2 are available to users logging in to VTY 0.

```
[Sysname-ui-vty0] user privilege level 2
```

Configure Telnet protocol is supported.

```
[Sysname-ui-vty0] protocol inbound telnet
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-vty0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-vty0] history-command max-size 20
```

Set the timeout time to 6 minutes.

```
[Sysname-ui-vty0] idle-timeout 6
```

Telnet Configuration with Authentication Mode Being Scheme

Configuration Procedure

Follow these steps to perform Telnet configuration (with authentication mode being **scheme**):

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Configure the authentication scheme	Enter the default ISP domain view	**domain** *domain name*	Optional By default, the local AAA scheme is applied. If you specify to apply the local AAA scheme, you need to perform the configuration concerning local user as well. If you specify to apply an existing scheme by providing the *radius-scheme-name* argument, you need to perform the following configuration as well: • Perform AAA-RADIUS configuration on the switch. (Refer to *AAA Configuration* in the *Security Volume* for details.) • Configure the user name and password accordingly on the AAA server. (Refer to the user manual of AAA server.)
	Configure the AAA scheme to be applied to the domain	**authentication default** { **hwtacacs-scheme** *hwtacacs-scheme- name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	
	Quit to system view	**quit**	
Create a local user and enter local user view		**local-user** *user-name*	No local user exists by default.
Set the authentication password for the local user		**password** { **simple** | **cipher** } *password*	Required
Specify the service type for VTY users		**service-type telnet**	Required
Quit to system view		**quit**	—
Enter one or more VTY user interface views		**user-interface vty** *first-number* [*last-number*]	—
Configure to authenticate users locally or remotely		**authentication-mode scheme**	Required The specified AAA scheme determines whether to authenticate users locally or remotely. Users are authenticated locally by default.
Configure the command level available to users logging in to the user interface		**user privilege level** *level*	Optional By default, commands of level 0 are available to users logging in to the VTY user interfaces.
Configure the supported protocol		**protocol inbound** { **all** | **ssh** | **telnet** }	Optional Both Telnet protocol and SSH protocol are supported by default.
Set the command that is automatically executed when a user logs into the user interface		**auto-execute command** *text*	Optional By default, no command is automatically executed when a user logs into a user interface.

To do…	Use the command…	Remarks
Define a shortcut key for aborting tasks	**escape-key** { **default** | *character* }	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
Make terminal services available	**shell**	Optional Terminal services are available in all use interfaces by default.
Set the maximum number of lines the screen can contain	**screen-length** *screen-length*	Optional By default, the screen can contain up to 24 lines. You can use the **screen-length** 0 command to disable the function to display information in pages.
Set history command buffer size	**history-command max-size** *value*	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface	**idle-timeout** *minutes* [*seconds*]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the **idle-timeout** 0 command to disable the timeout function.

Note that if you configure to authenticate the users in the scheme mode, the command level available to users logging in to a switch depends on the **authentication-mode scheme** [**command-authorization**] command and the **user privilege level** *level* command.

Note

Refer to *AAA Configuration* and *SSH2.0 Configuration* in the *Security Volume* for configuration about AAA, RADIUS and SSH..

Configuration Example

Network requirements

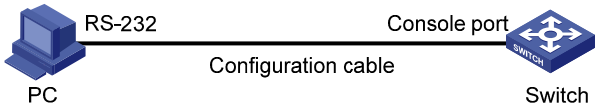
Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Configure the name of the local user to be “guest”.
- Set the authentication password of the local user to 123456 (in plain text).

- Set the service type of VTY users to Telnet.
- Configure to authenticate users logging in to VTY 0 in scheme mode.
- The commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported in VTY 0.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 3-3 Network diagram for Telnet configuration (with the authentication mode being **scheme**)



Configuration procedure

Enter system view, and enable the Telnet service.

```
<Sysname> system-view
[Sysname] telnet server enable
```

Create a local user named **guest** and enter local user view.

```
[Sysname] local-user guest
```

Set the authentication password of the local user to **123456** (in plain text).

```
[Sysname-luser-guest] password simple 123456
```

Set the service type to Telnet.

```
[Sysname-luser-guest] service-type
```

Enter VTY 0 user interface view.

```
[Sysname] user-interface vty 0
```

Configure to authenticate users logging in to VTY 0 in the scheme mode.

```
[Sysname-ui-vty0] authentication-mode scheme
```

Configure Telnet protocol is supported.

```
[Sysname-ui-vty0] protocol inbound telnet
```

Set the maximum number of lines the screen can contain to 30.

```
[Sysname-ui-vty0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[Sysname-ui-vty0] history-command max-size 20
```

Set the timeout time to 6 minutes.

```
[Sysname-ui-vty0] idle-timeout 6
```

Telnet Connection Establishment

Telnetting to a Switch from a Terminal

You can telnet to a switch and then configure the switch if the interface of the management VLAN of the switch is assigned with an IP address. (By default, VLAN 1 is the management VLAN.)

Following are procedures to establish a Telnet connection to a switch:

Step 1: Log in to the switch through the Console port, enable the Telnet server function and assign an IP address to the management VLAN interface of the switch.

- Connect to the Console port. Refer to Setting Up the Connection to the Console Port.
- Execute the following commands in the terminal window to enable the Telnet server function and assign an IP address to the management VLAN interface of the switch.

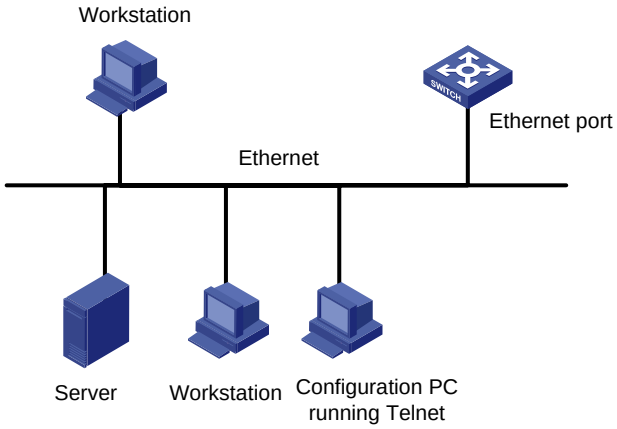
Enable the Telnet server function and configure the IP address of the management VLAN interface as 202.38.160.92, and .the subnet mask as 255.255.255.0.

```
<Sysname> system-view
[Sysname] telnet server enable
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] ip address 202.38.160.92 255.255.255.0
```

Step 2: Before Telnet users can log in to the switch, corresponding configurations should have been performed on the switch according to different authentication modes for them. Refer to Telnet Configuration with Authentication Mode Being None, Telnet Configuration with Authentication Mode Being Password, and Telnet Configuration with Authentication Mode Being Scheme for details. By default, Telnet users need to pass the password authentication to login.

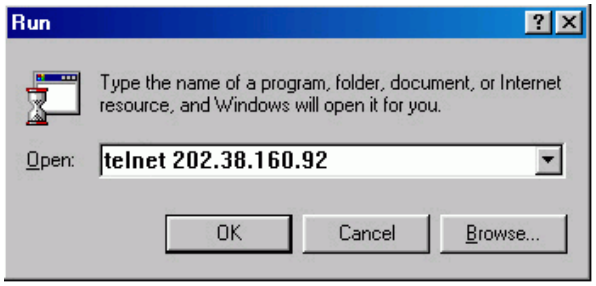
Step 3: Connect your PC to the Switch, as shown in Figure 3-4. Make sure the Ethernet port to which your PC is connected belongs to the management VLAN of the switch and the route between your PC and the switch is available.

Figure 3-4 Network diagram for Telnet connection establishment



Step 4: Launch Telnet on your PC, with the IP address of the management VLAN interface of the switch as the parameter, as shown in the following figure.

Figure 3-5 Launch Telnet



Step 5: Enter the password when the Telnet window displays “Login authentication” and prompts for login password. The CLI prompt (such as <4800G>) appears if the password is correct. If all VTY user interfaces of the switch are in use, you will fail to establish the connection and receive the message that says “All user interfaces are used, please try later!”. A 4800G Switch Ethernet switch can accommodate up to five Telnet connections at same time.

Step 6: After successfully Telnetting to a switch, you can configure the switch or display the information about the switch by executing corresponding commands. You can also type ? at any time for help. Refer to the following chapters for the information about the commands.

Note

- A Telnet connection will be terminated if you delete or modify the IP address of the VLAN interface in the Telnet session.
- By default, commands of level 0 are available to Telnet users authenticated by password. Refer to *Basic System Configuration* in the *System Volume* for information about command hierarchy.

Telnetting to Another Switch from the Current Switch

You can Telnet to another switch from the current switch. In this case, the current switch operates as the client, and the other operates as the server. If the interconnected Ethernet ports of the two switches are in the same LAN segment, make sure the IP addresses of the two management VLAN interfaces to which the two Ethernet ports belong to are of the same network segment, or the route between the two VLAN interfaces is available.

As shown in Figure 3-6, after Telnetting to a switch (labeled as Telnet client), you can Telnet to another switch (labeled as Telnet server) by executing the **telnet** command and then to configure the later.

Figure 3-6 Network diagram for Telnetting to another switch from the current switch

Step 1: Configure the user name and password for Telnet on the switch operating as the Telnet server. Refer to section Telnet Configuration with Authentication Mode Being None", section Telnet Configuration with Authentication Mode Being Password, and Telnet Configuration with Authentication Mode Being Scheme for details. By default, Telnet users need to pass the password authentication to login.

Step 2: Telnet to the switch operating as the Telnet client.

Step 3: Execute the following command on the switch operating as the Telnet client:

<Sysname> telnet xxxx

Where **xxxx** is the IP address or the host name of the switch operating as the Telnet server. You can use the **ip host** to assign a host name to a switch.

Step 4: Enter the password. If the password is correct, the CLI prompt (such as <4800G>) appears. If all VTY user interfaces of the switch are in use, you will fail to establish the connection and receive the message that says "All user interfaces are used, please try later!".

Step 5: After successfully Telnetting to the switch, you can configure the switch or display the information about the switch by executing corresponding commands. You can also type ? at any time for help. Refer to the following chapters for the information about the commands.

4 Logging in Through Web-based Network Management System

Introduction

An 3Com Switch 4800G has a Web server built in. You can log in to an Switch 4800G through a Web browser and manage and maintain the switch intuitively by interacting with the built-in Web server.

To log in to an Switch 4800G through the built-in Web-based network management system, you need to perform the related configuration on both the switch and the PC operating as the network management terminal.

Table 4-1 Requirements for logging in to a switch through the Web-based network management system

Item	Requirement
Switch	Start the Web server
	The IP address of the management VLAN of the switch is configured. The route between the switch and the network management terminal is available. (Refer to the module “IP Addressing and Performance” and “IP Routing” for more.)
	The user name and password for logging in to the Web-based network management system are configured.
PC operating as the network management terminal	IE is available.
	The IP address of the management VLAN interface of the switch is available.

HTTP Connection Establishment

Step 1: Log in to the switch through the console port and assign an IP address to the management VLAN interface of the switch. By default, VLAN 1 is the management VLAN.

- Connect to the console port. Refer to section Setting Up the Connection to the Console Port.
- Execute the following commands in the terminal window to assign an IP address to the management VLAN interface of the switch.

Configure the IP address of the management VLAN interface to be 10.153.17.82 with the mask 255.255.255.0.

```
<Sysname> system-view
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] ip address 10.153.17.82 255.255.255.0
```

Step 2: Configure the user name and the password for the Web-based network management system.

Configure the user name to be admin.

```
[Sysname] local-user admin
```

Set the password to admin.

```
[Sysname-luser-admin] password simple admin
```

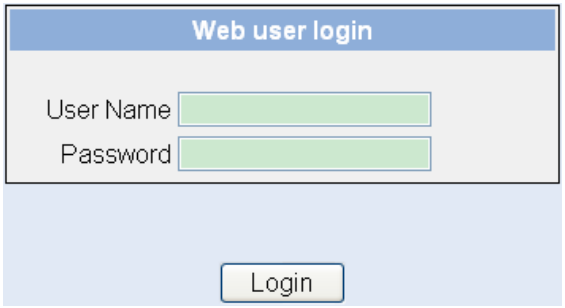
Step 3: Establish an HTTP connection between your PC and the switch, as shown in the following figure.

Figure 4-1 Establish an HTTP connection between your PC and the switch

Step 4: Log in to the switch through IE. Launch IE on the Web-based network management terminal (your PC) and enter the IP address of the management VLAN interface of the switch (here it is http://10.153.17.82). (Make sure the route between the Web-based network management terminal and the switch is available.)

Step 5: When the login interface (shown in Figure 4-2) appears, enter the user name and the password configured in step 2 and click <Login> to bring up the main page of the Web-based network management system.

Figure 4-2 The login page of the Web-based network management system



Web Server Shutdown/Startup

You can shut down or start up the Web server.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Shut down the Web server	**undo ip http enable**	Required Execute this command in system view. The Web server is started by default.
Start the Web server	**ip http enable**	Required Execute this command in system view.

Displaying Web Users

After the above configurations, execute the **display** command in any view to display the information about Web users, and thus to verify the configuration effect.

Table 4-2 Display information about Web users

To do…	Use the command…
Display information about Web users	**display web users**

5 Logging In Through NMS

When logging in through NMS, go to these sections for information you are interested in:

Introduction

You can also log in to a switch through an NMS (network management station), and then configure and manage the switch through the agent module on the switch.

- The agent here refers to the software running on network devices (switches) and as the server.
- SNMP (simple network management protocol) is applied between the NMS and the agent.

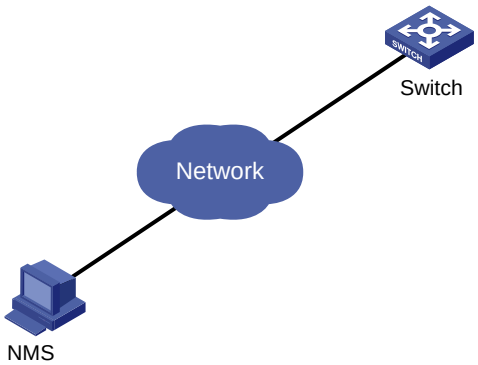
To log in to a switch through an NMS, you need to perform related configuration on both the NMS and the switch.

Table 5-1 Requirements for logging in to a switch through an NMS

Item	Requirement
Switch	The IP address of the management VLAN of the switch is configured. The route between the NMS and the switch is available.
	The basic SNMP functions are configured. (Refer to *SNMP Configuration* in the *System Volume* for details.)
NMS	The NMS is properly configured. (Refer to the user manual of the NMS for details.)

Connection Establishment Using NMS

Figure 5-1 Network diagram for logging in through an NMS



6 Specifying Source for Telnet Packets

When specifying source IP address/interface for Telnet packets, go to these sections for information you are interested in:

Introduction

To improve security and make it easier to manage services, you can specify source IP addresses/interfaces for Telnet clients.

Usually, VLAN interface IP addresses and Loopback interface IP addresses are used as the source IP addresses of Telnet packets. After you specify the IP address of a VLAN interface or a Loopback interface as the source IP address of Telnet packets, all the packets exchanged between the Telnet client and the Telnet server use the IP address as their source IP addresses, regardless of the ports through which they are transmitted. In such a way, the actual IP addresses used are concealed. This helps to improve security. Specifying source IP address/interfaces for Telnet packets also provides a way to successfully connect to servers that only accept packets with specific source IP addresses.

Specifying Source IP address/Interface for Telnet Packets

The configuration can be performed in user view and system view. The configuration performed in user view only applies to the current session. Whereas the configuration performed in system view applies to all the subsequent sessions. Priority in user view is higher than that in system view.

Specifying source IP address/interface for Telnet packets in user view

Follow these steps to specify source IP address/interface for Telnet packets in user view:

To do…	Use the command…	Remarks
Specify source IP address/interface for Telnet packets (the switch operates as a Telnet client)	**telnet** *remote-system* [*port-number*] [**source** { **ip** *ip-address* | **interface** *interface-type interface-number* }]	Optional By default, no source IP address/interface is specified.

Specifying source IP address/interface for Telnet packets in system view

Follow these steps to specify source IP address/interface for Telnet packets in system view:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—

To do…	Use the command…	Remarks
Specify source IP address/interface for Telnet packets	**telnet client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	Optional By default, no source IP address/interface is specified.

Note

- The IP address specified must be a local IP address.
- When specifying the source interface for Telnet packets, make sure the interface already exists.
- Before specifying the source IP address/interface for Telnet packets, make sure the route between the interface and the Telnet server is reachable.

Displaying the source IP address/Interface Specified for Telnet Packets

Follow these steps to display the source IP address/interface specified for Telnet packets:

To do…	Use the command…	Remarks
Display the source IP address/interface specified for Telnet packets	**display telnet client configuration**	Available in any view

7 Controlling Login Users

When controlling login users, go to these sections for information you are interested in:

Introduction

Multiple ways are available for controlling different types of login users, as listed in Table 7-1.

Table 7-1 Ways to control different types of login users

Login mode	Control method	Implementation	Related section
Telnet	By source IP addresses	Through basic ACLs	Controlling Telnet Users by Source IP Addresses
	By source and destination IP addresses	Through advanced ACLs	Controlling Telnet Users by Source and Destination IP Addresses
	By source MAC addresses	Through Layer 2 ACLs	Controlling Telnet Users by Source MAC Addresses
SNMP	By source IP addresses	Through basic ACLs	Controlling Network Management Users by Source IP Addresses

Controlling Telnet Users

Prerequisites

The controlling policy against Telnet users is determined, including the source and destination IP addresses to be controlled and the controlling actions (permitting or denying).

Controlling Telnet Users by Source IP Addresses

This configuration needs to be implemented by basic ACL; a basic ACL ranges from 2000 to 2999. For the definition of ACL, refer to *ACL Configuration* in the *Security Volume*.

Follow these steps to control Telnet users by source IP addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a basic ACL or enter basic ACL view	**acl** [**ipv6**] **number** *acl-number* [**match-order** { **config** | **auto** }]	As for the **acl number** command, the **config** keyword is specified by default.

To do…	Use the command…	Remarks
Define rules for the ACL	**rule** [*rule-id*] { **permit** | **deny** } [**source** { *sour-addr sour-wildcard* | **any** } | **time-range** *time-name* | **fragment** | **logging**]*	Required
Quit to system view	**quit**	—
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—
Apply the ACL to control Telnet users by source IP addresses	**acl** [**ipv6**] *acl-number* { **inbound** | **outbound** }	Required The **inbound** keyword specifies to filter the users trying to Telnet to the current switch. The **outbound** keyword specifies to filter users trying to Telnet to other switches from the current switch.

Controlling Telnet Users by Source and Destination IP Addresses

This configuration needs to be implemented by advanced ACL; an advanced ACL ranges from 3000 to 3999. For the definition of ACL, refer to *ACL Configuration* in the *Security Volume*.

Follow these steps to control Telnet users by source and destination IP addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an advanced ACL or enter advanced ACL view	**acl** [**ipv6**] **number** *acl-number* [**match-order** { **config** | **auto** }]	As for the **acl number** command, the **config** keyword is specified by default.
Define rules for the ACL	**rule** [*rule-id*] { **permit** | **deny** } *rule-string*	Required You can define rules as needed to filter by specific source and destination IP addresses.
Quit to system view	**quit**	—
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—
Apply the ACL to control Telnet users by specified source and destination IP addresses	**acl** [**ipv6**] *acl-number* { **inbound** | **outbound** }	Required The **inbound** keyword specifies to filter the users trying to Telnet to the current switch. The **outbound** keyword specifies to filter users trying to Telnet to other switches from the current switch.

Controlling Telnet Users by Source MAC Addresses

This configuration needs to be implemented by Layer 2 ACL; a Layer 2 ACL ranges from 4000 to 4999. For the definition of ACL, refer to *ACL Configuration* in the *Security Volume*.

Follow these steps to control Telnet users by source MAC addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a basic ACL or enter basic ACL view	**acl number** *acl-number* [**match-order** { **config** | **auto** }]	As for the **acl number** command, the **config** keyword is specified by default.
Define rules for the ACL	**rule** [*rule-id*] { **permit** | **deny** } *rule-string*	Required You can define rules as needed to filter by specific source MAC addresses.
Quit to system view	**quit**	—
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—
Apply the ACL to control Telnet users by source MAC addresses	**acl** *acl-number* **inbound**	Required The **inbound** keyword specifies to filter the users trying to Telnet to the current switch.

Note

Layer 2 ACL is invalid for this function if the source IP address of the Telnet client and the interface IP address of the Telnet server are not in the same subnet.

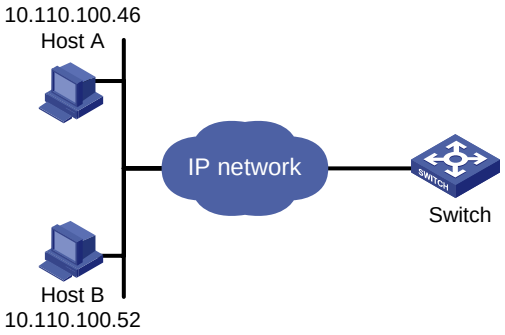
Configuration Example

Network requirements

Only the Telnet users sourced from the IP address of 10.110.100.52 and 10.110.100.46 are permitted to log in to the switch.

Network diagram

Figure 7-1 Network diagram for controlling Telnet users using ACLs



Configuration procedure

Define a basic ACL.

```
<Sysname> system-view
[Sysname] acl number 2000 match-order config
[Sysname-acl-basic-2000] rule 1 permit source 10.110.100.52 0
[Sysname-acl-basic-2000] rule 2 permit source 10.110.100.46 0
[Sysname-acl-basic-2000] rule 3 deny source any
[Sysname-acl-basic-2000] quit
```

Apply the ACL.

```
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] acl 2000 inbound
```

Controlling Network Management Users by Source IP Addresses

You can manage a 3Com Switch 4800G through network management software. Network management users can access switches through SNMP.

You need to perform the following two operations to control network management users by source IP addresses.

- Defining an ACL
- Applying the ACL to control users accessing the switch through SNMP

Prerequisites

The controlling policy against network management users is determined, including the source IP addresses to be controlled and the controlling actions (permitting or denying).

Controlling Network Management Users by Source IP Addresses

Follow these steps to control network management users by source IP addresses:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a basic ACL or enter basic ACL view	**acl number** *acl-number* [**match-order** { **config** | **auto** }]	As for the **acl number** command, the **config** keyword is specified by default.
Define rules for the ACL	**rule** [*rule-id*] { **permit** | **deny** } [**source** { *sour-addr sour-wildcard* | **any** } | **time-range** *time-name* | **fragment** | **logging**]*	Required
Quit to system view	**quit**	—
Apply the ACL while configuring the SNMP community name	**snmp-agent community** { **read** | **write** } *community-name* [**mib-view** *view-name* | **acl** *acl-number*]*	Required According to the SNMP version and configuration customs of NMS users, you can reference an ACL when configuring community name, group name or username. For the detailed configuration, refer to *SNMP Configuration* in the *System Volume*.
Apply the ACL while configuring the SNMP group name	**snmp-agent group** { **v1** | **v2c** } *group-name* [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*] **snmp-agent group v3** *group-name* [**authentication** | **privacy**] [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*]	
Apply the ACL while configuring the SNMP user name	**snmp-agent usm-user** { **v1** | **v2c** } *user-name group-name* [**acl** *acl-number*] **snmp-agent usm-user v3** *user-name group-name* [[**cipher**] **authentication-mode** { **md5** | **sha** } *auth-password* [**privacy-mode** { **aes128** | **des56** } *priv-password*]] [**acl** *acl-number*]	

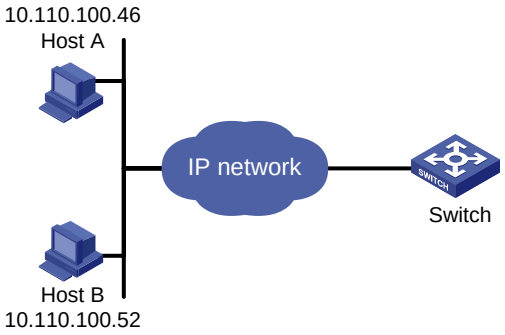
Configuration Example

Network requirements

Only SNMP users sourced from the IP addresses of 10.110.100.52 and 10.110.100.46 are permitted to access the switch.

Network diagram

Figure 7-2 Network diagram for controlling SNMP users using ACLs



Configuration procedure

Define a basic ACL.

```
<Sysname> system-view
[Sysname] acl number 2000 match-order config
[Sysname-acl-basic-2000] rule 1 permit source 10.110.100.52 0
[Sysname-acl-basic-2000] rule 2 permit source 10.110.100.46 0
[Sysname-acl-basic-2000] rule 3 deny source any
[Sysname-acl-basic-2000] quit
```

Apply the ACL to only permit SNMP users sourced from the IP addresses of 10.110.100.52 and 10.110.100.46 to access the switch.

```
[Sysname] snmp-agent community read 3com acl 2000
[Sysname] snmp-agent group v2c 3comgroup acl 2000
[Sysname] snmp-agent usm-user v2c 3comuser 3comgroup acl 2000
```

Table of Contents

1 Basic Configurations

While performing basic configurations of the system, go to these sections for information you are interested in:

- Configuration Display
- Basic Configurations
- CLI Features

Configuration Display

To avoid duplicate configuration, you can use the **display** commands to view the current configuration of the device before configuring the device. The configurations of a device fall into the following categories:

- Factory defaults: When devices are shipped, they are installed with some basic configurations, which are called factory defaults. These default configurations ensure that a device can start up and run normally when it has no configuration file or the configuration file is damaged.
- Current configuration: The currently running configuration on the device.
- Saved configuration: Configurations saved in the startup configuration file.

Follow these steps to display device configurations:

To do…	Use the command…	Remarks
Display the factory defaults of the device	**display default-configuration**	Available in any view.
Display the current validated configurations of the device	**display current-configuration** [[**configuration** [*configuration*] | **interface** [*interface-type*] [*interface-number*]] [**by-linenum**] [**|** { **begin** | **exclude** | **include** } *regular-expression*]]	
Display the configuration saved on the storage media of the device	**display saved-configuration** [**by-linenum**]	

Note

For details of the **display saved-configuration** command, refer to *File System Management Commands* in the *System Volume.*

Basic Configurations

This section covers the following topics:

- Entering/Exiting System View

Entering/Exiting System View

Follow these steps to enter/exit system view:

To do…	Use the command…	Remarks
Enter system view from user view	**system-view**	—
Return to user view from system view	**quit**	—

Note

With the **quit** command, you can return to the previous view. You can execute the **return** command or press the hot key **Ctrl+Z** to return to user view.

Configuring the Device Name

The device name is used to identify a device in a network. Inside the system, the device name corresponds to the prompt of the CLI. For example, if the device name is **Sysname**, the prompt of user view is <Sysname>.

Follow these steps to configure the device name:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the device name	**sysname** *sysname*	Optional The device name is “3Com” by default.

Configuring the System Clock

Configuring the system clock

The system clock, displayed by system time stamp, is decided by the configured relative time, time zone, and daylight saving time. You can view the system clock by using the **display clock** command.

Follow these steps to configure the system clock:

To do…	Use the command…	Remarks
Set time and date	**clock datetime** *time date*	Optional Available in user view.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the time zone	**clock timezone** *zone-name* { **add** | **minus** } *zone-offset*	Optional
Set a daylight saving time scheme	**clock summer-time** *zone-name* **one-off** *start-time start-date end-time end-date add-time*	Optional Use either command
	clock summer-time *zone-name* **repeating** *start-time start-date end-time end-date add-time*	

Displaying the system clock

The system clock is decided by the commands **clock datetime**, **clock timezone** and **clock summer-time**. If these three commands are not configured, the **display clock** command displays the original system clock. If you combine these three commands in different ways, the system clock is displayed in the ways shown in Table 1-1. The meanings of the parameters in the configuration column are as follows:

- 1 indicates date-time has been configured with the **clock datetime**.
- 2 indicates time-zone has been configured with the **clock timezone** command and the offset time is *zone-offset.*
- 3 indicates daylight saving time has been configured with the **clock summer-time** command and the offset time is *summer-offset.*
- [1] indicates the **clock datetime** command is an optional configuration.
- The default system clock is 2005/1/1 1:00:00 in the example.

Table 1-1 Relationship between the configuration and display of the system clock

Configuration	System clock displayed by the display clock command	Example
1	*date-time*	Configure: clock datetime 1:00 2007/1/1 Display: 01:00:00 UTC Mon 01/01/2007
2	The original system clock ± *zone-offset*	Configure: clock timezone zone-time add 1 Display: 02:00:00 zone-time Sat 01/01/2005
1 and 2	*date-time* ± *zone-offset*	Configure: clock datetime 2:00 2007/2/2 and clock timezone zone-time add 1 Display: 03:00:00 zone-time Fri 02/02/2007
[1], 2 and 1	*date-time*	Configure: clock timezone zone-time add 1 and clock datetime 3:00 2007/3/3 Display: 03:00:00 zone-time Sat 03/03/2007
3	If the original system clock is not in the daylight saving time range, the original system clock is displayed.	Configure: clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 Display: 01:00:00 UTC Sat 01/01/2005
	If the original system clock is in the daylight saving time range, the original system clock + *summer-offset* is displayed.	Configure: clock summer-time ss one-off 00:30 2005/1/1 1:00 2005/8/8 2 Display: 03:00:00 ss Sat 01/01/2005

Configuration	System clock displayed by the display clock command	Example
1 and 3	If *date-time* is not in the daylight saving time range, *date-time* is displayed.	Configure: clock datetime 1:00 2007/1/1 and clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 Display: 01:00:00 UTC Mon 01/01/2007
	If *date-time* is in the daylight saving time range, "*date-time"* + "*summer-offset"* is displayed.	Configure: clock datetime 8:00 2007/1/1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 10:00:00 ss Mon 01/01/2007
[1], 3 and 1	If *date-time* is not in the daylight saving time range, *date-time* is displayed.	Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 1:00 2008/1/1 Display: 01:00:00 UTC Tue 01/01/2008
	date-time is in the daylight saving time range: If the value of *"date-time"* - *"summer-offset"* is not in the summer-time range, *"date-time"* - *"summer-offset"* is displayed; If the value of *"date-time"* - *"summer-offset"* is in the summer-time range, *date-time* is displayed.	Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 1:30 2007/1/1 Display: 23:30:00 UTC Sun 12/31/2006
		Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 3:00 2007/1/1 Display: 03:00:00 ss Mon 01/01/2007
2 and 3 or 3 and 2	If the value of the original system clock ± "zone-offset" is not in the summer-time range, the original system clock ± "zone-offset" is displayed.	Configure: clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 02:00:00 zone-time Sat 01/01/2005
		Configure: clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2005/1/1 1:00 2005/8/8 2 Display: 04:00:00 ss Sat 01/01/2005
	If the value of the original system clock ± "zone-offset" is in the summer-time range, the original system clock ± "zone-offset" + "summer-offset" is displayed.	Configure: clock datetime 1:00 2007/1/1, clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 Display: 02:00:00 zone-time Mon 01/01/2007
1, 2 and 3 or 1, 3 and 2	If the value of "*date-time*"±"*zone-offset*" is not in the summer-time range, "date-time"±"zone-offset" is displayed.	Configure: clock datetime 1:00 2007/1/1, clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 04:00:00 ss Mon 01/01/2007
	If the value of "*date-time*"±"*zone-offset*" is in the summer-time range, "date-time"±"zone-offset"+"summer-offset" is displayed.	Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 1:00 2007/1/1 Display: 01:00:00 zone-time Mon 01/01/2007

Configuration	System clock displayed by the display clock command	Example
[1], 2, 3 and 1 or [1], 3, 2 and 1	If *date-time* is not in the daylight saving time range, *date-time* is displayed.	Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 1:30 2008/1/1 Display: 23:30:00 zone-time Mon 12/31/2007
	date-time is in the daylight saving time range: If the value of *"date-time"-"summer-offset"* is not in the summer-time range, *"date-time"-"summer-offset"* is displayed; If the value of *"date-time"-"summer-offset"* is in the summer-time range, *date-time* is displayed.	Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 3:00 2008/1/1 Display: 03:00:00 ss Tue 01/01/2008

Enabling/Disabling the Display of Copyright Information

- With the display of copyright information enabled, the copyright information is displayed when a user logs in through Telnet or SSH, or when a user quits user view after logging in to the device through the console port, AUX port, or asynchronous serial interface. The copyright information will not be displayed under other circumstances. The display format of copyright information is as shown below:

```
**************************************************************************
* Copyright (c) 2004-2008 3Com Corp. and its licensors. All rights reserved. *
* This software is protected by copyright law and international treaties.    *
* Without the prior written permission of 3Com Corporation and its licensors,*
* any reproduction republication, redistribution, decompiling, reverse       *
* engineering is strictly prohibited. Any unauthorized use of this software  *
* or any portion of it may result in severe civil and criminal penalties, and*
* will be prosecuted to the maximum extent possible under the applicable law.*
**************************************************************************
```

- With the display of copyright information disabled, under no circumstances will the copyright information be displayed.

Follow these steps to enable/disable the display of copyright information:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the display of copyright information	**copyright-info enable**	Optional Enabled by default.
Disable the display of copyright information	**undo copyright-info enable**	Required Enabled by default.

Configuring a Banner

Introduction to banners

Banners are prompt information displayed by the system when users are connected to the device, perform login authentication, and start interactive configuration. The administrator can set corresponding banners as needed.

At present, the system supports the following five kinds of welcome information.

- **shell** banner, also called session banner, displayed when a non TTY Modem user enters user view.
- **incoming** banner, also called user interface banner, displayed when a user interface is activated by a Modem user.
- **login** banner, welcome information at login authentications, displayed when password and scheme authentications are configured.
- **motd** (Message of the Day) banner, welcome information displayed before authentication.
- **legal** banner, also called authorization information. The system displays some copyright or authorization information, and then displays the **legal** banner before a user logs in, waiting for the user to confirm whether to continue the authentication or login. If entering Y or pressing the **Enter** key, the user enters the authentication or login process; if entering N, the user quits the authentication or login process. Y and N are case insensitive.

Configuring a banner

When you configure a banner, the system supports two input modes. One is to input all the banner information right after the command keywords. The start and end characters of the input text must be the same but are not part of the banner information. In this case, the input text, together with the command keywords, cannot exceed 510 characters. The other is to input all the banner information in multiple lines by pressing the **Enter** key. In this case, up to 2000 characters can be input.

The latter input mode can be achieved in the following three ways:

- Press the **Enter** key directly after the command keywords, and end the setting with the % character. The **Enter** and % characters are not part of the banner information.
- Input a character after the command keywords at the first line, and then press the **Enter** key. End the setting with the character input at the first line. The character at the first line and the end character are not part of the banner information.
- Input multiple characters after the command keywords at the first line (with the first and last characters being different), then press the **Enter** key. End the setting with the first character at the first line. The first character at the first line and the end character are not part of the banner information.

Follow these steps to configure a banner:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the banner to be displayed at login (available for Modem login users)	**header incoming** *text*	Optional
Configure the banner to be displayed at login authentication	**header login** *text*	Optional
Configure the authorization information before login	**header legal** *text*	Optional

To do…	Use the command…	Remarks
Configure the banner to be displayed when a user enters user view (non Modem login users)	**header shell** *text*	Optional
Configure the banner to be displayed before login	**header motd** *text*	Optional

Configuring CLI Hotkeys

Follow these steps to configure CLI hotkeys:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure CLI hotkeys	**hotkey** { **CTRL_G** | **CTRL_L** | **CTRL_O** | **CTRL_T** | **CTRL_U** } *command*	Optional The **Ctrl+G**, **Ctrl+L** and **Ctrl+O** hotkeys are specified with command lines by default.
Display hotkeys	**display hotkey**	Available in any view. Refer to Table 1-2 for hotkeys reserved by the system.

Note

By default, the **Ctrl+G**, **Ctrl+L** and **Ctrl+O** hotkeys are configured with command line and the **Ctrl+T** and **Ctrl+U** commands are NULL.

- **Ctrl+G** corresponds to the **display current-configuration** command.
- **Ctrl+L** corresponds to the **display ip routing-table** command.
- **Ctrl+O** corresponds to the **undo debugging all** command.

Table 1-2 Hotkeys reserved by the system

Hotkey	Function
Ctrl+A	Moves the cursor to the beginning of the current line.
Ctrl+B	Moves the cursor one character to the left.
Ctrl+C	Stops performing a command.
Ctrl+D	Deletes the character at the current cursor position.
Ctrl+E	Moves the cursor to the end of the current line.
Ctrl+F	Moves the cursor one character to the right.
Ctrl+H	Deletes the character to the left of the cursor.
Ctrl+K	Terminates an outgoing connection.
Ctrl+N	Displays the next command in the history command buffer.
Ctrl+P	Displays the previous command in the history command buffer.
Ctrl+R	Redisplays the current line information.
Ctrl+V	Pastes the content in the clipboard.
Ctrl+W	Deletes all the characters in a continuous string to the left of the cursor.

Hotkey	Function
Ctrl+X	Deletes all the characters to the left of the cursor.
Ctrl+Y	Deletes all the characters to the right of the cursor.
Ctrl+Z	Exits to user view.
Ctrl+]	Terminates an incoming connection or a redirect connection.
Esc+B	Moves the cursor to the leading character of the continuous string to the left.
Esc+D	Deletes all the characters of the continuous string at the current cursor position and to the right of the cursor.
Esc+F	Moves the cursor to the front of the next continuous string to the right.
Esc+N	Moves the cursor down by one line (available before you press **Enter**)
Esc+P	Moves the cursor up by one line (available before you press **Enter**)
Esc+<	Specifies the cursor as the beginning of the clipboard.
Esc+>	Specifies the cursor as the ending of the clipboard.

Note

These hotkeys are defined by the device. When you interact with the device from terminal software, these keys may be defined to perform other operations. If so, the definition of the terminal software will dominate.

Configuring User Privilege Levels and Command Levels

Introduction

To restrict the different users' access to the device, the system manages the users by their privilege levels. User privilege levels correspond to command levels. After users at different privilege levels log in, they can only use commands at their own, or lower, levels. All the commands are categorized into four levels, which are visit, monitor, system, and manage from low to high, and identified respectively by 0 through 3. Table 1-3 describes the levels of the commands.

Table 1-3 Default command levels

Level	Privilege	Description
0	Visit	Involves commands for network diagnosis and commands for accessing an external device. Commands at this level are not allowed to be saved after being configured. After the device is restarted, the commands at this level will be restored to the default settings. Commands at this level include **ping**, **tracert**, **telnet** and **ssh2**.
1	Monitor	Includes commands for system maintenance and service fault diagnosis. Commands at this level are not allowed to be saved after being configured. After the device is restarted, the commands at this level will be restored to the default settings. Commands at this level include **debugging**, **terminal**, **refresh**, **reset**, and **send**.

Level	Privilege	Description
2	System	Provides service configuration commands, including routing and commands at each level of the network for providing services. By default, commands at this level include all configuration commands except for those at manage level.
3	Manage	Influences the basic operation of the system and the system support modules for service support. By default, commands at this level involve file system, FTP, TFTP, Xmodem command download, user management, level setting, as well as parameter setting within a system (the last case involves those non-protocol or non RFC provisioned commands).

Configuring user privilege level

User privilege level can be configured by using AAA authentication parameters or under a user interface.

1) Configure user privilege level by using AAA authentication parameters

If the user interface authentication mode is **scheme** when a user logs in, and username and password are needed at login, then the user privilege level is specified in the configuration of AAA authentication.

Follow these steps to configure user privilege level by using AAA authentication parameters:

<table>
<tr><th colspan="2">To do…</th><th>Use the command…</th><th>Remarks</th></tr>
<tr><td colspan="2">Enter system view</td><td>system-view</td><td>—</td></tr>
<tr><td colspan="2">Enter user interface view</td><td>user-interface [type] first-number [last-number]</td><td>—</td></tr>
<tr><td colspan="2">Configure the authentication mode for logging in to the user interface as scheme</td><td>authentication-mode scheme [command-authorization]</td><td>Required
By default, the authentication mode for VTY and AUX users is password.</td></tr>
<tr><td colspan="2">Exit to system view</td><td>quit</td><td>—</td></tr>
<tr><td colspan="2">Configure the authentication mode for SSH users as password</td><td>For the details, refer to SSH2.0 Configuration in the Security Volume.</td><td>Required if users use SSH to log in, and username and password are needed at authentication</td></tr>
<tr><td rowspan="2">Configure the user privilege level by using AAA authentication parameters</td><td>Using local authentication</td><td>• Use the local-user command to create a local user and enter local user view.
• Use the level keyword in the authorization-attribute command to configure the user level.</td><td rowspan="2">User either approach
• For local authentication, if you do not configure the user level, the user level is 0, that is, users of this level can use commands with level 0 only.
• For remote authentication, if you do not configure the user level, the user level depends on the default configuration of the authentication server.</td></tr>
<tr><td>Using remote authentication (RADIUS, HWTACACS, and LDAP authentication s)</td><td>Configure user level on the authentication server</td></tr>
</table>

Note

- For the description of user interface, refer to *Login Configuration* in the *System Volume*; for the description of the **user-interface**, **authentication-mode** and **user privilege level** commands, refer to *User Interface Commands* in the *System Volume*.
- For the introduction to AAA authentication, refer to *AAA Configuration* in the *Security Volume*; for the description of the **local-user** and **authorization-attribute** commands, refer to *AAA Commands* in the *Security Volume*.
- For the introduction to SSH, refer to *SSH 2.0 Configuration* in the *Security Volume*.

2) Example of configuring user privilege level by using AAA authentication parameters

Authenticate the users telnetting to the device through VTY 1, verify their usernames and passwords locally, and specify the user privilege level as 3.

```
<Sysname> system-view
[Sysname] user-interface vty 1
[Sysname-ui-vty1] authentication-mode scheme
[Sysname-ui-vty1] quit
[Sysname] local-user test
[Sysname-luser-test] password cipher 123
[Sysname-luser-test] service-type telnet
```

After the above configuration, when users telnet to the device through VTY 1, they need to input username **test** and password **123**. After passing the authentication, users can only use the commands of level 0. If the users need to use commands of levels 0, 1, 2 and 3, the following configuration is required:

```
[Sysname-luser-test] authorization-attribute level 3
```

3) Configure the user privilege level under a user interface

If the user interface authentication mode is **scheme** when a user logs in, and SSH **publickey** authentication type (only username is needed for this authentication type) is adopted, then the user privilege level is the user interface level; if a user logs in using the **none** or **password** mode (namely, no username is needed), the user privilege level is the user interface level.

Follow these steps to configure the user privilege level under a user interface (SSH **publickey** authentication type):

To do…	Use the command…	Remarks
Configure the authentication type for SSH users as **publickey**	For the details, refer to *SSH2.0 Configuration* in the *Security Volume*.	Required if users adopt the SSH login mode, and only username, instead of password is needed at authentication. After the configuration, the authentication mode of the corresponding user interface must be set to **scheme**.
Enter system view	**system-view**	—
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—

To do…	Use the command…	Remarks
Configure the authentication mode when a user uses the current user interface to log in to the device	**authentication-mode scheme** [**command-authorization**]	Optional By default, the authentication mode for VTY and AUX user interfaces is **password**.
Configure the privilege level of the user logging in from the current user interface	**user privilege level** *level*	Optional By default, the user privilege level for users logging in from the console user interface is 3, and that for users logging from the other user interfaces is 0.

Follow these steps to configure the user privilege level under a user interface (**none** or **password** authentication mode):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter user interface view	**user-interface** [*type*] *first-number* [*last-number*]	—
Configure the authentication mode when a user uses the current user interface to log in to the device	**authentication-mode** { **none** | **password** }	Optional By default, the authentication mode for VTY and AUX user interfaces is **password**.
Configure the privilege level of the user logging in from the current user interface	**user privilege level** *level*	Optional By default, the user privilege level for users logging in from the console user interface is 3, and that for users logging from the other user interfaces is 0.

4) Example of configuring user privilege level under a user interface

- Perform no authentication to the users telnetting to the device, and specify the user privilege level as 1. (This configuration brings potential security problem. Therefore, you are recommended to use it only in a lab environment.)

```
<Sysname> system-view
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode none
[Sysname-ui-vty0-4] user privilege level 1
```

By default, when users telnet to the device, they can only use the following commands after passing the authentication:

```
<Sysname> ?
User view commands:
  cluster        Run cluster command
  display        Display current system information
  ping           Ping function
  quit           Exit from current command view
  ssh2           Establish a secure shell client connection
  super          Set the current user priority level
```

```
  telnet          Establish one TELNET connection
  tracert         Trace route function
```

After you set the user privilege level under the user interface, users can log in to the device through Telnet without any authentication and use the following commands:

```
<Sysname> ?
User view commands:
  cluster         Run cluster command
  debugging       Enable system debugging functions
  display         Display current system information
  ipc             Interprocess communication
  ping            Ping function
  quit            Exit from current command view
  refresh         Do soft reset
  reset           Reset operation
  screen-length   Specify the lines displayed on one screen
  send            Send information to other user terminal interface
  ssh2            Establish a secure shell client connection
  super           Set the current user priority level
  telnet          Establish one TELNET connection
  terminal        Set the terminal line characteristics
  tracert         Trace route function
  undo            Cancel current setting
```

- Authenticate the usesr logging in to the device through Telnet, verify their passwords, and specify the user privilege levels as 2.

```
<Sysname> system-view
[Sysname] user-interface vty 0 4
[Sysname-ui-vty1] authentication-mode password
[Sysname-ui-vty0-4] set authentication password cipher 123
[Sysname-ui-vty0-4] user privilege level 2
```

By default, when users log in to the device through Telnet, they can use the commands of level 0 after passing the authentication. After you set the user privilege level under the user interface, when users log in to the device through Telnet, they need to input password **123**, and then they can use commands of levels 0, 1, and 2.

Switching user privilege level

Users can switch their user privilege level temporarily without logging out and disconnecting the current connection; after the switch, users can continue to configure the device without the need of relogin and reauthentication, but the commands that they can execute have changed. For example, if the current user privilege level is 3, the user can configure system parameters; after switching the user privilege level to 0, the user can only execute some simple commands, like **ping** and **tracert**, and only a few **display** commands. The switching of user privilege level is temporary, and effective for the current login; after the user relogs in, the user privilege restores to the original level.

To avoid misoperations, the administrators are recommended to log in to the device by using a lower privilege level and view device operating parameters, and when they have to maintain the device, they can switch to a higher level temporarily; when the administrators need to leave for a while or ask someone else to manage the device temporarily, they can switch to a lower privilege level before they leave to restrict the operation by others.

Users can switch from a high user privilege level to a low user privilege level without entering a password; when switching from a low user privilege level to a high user privilege level, only the console login users do not have to enter the password, and users that log in from VTY user interfaces need to enter the password for security's sake. This password is for level switching only and is different from the login password. If the entered password is incorrect or no password is configured, the switching fails. Therefore, before switching a user to a higher user privilege level, you should configure the password needed.

Follow these steps to switch user privilege level:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the password for switching the user privilege level	**super password** [**level** *user-level*] { **simple** | **cipher** } *password*	Required By default, no password is configured.
Exit to user view	**quit**	—
Switch the user privilege level	**super** [*level*]	Required When logging in to the device, a user has a user privilege level, which is decided by user interface or authentication user level.

Caution

- When you configure the password for switching user privilege level with the **super password** command, the user privilege level is 3 if no user privilege level is specified.
- The password for switching user privilege level can be displayed in both cipher text and simple text. You are recommended to adopt the former as the latter is easily cracked.

Modifying command level

All the commands in a view are defaulted to different levels, as shown in Table 1-3. The administrator can modify the command level based on users' needs to make users of a lower level use commands with a higher level or improve device security.

Follow these steps to modify the command level:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the command level in a specified view	**command-privilege level** *level* **view** *view command*	Required Refer to Table 1-3 for the default settings.

Caution

You are recommended to use the default command level or modify the command level under the guidance of professional staff; otherwise, the change of command level may bring inconvenience to your maintenance and operation, or even potential security problem.

Displaying and Maintaining Basic Configurations

To do…	Use the command…	Remarks
Display information on system version	**display version**	Available in any view
Display information on the system clock	**display clock**	
Display information on terminal users	**display users** [**all**]	
Display the valid configuration under current view	**display this** [**by-linenum**]	
Display clipboard information	**display clipboard**	
Display and save statistics of each module's running status	**display diagnostic-information**	

During daily maintenance or when the system is operating abnormally, you need to view each module's running status to find the problem. Therefore, you are required to execute the corresponding **display** commands one by one. To collect more information one time, you can execute the **display diagnostic-information** command in any view to display or save statistics of each module's running status. The execution of the **display diagnostic-information** command has the same effect as that of the commands **display clock**, **display version**, **display device**, and **display current-configuration**.

Note

- For the detailed description of the display users command, refer to *Login Commands* in the System Volume.
- Support for the display configure-user and display current-configuration command depends on the device model.
- The display commands discussed above are for the global configuration. Refer to the corresponding section for the display command for specific protocol and interface.

CLI Features

This section covers the following topics:

- Editing Features
- CLI Display
- Saving History Command
- Command Line Error Information

Introduction to CLI

CLI is an interaction interface between devices and users. Through CLI, you can configure your devices by entering commands and view the output information and verify your configurations, thus facilitating your configuration and management of your devices.

CLI provides the following features for you to configure and manage your devices:

- Hierarchical command protection where you can only execute the commands at your own or lower levels. Refer to Configuring User Privilege Levels and Command Levels for details.
- Easy access to on-line help by entering "?"
- Abundant debugging information for fault diagnosis
- Saving and executing commands that have been executed
- Fuzzy match for convenience of input. When you execute a command, you can input part of the characters in a keyword. However, to enable you to confirm your operation, the command can be executed only when you input enough characters to make the command unique. Take the commands **save**, **startup saved-configuration**, and **system-view** which start with **s** as an example. To save the current configuration, you need to input **sa** at least; to set the configuration file for next startup, you need to input **st s** at least; to enter system view, you need to input **sy** at least. You can press **Tab** to complement the command, or you can input the complete command.

Online Help with Command Lines

The following are the types of online help available with the CLI:

- Full help
- Fuzzy help

To obtain the desired help information, you can:

1) Enter **?** in any view to access all the commands in this view and brief description about them as well.

```
<Sysname> ?
User view commands:
  backup              Backup next startup-configuration file to TFTP server
  boot-loader         Set boot loader
  bootrom             Update/read/backup/restore bootrom
  cd                  Change current directory
  clock               Specify the system clock
  cluster             Run cluster command
  copy                Copy from one file to another
  debugging           Enable system debugging functions
  delete              Delete a file
  dir                 List files on a file system
  display             Show running system information
......omitted......
```

2) Enter a command and a **?** separated by a space. If **?** is at the position of a keyword, all the keywords are given with a brief description.

```
<Sysname> terminal ?
  debugging  Send debug information to terminal
  logging    Send log information to terminal
  monitor    Send information output to current terminal
  trapping   Send trap information to terminal
```

3) Enter a command and a **?** separated by a space. If **?** is at the position of a parameter, the description about this parameter is given.

```
<Sysname> system-view
[Sysname] interface vlan-interface ?
  <1-4094>  VLAN interface number
[Sysname] interface vlan-interface 1 ?
  <cr>
[Sysname] interface vlan-interface 1
```

Where, <cr> indicates that there is no parameter at this position. The command is then repeated in the next command line and executed if you press **Enter**.

4) Enter a character string followed by a **?**. All the commands starting with this string are displayed.

```
<Sysname> c?
   cd
   clock
   copy
```

5) Enter a command followed by a character string and a **?.** All the keywords starting with this string are listed.

```
<Sysname> display ver?
   version
```

6) Press **Tab** after entering the first several letters of a keyword to display the complete keyword, provided these letters can uniquely identify the keyword in this command. If several matches are found, the complete keyword which is matched first is displayed (the matching rule is: the letters next to the input letters are arranged in alphabetic order, and the letter in the first place is matched first.). If you repeatedly press **Tab**, all the keywords starting with the letter that you enter are displayed in cycles.

Synchronous Information Output

Synchronous information output refers to the feature that if the user's input is interrupted by system output, then after the completion of system output the system will display a command line prompt and your input so far, and you can continue your operations from where you were stopped.

You can use the **info-center synchronous** command to enable synchronous information output. For the detailed description of this function, refer to *Information Center Configuration* in the *System Volume*.

Undo Form of a Command

Adding the keyword **undo** can form an **undo** command. Almost every configuration command has an **undo** form. **undo** commands are generally used to restore the system default, disable a function or cancel a configuration. For example, the **info-center enable** command is used to enable the information center, while the **undo info-center enable** command is used to disable the information center. (By default, the information center is enabled.)

Editing Features

The CLI provides the basic command editing functions and supports multi-line editing. When you execute a command, the system automatically goes to the next line if the maximum length of the command is reached. You cannot press **Enter** to go to the next line; otherwise, the system will automatically execute the command. The maximum length of each command is 510 characters. Table 1-4 lists these functions.

Table 1-4 Edit functions

Key	Function
Common keys	If the editing buffer is not full, insert the character at the position of the cursor and move the cursor to the right.
Backspace	Deletes the character to the left of the cursor and move the cursor back one character.
Left-arrow key or **Ctrl+B**	The cursor moves one character space to the left.
Right-arrow key or **Ctrl+F**	The cursor moves one character space to the right.
Up-arrow key or **Ctrl+P**	Displays history commands
Down-arrow key or **Ctrl+N**	
Tab	Pressing **Tab** after entering part of a keyword enables the fuzzy help function. If finding a unique match, the system substitutes the complete keyword for the incomplete one and displays it in the next line; when there are several matches, if you repeatedly press **Tab**, all the keywords starting with the letter that you enter are displayed in cycles. If there is no match at all, the system does not modify the incomplete keyword and displays it again in the next line.

Note

When editing the command line, you can use other shortcut keys (For details, see Table 1-2) besides the shortcut keys defined in Table 1-4, or you can define shortcut keys by yourself. (For details, see Configuring CLI Hotkeys.)

CLI Display

By filtering the output information, you can find the wanted information effectively. If there is a lot of information to be displayed, the system displays the information in multiple screens. When the information is displayed in multiple screens, you can also filter the output information to pick up the wanted information.

Filtering the output information

The device provides the function to filter the output information. You can specify a regular expression (that is, the output rule) to search information you need.

You can use one of the following two ways to filter the output information:

- Input the keyword **begin**, **exclude**, or **include** as well as the regular expression at the command line to filter the output information.
- Input slash (/), minus (-), or plus (+) as well as the regular expression to filter the rest output information. Slash (/) is equal to the keyword **begin**, minus (-) is equal to the keyword **exclude**, and plus (+) is equal to the keyword **include**.

Keywords **begin**, **exclude**, and **include** have the following meanings:

- **begin**: Displays the line that matches the regular expression and all the subsequent lines.
- **exclude**: Displays the lines that do not match the regular expression.
- **include**: Displays only the lines that match the regular expression.

The regular expression is a string of 1 to 256 characters, case sensitive. It also supports special characters as shown in Table 1-5.

Table 1-5 Special characters in a regular expression

Character	Meaning	Remarks
^*string*	Starting sign, *string* appears only at the beginning of a line.	For example, regular expression “^user” only matches a string beginning with “user”, not “Auser”.
string$	Ending sign, *string* appears only at the end of a line.	For example, regular expression "user$” only matches a string ending with “user”, not “userA”.
.	Full stop, a wildcard used in place of any character, including single character, special character and blank.	For example, “.l” can match “vlan” or “mpls”.
*	Asterisk, used to match a character or character group before it zero or multiple times.	For example, “zo*” can match “z” and “zoo”; (zo)* can match “zo” and “zozo”.
+	Addition, used to match a character or character group one or multiple times before it	For example, “zo+” can match “zo” and “zoo”, but not “z”.
\| | Vertical bar, used to match the whole string on the left or right of it | For example, “def\|int” can only match a character string containing “def” or “int”. |
_ | Underline. If it is at the beginning or the end of a regular expression, it equals ^ or $; in other cases, it equals comma, space, round bracket, or curly bracket. | For example, “a_b” can match “a b” or “a(b”; “_ab” can only match a line starting with “ab”; “ab_” can only match a line ending with “ab”. |
- | Hyphen. It connects two values (the smaller one before it and the bigger one after it) to indicate a range together with [ ]. | For example, “1-9” means numbers from 1 to 9 (inclusive); “a-h” means from a to h (inclusive). |
[ ] | A range of characters, Matches any character in the specified range. | For example, [16A] can match a string containing any character among 1, 6, and A; [1-36A] can match a string containing any character among 1, 2, 3, 6, and A (with - being a hyphen).
“]” can be matched only when it is put at the beginning of [ ] if it is used as a common character in [ ], for example [ ]*string*]. There is no such limit on “[”. |

Character	Meaning	Remarks
( )	A character group. It is usually used with “+” or “*”.	For example, (123A) means a character group “123A”; “408(12)+” can match 40812 or 408121212. But it cannot match 408.
index	Repeats a specified character group for once. A character group refers to the string in () before \. *index* refers to the sequence number (starting from 1 from left to right) of the character group before \: if only one character group appears before \, then *index* can only be 1; if n character groups appear before *index*, then *index* can be any integer from 1 to n.	For example, (*string*)\1 means to repeat *string* for once, and (*string*)\1 must match a string containing *stringstring;* (*string1*)(*string2*)\2 means to repeat *string2* for once, and (*string1*)(*string2*)\2 must match a string containing *string1string2string2;* (*string1*)(*string2*)\1\2 means to repeat *string1* for once first*,* and then repeat string2 for once*, and* (*string1*)(*string2*)\1\2 must match a string containing *string1string2string1string2.*
[^]	Used to match any character not in a specified range.	For example, [^16A] means to match a string containing any character except 1, 6 or A, and the string can also contain 1, 6 or A, but cannot contain these three characters only. For example, [^16A] can match “abc” and “m16”, but not 1, 16, or 16A.
\<*string*	Used to match a character string starting with *string*.	For example, “\<do” can match word “domain” or string “doa”.
string\>	Used to match a character string ending with *string*.	For example, “do\>” can match word “undo” or string “abcdo”.
\b*character2*	Used to match *character1character2. character1* can be any character except number, letter or underline, and \b equals [^A-Za-z0-9_].	For example, \ba can match -a, with - represents *character1,* and a represents *character2;* while \ba cannot match “2a” or “ba”*.*
\B*character*	It must match a string containing *character*, and there can no spaces before *character.*	For example, “\Bt” can match “t” in “install”, but not “t” in “big top”.
*character1**w*	Used to match *character1character2. character2* must be a number, letter or underline, and *w* equals [^A-Za-z0-9_].	For example, “v\w” can match “vlan”, with “v” being *character1,* and “l” being *character2.* v\w can also match “service”, with “i” being *character2.*
\W	Equals \b.	For example, “\Wa” can match “-a”, with “-” representing *character1,* and “a” representing *character2;* while “\ba” cannot match “2a” or “ba”*.*
\	Escape character. If single special characters listed in this table follow \, the specific meanings of the characters will be removed.	For example, “\\” can match a string containing “\”, “\^” can match a string containing “^”, and “\\b” can match a string containing “\b”.

Multiple-screen output

When there is a lot of information to be output, the system displays the information in multiple screens. Generally, 24 lines are displayed on one screen, and you can also use the **screen-length** command to set the number of lines displayed on the next screen. (For the details of this command, refer to *Login*

Commands in the *System Volume*.) You can follow the step below to disable the multiple-screen output function of the current user.

To do…	Use the command…	Remarks
Disable the multiple-screen output function of the current user	**screen-length disable**	Required By default, a login user uses the settings of the **screen-length** command. The default settings of the **screen-length** command are: multiple-screen output is enabled and 24 lines are displayed on the next screen. This command is executed in user view, and therefore is applicable to the current user only. When a user re-logs in, the settings restore to the system default.

Display functions

CLI offers the following feature:

When the information displayed exceeds one screen, you can pause using one of the methods shown in Table 1-6.

Table 1-6 Display functions

Action	Function
Press **Space** when information display pauses	Continues to display information of the next screen page.
Press **Enter** when information display pauses	Continues to display information of the next line.
Press **Ctrl+C** when information display pauses	Stops the display and the command execution.
Ctrl+E	Moves the cursor to the end of the current line.
PageUp	Displays information on the previous page.
PageDown	Displays information on the next page.

Saving History Commands

The CLI can automatically save the commands that have been used lately to the history buffer. You can know the operations that have been executed successfully, invoke and repeatedly execute them as needed. By default, the CLI can save up to ten commands for each user. You can use the **history-command max-size** command to set the capacity of the history commands buffer for the current user interface (For the detailed description of the **history-command max-size** command, refer to *Login Commands* in the *System Volume).* In addition:

- The CLI saves the commands in the format that you have input, that is, if you input a command in its incomplete form, the saved history command is also incomplete.
- If you execute a command for multiple times successively, the CLI saves the earliest one. However, if you execute the different forms of a command, the CLI saves each form of this command. For example, if you execute the **display cu** command for multiple times successively, the CLI saves only one history command; if you execute the **display cu** command and then the **display current-configuration** command, the CLI saves two history commands.

Follow these steps to access history commands:

To do…	Use the key/command…	Result
View the history commands	**display history-command**	Displays the commands that you have entered
Access the previous history command	Up-arrow key or **Ctrl+P**	Displays the earlier history command, if there is any.
Access the next history command	Down-arrow key or **Ctrl+N**	Displays the next history command, if there is any.

Note

You may use arrow keys to access history commands in Windows 200X and XP Terminal or Telnet. However, the up-arrow and down-arrow keys are invalid in Windows 9X HyperTerminal, because they are defined in a different way. You can press **Ctrl+P** or **Ctrl+N** instead.

Command Line Error Information

The commands are executed only if they have no syntax error. Otherwise, error information is reported. Table 1-7 lists some common errors.

Table 1-7 Common command line errors

Error information	Cause
% Unrecognized command found at '^' position.	The command was not found.
	The keyword was not found.
	Parameter type error
	The parameter value is beyond the allowed range.
% Incomplete command found at '^' position.	Incomplete command
% Ambiguous command found at '^' position.	Ambiguous command,
Too many parameters	Too many parameters
% Wrong parameter found at '^' position.	Wrong parameter

Table of Contents

1 Device Management

When configuring device management, go to these sections for information you are interested in:

Device Management Overview

Through the device management function, you can view the current working state of a device, configure running parameters, and perform daily device maintenance and management.

Device Management Configuration Task List

Complete these tasks to configure device management:

Task	Remarks
Configuring the Exception Handling Method	Optional
Rebooting a Device	Optional
Configuring the Scheduled Automatic Execution Function	Optional
Specifying a Boot File for the Next Device Boot	Optional
Disabling Boot ROM Access	Optional
Upgrading Boot ROM	Optional
Configuring a Detection Interval	Optional
Clearing the 16-bit Interface Indexes Not Used in the Current System	Optional
Identifying and Diagnosing Pluggable Transceivers	Optional

Configuring the Exception Handling Method

When the system detects any software abnormality, it handles the situation with one of the following two methods:

- **reboot**: The system recovers itself through automatic reboot.
- **maintain**: The system maintains the current situation, and does not take any measure to recover itself. Therefore, you need to recover the system manually, such as reboot the system. Sometimes, it is difficult for the system to recover, or some prompts that are printed during the failure are lost after the reboot. In this case, you can use this method to maintain the abnormal state to locate problems and recover the system.

Follow these steps to configure exception handling method:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure exception handling method on all member devices	**system-failure** { **maintain** | **reboot** }	Optional By default, all member devices adopt the **reboot** method to handle exceptions.

Note

- After this command is configured, all the member devices adopt the same method to handle exceptions.
- The exception handling method is effective to the failed member device only, and does not influence the operations of other stack members.

Rebooting a Device

When a fault occurs to a running device, you can remove the fault by rebooting the device, depending on the actual situation. This operation equals to powering on the device after powering it off. It is mainly used to reboot a device in remote maintenance, without performing hardware reboot of the device.

According to the actual environment:

- You can reboot a member device or the whole system.
- You can trigger the immediate reboot through command lines, or set a time at which the device can automatically reboot, or you can also set a delay so that the device can automatically reboot in the delay.

Follow these steps to reboot a device:

To do…	Use the command…	Remarks
Reboot a member device or the whole stack system	**reboot** [**slot** *slot-number*]	Optional Available in user view
Enable the scheduled reboot function of all member devices and specify a specific reboot time and date	**schedule reboot at** *hh:mm* [*date*]	Optional The scheduled reboot function is disabled by default. Available in user view
Enable the scheduled reboot function of all member devices and specify a reboot waiting time	**schedule reboot delay** { *hh:mm* | *mm* }	

Caution

- Use the **save** command to save the current configuration before you reboot the device to avoid configuration lost. (For details of the **save** command, refer to *File System Management Configuration* in the *System Volume*.)
- Use the **display startup** command and the **display boot-loader** command to verify the configuration files and the startup file to be used at the next system startup before you reboot the device. (For details of the **display startup** command, refer to *File System Management Configuration* in the *System Volume*.)
- The precision of the rebooting timer is 1 minute. One minute before the rebooting time, the device will prompt “REBOOT IN ONE MINUTE” and will reboot in one minute.
- When you execute the **reboot** command on the master, if you specify the **slot** keyword, the member device with the specified number will reboot; if you do not specify the **slot** keyword, all member devices in the stack will reboot.
- Device reboot may result in the interruption of the ongoing services. Use these commands with caution.
- If a main boot file fails or does not exist, the device cannot be rebooted with the **reboot** command. In this case, you can re-specify a main boot file to reboot the device, or you can power off the device then power it on and the system automatically uses the backup boot file to restart the device.
- If you are performing file operations when the device is to be rebooted, the system does not execute the command for the sake of security.

Configuring the Scheduled Automatic Execution Function

The scheduled automatic execution function means that the system automatically executes a specified command at a specified time in a specified view. This function is used for scheduled system upgrade or configuration.

Follow these steps to configure the scheduled automatic execution function:

To do…	Use the command…	Remarks
Automatically execute the specified command at the specified time	**schedule job at** *time* [*date*] **view** *view command*	Optional If you configure the function, use either command Available in user view
Automatically execute the specified command after the specified delay	**schedule job delay** *time* **view** *view command*	

Note that:

- At present, you can specify user view and system view only. To automatically execute the specified command in another view or automatically execute multiple commands at a time, you can configure the system to automatically execute a batch file at the specified time (note that you must provide a complete file path for the system to execute the batch file.).
- The system does not check the values of the *view* and *command* arguments. Therefore, ensure the correctness of the *command* argument (including the correct format of *command* and the correct relationship between the *command* and *view* arguments).

- After the specified automatic execution time is reached, the system executes the specified command in the background without displaying any information except system information such as log, trap and debug.
- The system does not require any interactive information when it is executing the specified command. If there is information for you to confirm, the system automatically inputs **Y** or **Yes**; if characters need to be input, the system automatically inputs a default character string, or inputs an empty character string when there is no default character string.
- For the commands used to switch user interfaces, such as **telnet**, **ftp**, and **ssh2**, the commands used to switch views, such as **system-view**, **quit**, and the commands used to modify status of a user that is executing commands, such as **super**, the operation interface, command view and status of the current user are not changed after the automatic execution function is performed.
- If the system time is modified after the automatic execution function is configured, the scheduled automatic execution configuration turns invalid automatically.
- Only the last configuration takes effect if you execute the **schedule job** command repeatedly.
- After you configure this feature on the master, the configuration is not backed up to the slaves; after the change of the master, this configuration will be ineffective.

Specifying a Boot File for the Next Device Boot

A Boot ROM file, also known as the system software or device software, is an application file used to boot the device. A main boot file is used to boot a device and a backup boot file is used to boot a device only when a main boot file is unavailable. When multiple Boot ROM files are available on the storage media, you can specify a file for the next device boot by executing the following command.

Follow the step below to specify a boot file for the next device boot:

To do…	Use the command…	Remarks
Specify a boot file for a member device for the next device boot	**boot-loader file** *file-url* **slot** { **all** | *slot-number* } { **main** | **backup** }	Required Available in user view.

Caution

- To execute the **boot-loader** command successfully, you must save the file for the next device boot under the root directory of the storage media on a member device. You can copy or move a file to change the path of it to the root directory.
- The names of the files for the next boot of the master and slaves may be different, but the versions of the files must be the same; otherwise, a slave will reboot by using the master's boot file and join the stack again.

Disabling Boot ROM Access

By default, you can press **Ctrl+B** to enter the Boot ROM menu to configure the Boot ROM. However, this may bring security problems to the device. Therefore, the device provides the function of disabling the Boot ROM access to enhance security of the device. After this function is configured, no matter whether you press **Ctrl+B** or not, the system does not enter the Boot ROM menu, but enters the command line configuration interface directly.

In addition, you need to set the Boot ROM access password when you enter the Boot ROM menu for the first time to protect the Boot ROM against operations of illegal users.

You can use the **display startup** command to view the status of the Boot ROM access function. For the detailed description of the **display startup** command, refer to *File System Management* in the *System Volume*.

Follow the step below to disable Boot ROM access:

To do…	Use the command…	Remarks
Disable Boot ROM access	**undo startup bootrom-access enable**	Required By default, Boot ROM access is enabled. Available in user view

Upgrading Boot ROM

During the operation of the device, you can use the Boot ROM in the storage media to upgrade those that are running on the device.

Since the Boot ROM programs of the member devices vary with devices, users are easily confused to make mistakes when upgrading the Boot ROM. After the validity check function is enabled, the device will strictly check the Boot ROM upgrade files for correctness and version configuration information to ensure a successful upgrade.

Follow these steps to upgrade Boot ROM:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the validity check function when upgrading Boot ROM	**bootrom-update security-check enable**	Optional By default, the validity check function is enabled at the time of upgrading Boot ROM.
Return to user view	**quit**	—
Upgrade the Boot ROM program on a member device(s)	**bootrom update file** *file-url* **slot** *slot-number-lis*	Required Available in user view.

Note

Restart the device to validate the upgraded Boot ROM.

Configuring a Detection Interval

When detecting an exception on a port, the operation, administration and maintenance (OAM) module will automatically shut down the port. The device will detect the status of the port when a detection interval elapses. If the port is still shut down, the device will recover it.

Follow these steps to configure a detection interval:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a detection interval	**shutdown-interval** *time*	Optional The detection interval is 30 seconds by default.

Clearing the 16-bit Interface Indexes Not Used in the Current System

In practical networks, the network management software requires the device to provide a uniform, stable 16-bit interface index. That is, a one-to-one relationship should be kept between the interface name and the interface index in the same device.

For the purpose of the stability of an interface index, the system will save the 16-bit interface index when a board or logical interface is removed.

If you repeatedly insert and remove different subboards to create or delete a large number of logical interfaces, the interface indexes will be used up, which will result in interface creation failures. To avoid such a case, you can clear all 16-bit interface indexes saved but not used in the current system in user view.

After the above operation,

- For a re-created interface, the new interface index may not be consistent with the original one.
- For existing interfaces, their interface indexes remain unchanged.

Follow these steps to clear the 16-bit interface indexes not used in the current system:

To do…	Use the command…	Remarks
Clear the 16-bit interface indexes saved but not in use in the current systems of all member devices	**reset unused porttag**	Required Available in user view

Caution

A confirmation is required when you execute this command. If you fail to make a confirmation within 30 seconds or enter **N** to cancel the operation, the command will not be executed.

Identifying and Diagnosing Pluggable Transceivers

Introduction to pluggable transceivers

At present, four types of pluggable transceivers are commonly used, as shown in Table 1-1. They can be further divided into optical transceivers and electrical transceivers based on transmission medium.

Table 1-1 Commonly used pluggable transceivers

Transceiver type	Application environment	Whether can be an optical transceiver	Whether can be an electrical transceiver
SFP (Small Form-factor Pluggable)	Generally used for 100M/1000M Ethernet interfaces or POS 155M/622M/2.5G interfaces	Yes	Yes
GBIC (Gigabit Interface Converter)	Generally used for 1000M Ethernet interfaces	Yes	Yes
XFP (10-Gigabit small Form-factor Pluggable)	Generally used for 10G Ethernet interfaces	Yes	No
XENPAK (10-Gigabit Ethernet Transceiver Package)	Generally used for 10G Ethernet interfaces	Yes	Yes

Identifying pluggable transceivers

As pluggable transceivers are of various types and from different vendors, you can use the following commands to view the key parameters of the pluggable transceivers, including transceiver type, connector type, central wavelength of the laser sent, transfer distance and vendor name or name of the vendor who customizes the transceivers to identify the pluggable transceivers.

Follow these steps to identify pluggable transceivers:

To do…	Use the command…	Remarks
Display key parameters of the pluggable transceiver(s)	**display transceiver interface** [*interface-type interface-number*]	Available for all pluggable transceivers.
Display part of the electrical label information of the anti-spoofing transceiver(s) customized	**display transceiver manuinfo interface** [*interface-type interface-number*]	Available for anti-spoofing pluggable transceiver(s) customized by H3C only.

- You can use the **Vendor Name** field in the prompt information of the **display transceiver** command to identify an anti-spoofing pluggable transceiver customized by H3C. If the field is **H3C**, it is considered an H3C-customized pluggable transceiver.
- Electrical label information is also called permanent configuration data or archive information, which is written to the storage component of a board during device debugging or testing. The information includes name of the board, device serial number, and vendor name or name of the vendor who customizes the transceiver.

Diagnosing pluggable transceivers

The system outputs alarm information for you to diagnose and troubleshoot faults of pluggable transceivers. Optical transceivers customized by H3C also support the digital diagnosis function, which monitors the key parameters of a transceiver, such as temperature, voltage, laser bias current, TX

power, and RX power. When these parameters are abnormal, you can take corresponding measures to prevent transceiver faults.

Follow these steps to diagnose pluggable transceivers:

To do…	Use the command…	Remarks
Display the current alarm information of the pluggable transceiver(s)	**display transceiver alarm interface** [*interface-type interface-number*]	Available for all pluggable transceivers.
Display the currently measured value of the digital diagnosis parameters of the anti-spoofing optical transceiver(s) customized	**display transceiver diagnosis interface** [*interface-type interface-number*]	Available for anti-spoofing pluggable optical transceiver(s) customized by H3C only.

Displaying and Maintaining Device Management Configuration

Follow these steps to display and maintain device management configuration:

To do…	Use the command…	Remarks
Display information of the boot file	**display boot-loader** [**slot** *slot-number*]	Available in any view
Display the statistics of the CPU usage	**display cpu-usage** [*number* [*offset*] [**verbose**] [**from-device**]]	Available in any view
Display history statistics of the CPU usage in a chart	**display cpu-usage history** [**task** *task-id*] [**slot** *slot-number* [**cpu** *cpu-number*]]	Available in any view
Display information about a board, subboard, CF board, USB or hardware on the device	**display device** [[**shelf** *shelf-number*] [**frame** *frame-number*] [**slot** *slot-number* [**subslot** *subslot-number*]] | **verbose**]	Available in any view
Display electrical label information of the device	**display device manuinfo**	Available in any view
Display the temperature information of devices	**display environment**	Available in any view
Display the operating state of fans in the device	**display fan** [**slot** *slot-number* [*fan-id*]]	Available in any view
Display the usage of the memory of the device	**display memory** [**slot** *slot-number* [**cpu** *cpu-number*]]	Available in any view
Display the power state of a device	**display power** [**slot** *slot-number* [*power-id*]]	Available in any view
Display state of the RPS	**display rps** [**slot** *slot-number* [*rps-id*]]	Available in any view
Display the reboot mode of a device	**display reboot-type** [**slot** *slot-number*]	Available in any view
Display the reboot time of a device	**display schedule reboot**	Available in any view
Display detailed configurations of the scheduled automatic execution function	**display schedule job**	Available in any view

To do…	Use the command…	Remarks
Display the exception handling methods	**display system-failure**	Available in any view

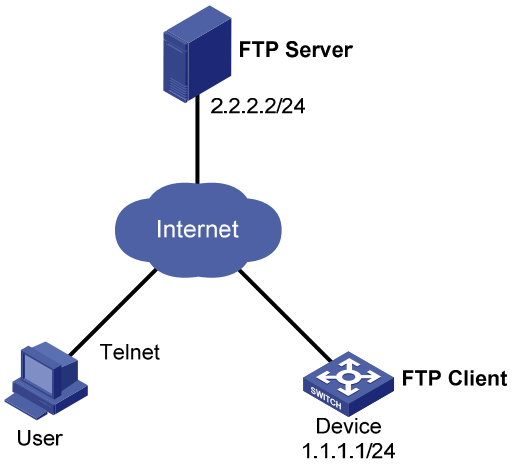
Device Management Configuration Examples

Remote Scheduled Automatic Upgrade Configuration Example (Centralized Device)

Network requirement

- As shown in Figure 1-1, the current software version is **soft-version1** for Device. Upgrade the software version of Device to **soft-version2** and configuration file to **new-config** at a time when few services are processed (for example, at 3 am) through remote operations.
- The newest application **soft-version2.bin** and the newest configuration file **new-config.cfg** are both saved under the **aaa** directory of the FTP server.
- The IP address of Device is 1.1.1.1/24, the IP address of the FTP server is 2.2.2.2/24, and the FTP server is reachable.
- User can log in to Device via Telnet and a route exists between User and Device.

Figure 1-1 Network diagram for remote scheduled automatic upgrade



Configuration procedure

1) Configuration on the FTP server (Note that configurations may vary with different types of servers)

- Set the access parameters for the FTP client (including enabling the FTP server function, setting the FTP username to **aaa** and password to **hello**, and setting the user to have access to the **flash:/aaa** directory).

```
<FTP-Server> system-view
[FTP-Server] ftp server enable
[FTP-Server] local-user aaa
[FTP-Server-luser-aaa] password cipher hello
[FTP-Server-luser-aaa] service-type ftp
[FTP-Server-luser-aaa] authorization-attribute work-directory flash:/aaa
```

- Use text editor on the FTP server to edit batch file **auto-update.txt**. The following is the content of the batch file:

```
return
```

```
startup saved-configuration new-config.cfg
boot-loader file soft-version2.bin main
reboot
```

2) Configuration on Device

Log in to the FTP server (note that the prompt may vary with servers.)

```
<Device> ftp 2.2.2.2
Trying 2.2.2.2 ...
Press CTRL+K to abort
Connected to 2.2.2.2.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(2.2.2.2:(none)):aaa
331 Give me your password, please
Password:
230 Logged in successfully
[ftp]
```

Download file **auto-update.txt** on the FTP server.

```
[ftp] ascii
[ftp] get auto-update.txt
```

Download file **new-config.cfg** on the FTP server.

```
[ftp]get new-config.cfg
```

Download file **soft-version2.bin** on the FTP server.

```
[ftp] binary
[ftp] get soft-version2.bin
[ftp] bye
<Device>
```

Modify the extension of file **auto-update.txt** as **.bat**.

```
<Device> rename auto-update.txt auto-update.bat
```

To ensure correctness of the file, you can use the **more** command to view the content of the file.

Execute the scheduled automatic execution function to enable the device to be automatically upgraded at 3 am.

```
<Device> schedule job at 03:00 view system execute auto-update.bat
Info: Command execute auto-update.bat in system view will be executed at 03:00 12/11/2007(in
12 hours and  0 minutes).
```

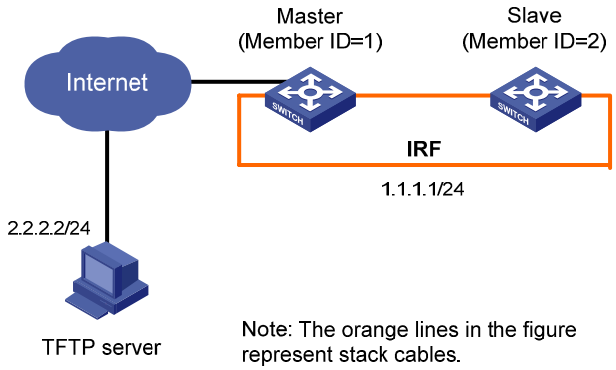
Remote Scheduled Automatic Upgrade Configuration Example (Centralized Stacking Device)

Network requirement

- As shown in Figure 1-2, the current software version is **soft-version1** for the IRF stack system. Upgrade the software version of the stack system to **soft-version2** and configuration file to **new-config**.
- The newest application **soft-version2.bin** and the newest configuration file **new-config.cfg** are both saved under the TFTP server.

- The IP address of the IRF stack system is 1.1.1.1/24, the IP address of the TFTP server is 2.2.2.2/24, and the TFTP server is reachable.

Figure 1-2 Network diagram for remote scheduled automatic upgrade



Configuration procedure

1) Configuration on the TFTP server (Note that configurations may vary with different types of servers)

Obtain the boot file and configuration file through legitimate channels, such as the official website of 3Com, agents, and technical staff. Save these files under the working path of the TFTP server for the access of the TFTP clients.

2) Configuration on the IRF stack members

Download file **new-config.cfg** on the TFTP server to Master (Note that configurations may vary with different types of servers).

```
<IRF> tftp 2.2.2.2 get new-config.cfg
  ..
  File will be transferred in binary mode
  Downloading file from remote TFTP server, please wait.....
  TFTP:      917 bytes received in 1 second(s)

File downloaded successfully.
```

Download file **new-config.cfg** to Slave with the member ID of 2.

```
<IRF> tftp 2.2.2.2 get new-config.cfg slot2#flash:/new-config.cfg
```

Download file **soft-version2.bin** on the TFTP server to Master and Slave.

```
<IRF> tftp 2.2.2.2 get soft-version2.bin
  ...
  File will be transferred in binary mode
  Downloading file from remote TFTP server, please wait............
  TFTP: 10058752 bytes received in 141 second(s)

  File downloaded successfully.
<IRF> tftp 2.2.2.2 get soft-version2.bin slot2#flash:/soft-version2.bin
```

Specify file **new-config.cfg** as the configuration file for the next boot for all members.

```
<IRF> startup saved-configuration new-config.cfg main
Please wait ...
Setting the master board ...
```

```
... Done!
Setting the slave board ...
Slot 2:
 Set next configuration file successfully
```

Specify file **soft-version2.bin** as the boot file for the next boot for all members.

```
<IRF> boot-loader file soft-version2.bin slot all main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 1!
The specified file will be used as the main boot file at the next reboot on slot 2!
```

Reboot the device. The software version is upgraded now.

```
<IRF> reboot
```

Table of Contents

1 File System Management Configuration

When configuring file system management, go to these sections for information you are interested in:

- File System Management
- Configuration File Management
- Displaying and Maintaining Device Configuration

File System Management

This section covers these topics:

- File System Overview
- Filename Formats
- Directory Operations
- File Operations
- Batch Operations
- Storage Medium Operations
- Setting File System Prompt Modes
- File System Operations Example

File System Overview

A major function of the file system is to manage storage media. It allows you to perform operations such as directory create and delete, and file copy and display. If an operation, delete or overwrite for example, causes problems such as data loss or corruption, the file system will prompt you to confirm the operation by default.

Depending on the managed object, file system operations fall into Directory Operations, File Operations, Batch Operations, Storage Medium Operations, and Setting File System Prompt Modes.

Filename Formats

When you specify a file, you must enter the filename in one of the following formats.

Filename formats:

Format	Description	Length	Example
file-name	Specifies a file under the current working directory.	1 to 91 characters	a.txt: Indicates that a file named **a.txt** is under the current working directory. If the current working directory is on the master, **a.txt** represents file **a.txt** on the master; if the current working directory is on a slave, **a.txt** represents file **a.txt** on the slave.

Format	Description	Length	Example
path/*file-name*	Specifies a file in the specified folder under the current working directory. *path* represents the folder name. You can specify multiple folders, indicating a file under a multi-level folder.	1 to 135 characters	test/a.txt: Indicates that a file named **a.txt** is in the **test** folder under the current working directory.
drive:/[*path*]/*file-name*	Specifies a file in the specified storage medium on the device. *drive* represents the storage medium name. The Switch 4800G use flashes as their storage media. The storage medium on the master is flash; the storage medium on a slave is slot2#flash, where 2 represents the member ID of the slave. You can use the **display stack** command to view the correspondence between a device and its member ID.	1 to 135 characters	flash:/test/a.txt: Indicates that a file named **a.txt** is in the **test** folder under the root directory of the flash memory on the master. To read and write the a.txt file under the root directory of the flash on a slave (with the member ID 2), input **slot2#flash:/a.txt** for the filename.

Note

For the Switch 4800G, when you specify a configuration file (.cfg file), startup file (.bin file), or Boot ROM file by inputting its name in the format of *drive*:/[*path*]/*file-name*), the total length of the name cannot exceed 63 characters.

Directory Operations

Directory operations include creating/removing a directory, displaying the current working directory, displaying the specified directory or file information, and so on.

Displaying directory information

To do…	Use the command…	Remarks
Display directory or file information	**dir** [**/all**] [*file-url*]	Required Available in user view

Displaying the current working directory

To do…	Use the command…	Remarks
Display the current working directory	**pwd**	Required Available in user view

Changing the current working directory

To do…	Use the command…	Remarks
Change the current working directory	**cd** { *directory* | **..** | **/** }	Required Available in user view

Creating a directory

To do…	Use the command…	Remarks
Create a directory	**mkdir** *directory*	Required Available in user view

Removing a directory

To do…	Use the command…	Remarks
Remove a directory	**rmdir** *directory*	Required Available in user view

Note

- The directory to be removed must be empty, meaning that before you remove a directory, you must delete all the files and the subdirectory under this directory. For file deletion, refer to the **delete** command; for subdirectory deletion, refer to the **rmdir** command.
- After you execute the **rmdir** command successfully, the files in the recycle bin under the directory will be automatically deleted.

File Operations

File operations include displaying the specified directory or file information; displaying file contents; renaming, copying, moving, removing, restoring, and deleting files.

Note

You can create a file by copying, downloading or using the **save** command.

Displaying file information

To do…	Use the command…	Remarks
Display file or directory information	**dir** [/**all**] [*file-url*]	Required Available in user view

Displaying the contents of a file

To do…	Use the command…	Remarks
Display the contents of a file	**more** *file-url*	Required Currently only a .txt file can be displayed. Available in user view

Renaming a file

To do…	Use the command…	Remarks
Rename a file	**rename** *fileurl-source fileurl-dest*	Required Available in user view

Copying a file

To do…	Use the command…	Remarks
Copy a file	**copy** *fileurl-source fileurl-dest*	Required Available in user view

Moving a file

To do…	Use the command…	Remarks
Move a file	**move** *fileurl-source fileurl-dest*	Required Available in user view

Deleting a file

To do…	Use the command…	Remarks
Move a file to the recycle bin or delete it permanently	**delete** [/**unreserved**] *file-url*	Required Available in user view

Caution

- The files in the recycle bin still occupy storage space. To delete a file in the recycle bin, you need to execute the **reset recycle-bin** command in the directory that the file originally belongs. It is recommended to empty the recycle bin timely with the **reset recycle-bin** command to save storage space.
- The **delete /unreserved** *file-url* command deletes a file permanently and the action cannot be undone. Execution of this command equals that you execute the **delete** *file-url* command and then the **reset recycle-bin** command in the same directory.

Restoring a file from the recycle bin

To do…	Use the command…	Remarks
Restore a file from the recycle bin	**undelete** *file-url*	Required Available in user view

Emptying the recycle bin

To do…	Use the command…	Remarks
Enter the original working directory of the file to be deleted	**cd** { *directory* | **..** | **/** }	Optional If the original directory of the file to be deleted is not the current working directory, this command is required. Available in user view
Delete the file under the current directory and in the recycle bin	**reset recycle-bin** [**/force**]	Required Available in user view

Batch Operations

A batch file is a set of executable commands. Executing a batch file equals executing the commands in the batch file one by one.

The following steps are recommended to execute a batch file:

1) Edit the batch file on your PC.
2) Download the batch file to the device. If the suffix of the file is not **.bat**, use the **rename** command to change the suffix to **.bat**.
3) Execute the batch file.

Follow the steps below to execute a batch file:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Execute a batch file	**execute** *filename*	Required

Caution

Execution of a batch file does not guarantee the successful execution of every command in the batch file. If a command has error settings or the conditions for executing the command are not satisfied, the system will skip the command to the next one.

Storage Medium Operations

Managing space of the storage medium

When some space of a storage medium becomes inaccessible due to abnormal operations for example, you can use the **fixdisk** command to restore the space of the storage medium. The execution of the **format** command will format the storage medium, and all the data on the storage medium will be deleted.

Use the following commands to manage the storage medium space:

To do…	Use the command…	Remarks
Restore the space of a storage medium	**fixdisk** *device*	Optional Available in user view
Format a storage medium	**format** *device*	Optional Available in user view

Caution

- When you format a storage medium, all the files stored on it are erased and cannot be restored. In particular, if there is a startup configuration file on the storage medium, formatting the storage medium results in loss of the startup configuration file.
- You can execute the **fixdisk** command for the storage medium on the master, but you cannot execute the command for a storage medium on the slave.

Setting File System Prompt Modes

The file system provides the following two prompt modes:

- **alert**: In this mode, the system warns you about operations that may bring undesirable consequences such as file corruption or data loss.
- **quiet**: In this mode, the system does not prompt confirmation for any operation.

To prevent undesirable consequence resulting from misoperations, the **alert** mode is preferred.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the operation prompt mode of the file system	**file prompt** { **alert** | **quiet** }	Optional The default is **alert**.

File System Operations Example

Display the files and the subdirectories under the current directory.

```
<Sysname> dir
Directory of flash:/
   0     -rw-  10197108  Jul 17 2007 18:30:04   s4800g.bin
   1     -rw-    478164  Apr 26 2007 14:40:07   s4800g _505.btm
   2     -rw-      1586  Aug 24 2007 12:00:03   startup.cfg
   3     -rw-  11053555  Aug 22 2007 17:25:16   s4800g1.bin
   4     drw-         -  Apr 26 2007 19:58:11   test
31496 KB total (9943 KB free)
```

Create a new folder called **mytest** under the test directory.

```
<Sysname> cd test
<Sysname> mkdir mytest
%Created dir flash:/test/mytest.
```

Display the current working directory.

```
<Sysname> pwd
flash:/test
```

Display the files and the subdirectories under the test directory.

```
<Sysname> dir
Directory of flash:/test/

0     drw-         -  Apr 26 2007 19:58:39   mytest

31496 KB total (9942 KB free)
```

Return to the upper directory.

```
<Sysname> cd ..
```

Display the current working directory.

```
<Sysname> pwd
flash:
```

Configuration File Management

The device provides the configuration file management function with a user-friendly command line interface (CLI) for you to manage the configuration files conveniently.

This section covers these topics:

- Configuration File Overview

Configuration File Overview

A configuration file saves the device configurations in command lines in text format. You can view configuration information conveniently through configuration files.

Types of configuration

The configuration of a device falls into two types:

- Startup configuration, a configuration file used for initialization when the device boots. If this file does not exist, the system boots using null configuration, that is, using the default parameters.
- Current configuration, which refers to the currently running configuration of the system. The current configuration may include the startup configuration if the startup configuration is not modified during system operation, and it also includes the new configuration added during the system operation. The current configuration is stored in the temporary storage medium of the device, and will be removed when the device reboots if not saved.

Format of a configuration file

A configuration file is saved as a text file. It:

- Saves configuration in the form of commands.
- Saves only non-default configuration settings.
- Lists commands in sections by views, usually in the order of system view, interface view, and routing protocol view. Sections are separated with one or multiple blank lines or comment lines that start with a pound sign #.
- Ends with a return.

Coexistence of multiple configuration files

Multiple configuration files can be stored on a storage medium of a device. You can save the configuration used in different environments as different configuration files. In this case, when the device moves between these networking environments, you just need to specify the corresponding configuration file as the startup configuration file for the next boot of the device and restart the device, so that the device can adapt to the network rapidly, saving the configuration workload.

A device boots using only one configuration file. However, you can specify two startup configuration files, main and backup startup configuration file, for the next startup of the device as needed and when the device supports this feature. When the device boots, the system uses the main startup configuration file, and if the main startup configuration file is corrupted or lost, the system will use the backup startup configuration file for device boot and configuration. The devices supporting the configuration of the main and backup startup configuration files, compared with the devices that do not support this feature, are more secure and reliable.

At a moment, there are at most one main startup configuration file and one backup startup configuration file. You can specify neither of the two files (displayed as NULL), or specify the two files as the same configuration file.

You can specify the main and backup startup configuration files for the next boot of the device in the following two methods:

- Specify them when saving the current configuration. For detailed configuration, refer to Saving the Current Configuration.
- Specify them when specifying the startup configuration file for the next system startup. For detailed configuration, refer to Specifying a Startup Configuration File for the Next System Startup.

Startup with the configuration file

The device takes the following steps when it boots:

1) If the main startup configuration file exists, the device initializes with this configuration file.
2) If the main startup configuration file does not exist but the backup startup configuration file exists, the device initializes with the backup startup configuration file.
3) If neither the main nor the backup startup configuration file exists, the device will boot with null configuration (boot with null configuration means to boot with the factory default configuration).

Saving the Current Configuration

Introduction

You can modify the current configuration on your device using command line interface. However, the current configuration is temporary. To make the modified configuration take effect at the next boot of the device, you must save the current configuration to the startup configuration file before the device reboots.

Complete these tasks to save the current configuration:

Task	Remarks
Enabling configuration file auto-save	Optional
Modes in saving the configuration	Required

Enabling configuration file auto-save

- After the configuration file auto-save function is enabled, when you save the current configuration by executing the **save** [**safely**] [**backup** | **main**] command or executing the **save** *filename* **all** command and then pressing **Enter**, the master and a slave will automatically save the current configuration to the specified configuration file, and use the file as the configuration file for the next startup, thus keeping the consistency of the configuration files on the master and the slave.
- If the configuration file auto-save function is not enabled, when you save the current configuration by executing the **save** [**safely**] [**backup** | **main**] command or executing the **save** *filename* **all** command and then pressing **Enter**, only the master will automatically save the current configuration to the specified configuration file, and use the file as the configuration file for the next startup; the slaves will neither save the configuration file nor configure the file for the next startup.

Follow these steps to configure the configuration file auto-save function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable configuration file auto-save	**slave auto-update config**	Optional Enabled by default.

Modes in saving the configuration

- Fast saving mode. This is the mode when you use the **save** command without the **safely** keyword. The mode saves the file more quickly but is likely to lose the existing configuration file if the device reboots or the power fails during the process.
- Safe mode. This is the mode when you use the **save** command with the **safely** keyword. The mode saves the file more slowly but can retain the configuration file in the device even if the device reboots or the power fails during the process.

The fast saving mode is suitable for environments where power supply is stable. The safe mode, however, is preferred in environments where stable power supply is unavailable or remote maintenance is involved.

To do…	Use the command…	Remarks
Save the current configuration to the specified file, but the configuration file will not be set as the file for the next startup	**save** *file-url* [**all** | **slot** *slot-number*]	Required Use either command Available in any view
Save the current configuration to the root directories of the storage media of all the member devices and specify the file as the startup configuration file that will be used at the next system startup	**save** [**safely**] [**backup** | **main**]	

Note

- The configuration file must be with extension **.cfg**.
- Whether the **save** [**safely**] [**backup** | **main**] command or the **save** *filename* **all** command+**Enter** takes effect on all the member devices or on the master only depends on whether the configuration file auto-save function is enabled. For the configuration file auto-save function, refer to Enabling configuration file auto-save. (centralized stacking device)
- The execution of the **save** [**safely**] and **save** [**safely**] **main** commands has the same effect: The system will save the current configuration and specify the configuration file as the main startup configuration file to be used at the next system startup.
- During the execution of the **save** [**safely**] [**backup** | **main**] command, the startup configuration file to be used at the next system startup may be lost if the device reboots or the power supply fails. In this case, the device will boot with the null configuration, and after the device reboots, you need to re-specify a startup configuration file for the next system startup (refer to Specifying a Startup Configuration File for the Next System Startup).

Specifying a Startup Configuration File for the Next System Startup

A startup configuration file is the configuration file to be used at the next system startup. You can specify a configuration file as the startup configuration file to be used at the next system startup in the following two ways:

- Use the **save** command. If you save the current configuration to the specified configuration file in the interactive mode, the system automatically sets the file as the main startup configuration file to be used at the next system startup.
- Use the command dedicated to specify a startup configuration file, which is described in the following table:

Follow the step below to specify a configuration file as the startup configuration file for the next system startup:

To do…	Use the command…	Remarks
Specify a startup configuration file for the next system startup of all the member devices	**startup saved-configuration** *cfgfile* [**backup** | **main**]	Required Available in user view

Caution

A configuration file must use **.cfg** as its extension name and the startup configuration file must be saved under the root directory of the storage medium.

Backing Up the Startup Configuration File

The backup function allows you to copy the startup configuration file to be used at the next system startup from the device to the TFTP server for backup.

The backup operation backs up the startup configuration file to the TFTP server.

Follow the step below to back up the startup configuration file to be used at the next system startup:

To do…	Use the command…	Remarks
Back up the configuration file to be used at the next system startup to the specified TFTP server	**backup startup-configuration to** *dest-addr* [*dest- filename*]	Required Available in user view

Note

Before the backup operation, you should:

- Ensure that the server is reachable, the server is enabled with TFTP service, and the client has permission to read and write.
- Use the **display startup** command (in user view) to see whether you have set the startup configuration file, and use the **dir** command to see whether this file exists. If the file is set as NULL or does not exist, the backup operation will fail.

Deleting the Startup Configuration File for the Next Startup

You can delete the startup configuration file to be used at the next system startup using commands. On a device that has the main and backup startup configuration files, you can choose to delete either the

main or backup startup configuration file. However, in the case that the main and backup startup configuration files are the same, if you perform the delete operation for once, the system will not delete the configuration file but only set the corresponding startup configuration file (main or backup, according to which one you specified in the command) to NULL.

You may need to delete the startup configuration file for the next startup for one of these reasons:

- After you upgrade system software, the existing configuration file does not match the new system software.
- The configuration file is corrupted (often caused by loading a wrong configuration file).

After the startup configuration file is deleted, the system will use the null configuration when the device reboots.

Follow the step below to delete the startup configuration file for the next startup:

To do…	Use the command…	Remarks
Delete the startup configuration file for the next startup from the storage medium	**reset saved-configuration** [**backup** | **main**]	Required Available in user view

Caution

This command will permanently delete the configuration file from all the member devices. Use it with caution.

Restoring the Startup Configuration File

- The restore function allows you to copy a configuration file from TFTP server to the root directory of the storage media of all the member devices and specify the file as the startup configuration file to be used at the next system startup.

Follow the step below to restore the startup configuration file to be used at the next system startup:

To do…	Use the command…	Remarks
Restore the startup configuration file to be used at the next system startup	**restore startup-configuration from** *src-addr src-filename*	Required Available in user view

Note

- The restore operation restores the main startup configuration file.
- Before restoring a configuration file, you should ensure that the server is reachable, the server is enabled with TFTP service, and the client has read and write permission.
- After the command is successfully executed, you can use the **display startup** command (in user view) to verify that the filename of the configuration file to be used at the next system startup is the same with that specified by the *filename* argument, and use the **dir** command to verify that the restored startup configuration file exists.

Displaying and Maintaining Device Configuration

To do…	Use the command…	Remarks
Display the currently running configuration file saved on the storage medium of the device	**display saved-configuration** [**by-linenum**]	Available in any view
Display the configuration files for this and the next system startup	**display startup**	Available in any view
Display the validated configuration in current view	**display this** [**by-linenum**]	Available in any view
Display the current configuration	**display current-configuration** [[**configuration** [*configuration*] | **interface** [*interface-type*] [*interface-number*]] [**by-linenum**] [**|** { **begin** | **include** | **exclude** } *text*]]	Available in any view

Note

For detailed description of the **display this** and **display current-configuration** commands, refer to *Basic System Configuration Commands* in the *System Volume.*

2 FTP Configuration

When configuring FTP, go to these sections for information you are interested in:

FTP Overview

Introduction to FTP

The File Transfer Protocol (FTP) is an application layer protocol for sharing files between server and client over a TCP/IP network.

FTP uses TCP ports 20 and 21 for file transfer. Port 20 is used to transmit data, and port 21 to transmit control commands. Refer to RFC 959 for details of FTP basic operation.

FTP transfers files in two modes:

- Binary mode for program file transmission, like files with the suffixes **.app**, **.bin**, or **.btm.**
- ASCII mode for text file transmission, like files with the suffixes **.txt**, **.bat**, or **.cfg**.

Operation of FTP

FTP adopts the client/server model. Your device can function either as the client or as the server (as shown in Figure 2-1).

- When the device serves as the FTP client, the user first connects to the device from a PC through Telnet or an emulation program, and then executes the **ftp** command to establish a connection to the remote FTP server and gain access to the files on the server.
- When the device serves as the FTP server, FTP clients (users running the FTP client program) log in to the device to access files on the device (the administrator must configure the IP address of the device as the FTP server IP address before user login).

Figure 2-1 Network diagram for FTP

When the device serves as the FTP client, you need to perform the following configuration:

Table 2-1 Configuration when the device serves as the FTP client

Device	Configuration	Remarks
Device (FTP client)	Use the **ftp** command to establish the connection to the remote FTP server	If the remote FTP server supports anonymous FTP, the device can log in to it directly; if not, the device must obtain the FTP username and password first to log in to the remote FTP server.
PC (FTP server)	Enable FTP server on the PC, and configure the username, password, user privilege level, and so on.	—

When the device serves as the FTP server, you need to perform the following configuration:

Table 2-2 Configuration when the device serves as the FTP server

Device	Configuration	Remarks
Device (FTP server)	Enable the FTP server function	Disabled by default. You can use the **display ftp-server** command to view the FTP server configuration on the device.
	Configure authentication and authorization	Configure the username, password, authorized working directory for an FTP user. The device does not support anonymous FTP for security reasons. Therefore, you must use a valid username and password. By default, authenticated users can access the root directory of the device.
	Configure the FTP server operating parameters	Parameters such as the FTP connection timeout time
PC (FTP client)	Use the FTP client program to log in to the FTP server.	You can log in to the FTP server only after you input the correct FTP username and password.

Caution

- The FTP function is available when a reachable route exists between the FTP server and the FTP client.
- When you use IE to log in to the device serving as the FTP server, part of the FTP functions is not available. This is because multiple connections are established during the login process but the device supports only one connection at a time.

Configuring the FTP Client

Establishing an FTP Connection

To access an FTP server, an FTP client must establish a connection with the FTP server. Two ways are available to establish a connection: using the **ftp** command to establish the connection directly; using the **open** command in FTP client view.

Source address binding means to configure an IP address on a stable interface such as a loopback interface or Dialer interface, and then use this IP address as the source IP address of an FTP connection. The source address binding function simplifies the configuration of ACL rules and security policies. You just need to specify the source or destination address argument in an ACL rule as this address to filter inbound and outbound packets on the device, ignoring the difference between interface IP addresses as well as the affect of interface statuses. You can configure the source address by configuring the source interface or source IP address. The primary IP address configured on the source interface is the source address of the transmitted packets. The source address of the transmitted packets is selected following these rules:

- If no source address is specified, the FTP client uses the IP address of the interface determined by the matched route as the source IP address to communicate with an FTP server.
- If the source address is specified with the **ftp client source** or **ftp** command, this source address is used to communicate with an FTP server.
- If you use the **ftp client source** command and the **ftp** command to specify a source address respectively, the source address specified with the **ftp** command is used to communicate with an FTP server.

The source address specified with the **ftp client source** command is valid for all FTP connections and the source address specified with the **ftp** command is valid only for the current FTP connection.

Follow these steps to establish an FTP connection (In IPv4 networking):

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the source address of the FTP client	**ftp client source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }	Optional A device uses the IP address of the interface determined by the matched route as the source IP address to communicate with the FTP server by default.
Exit to system view	**quit**	—
Log in to the remote FTP server directly in user view	**ftp** [*server-address* [*service-port*] [**source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }]]	Use either approach. The **ftp** command is available in user view; and the **open** command is available in FTP client view.
Log in to the remote FTP server indirectly in FTP client view	**ftp**	
	open *server-address* [*service-port*]	

Note

- If no primary IP address is configured on the specified source interface, no FTP connection can be established.
- If you use the **ftp client source** command to first configure the source interface and then the source IP address of the transmitted packets, the newly configured source IP address will take effect instead of the current source interface, and vice versa.

Follow these steps to establish an FTP connection (In IPv6 networking):

To do…	Use the command…	Remarks
Log in to the remote FTP server directly in user view	**ftp ipv6** [*server-address* [*service-port*] [**source ipv6** *source-ipv6-address*] [**-i** *interface-type interface-number*]]	Use either approach. The **ftp ipv6** command is available in user view; and the **open ipv6** command is available in FTP client view.
Log in to the remote FTP server indirectly in FTP client view	**ftp ipv6**	
	open ipv6 *server-address* [*service-port*] [**-i** *interface-type interface-number*]	

Configuring the FTP Client

After a device serving as the FTP client has established a connection with the FTP server (For how to establish an FTP connection, refer to Establishing an FTP Connection.), you can perform the following operations in the authorized directories of the FTP server:

To do…	Use the command…	Remarks
Display help information of FTP-related commands supported by the remote FTP server	**remotehelp** [*protocol-command*]	Optional
Enable information display in a detailed manner	**verbose**	Optional Enabled by default
Enable FTP related debugging when the device acts as the FTP client	**debugging**	Optional Disabled by default
Use another username to relog after logging in to the FTP server successfully	**user** *username* [*password*]	Optional
Set the file transfer mode to ASCII	**ascii**	Optional ASCII by default
Set the file transfer mode to binary	**binary**	Optional ASCII by default
Change the working path on the remote FTP server	**cd** *directory*	Optional
Exit the current directory and enter the upper level directory	**cdup**	Optional

To do…	Use the command…	Remarks
View the detailed information of the files/directories on the FTP server	**dir** [*remotefile* [*localfile*]]	Optional
View the names of the files/directories on the FTP server	**ls** [*remotefile* [*localfile*]]	Optional
Download a file from the FTP server	**get** *remotefile* [*localfile*]	Optional
Upload a file to the FTP server	**put** *localfile* [*remotefile*]	Optional
View the currently accessed directory on the remote FTP server	**pwd**	Optional
View the working directory of the FTP client	**lcd**	Optional
Create a directory on the FTP server	**mkdir** *directory*	Optional
Set the data transfer mode to passive	**passive**	Optional Passive by default
Permanently delete the specified file on the FTP server	**delete** *remotefile*	Optional
Delete specified directory on the FTP server	**rmdir** *directory*	Optional
Disconnect from the FTP server without exiting the FTP client view	**disconnect**	Optional Equal to the **close** command
Disconnect from the FTP server without exiting the FTP client view	**close**	Optional Equal to the **disconnect** command
Disconnect from the FTP server and exit to user view	**bye**	Optional
Terminate the connection with the remote FTP server, and exit to user view	**quit**	Optional Available in FTP client view, equal to the **bye** command

Note

- FTP uses two modes for file transfer: ASCII mode and binary mode.
- The **ls** command can only display the file/directory name, while the **dir** command can display more information, such as the sizes of and date of creation of files or directories.
- The commands listed in the above table are only available for level 3 (manage level) users logging in to the device which serves as the FTP client. However, whether the users can successfully execute the commands depends on the FTP server's authorization.

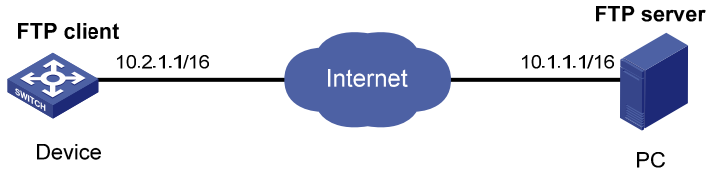
FTP Client Configuration Example

Single Device Upgrade

Network requirements

- As shown in Figure 2-2, use Device as an FTP client and PC as the FTP server. Their IP addresses are 10.2.1.1/16 and 10.1.1.1/16 respectively. An available route exists between Device and PC.
- Device downloads a startup file from PC for device upgrade, and uploads the configuration file to PC for backup.
- On PC, an FTP user account has been created for the FTP client, with the username being **abc** and the password being **pwd**.

Figure 2-2 Network diagram for FTPing a startup file from an FTP server



Configuration procedure

Caution

If the available memory space of the device is not enough, use the **fixdisk** command to clear the memory or use the **delete /unreserved** *file-url* command to delete the files not in use and then perform the following operations.

Log in to the server through FTP.

```
<Sysname> ftp 10.1.1.1
Trying 10.1.1.1
Connected to 10.1.1.1
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(10.1.1.1:(none)):abc
331 Give me your password, please
Password:
230 Logged in successfully
```

Set the file transmission mode to binary to transmit startup file.

```
[ftp] binary
200 Type set to I.
```

Download the startup file **newest.bin** from PC to Device.

```
[ftp] get newest.bin
```

Upload the configuration file **config.cfg** of Device to the server for backup.

```
[ftp] ascii
```

```
[ftp] put config.cfg back-config.cfg
227 Entering Passive Mode (10,1,1,1,4,2).
125 ASCII mode data connection already open, transfer starting for /config.cfg.
226 Transfer complete.
FTP: 3494 byte(s) sent in 5.646 second(s), 618.00 byte(s)/sec.
[ftp] bye
```

Specify **newest.bin** as the main startup file to be used at the next startup.

```
<Sysname> boot-loader file newest.bin main
```

Reboot the device, and the startup file is updated at the system reboot.

```
<Sysname> reboot
```

Caution

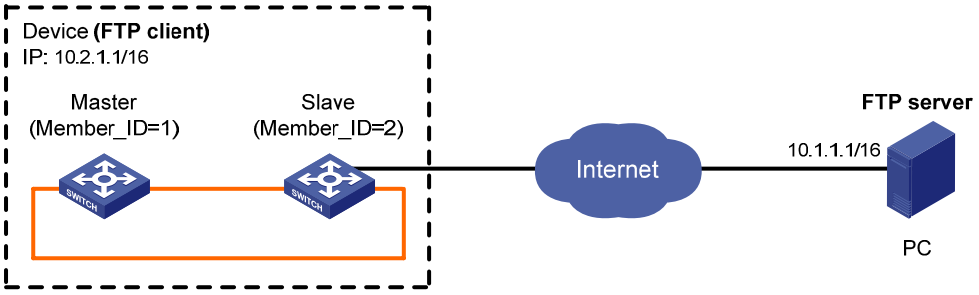
The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Stacking System Upgrade

Network requirements

- As shown in Figure 2-3, use Device as an FTP client and PC as the FTP server. Their IP addresses are 10.2.1.1/16 and 10.1.1.1/16 respectively. An available route exists between Device and PC.
- Device downloads a startup file from PC for device upgrade, and uploads the configuration file to PC for backup.
- On PC, an FTP user account has been created for the FTP client, with the username being **abc** and the password being **pwd**.

Figure 2-3 Network diagram for FTPing a startup file from an FTP server



Note: The orange lines represent stack cables.

Configuration procedure

Caution

If the available memory space of the device is not enough, use the **fixdisk** command to clear the memory or use the **delete /unreserved** *file-url* command to delete the files not in use and then perform the following operations.

Log in to the server through FTP.

```
<Sysname> ftp 10.1.1.1
Trying 10.1.1.1 ...
Connected to 10.1.1.1.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(10.1.1.1:(none)):abc
331 Give me your password, please
Password:
230 Logged in successfully
```

Set the file transmission mode to binary to transmit startup file.

```
[ftp] binary
200 Type set to I.
```

Download the startup file **newest.bin** from PC to the device.

- Download the startup file **newest.bin** from PC to the root directory of the storage medium on the master.

```
[ftp] get newest.bin
```

- Download the startup file **newest.bin** from PC to the root directory of the storage medium of a slave (with member ID of 2).

```
[ftp] get newest.bin slot2#flash:/newest.bin
```

Upload the configuration file **config.cfg** of the device to the server for backup.

```
[ftp] ascii
[ftp] put config.cfg back-config.cfg
227 Entering Passive Mode (10,1,1,1,4,2).
125 ASCII mode data connection already open, transfer starting for /config.cfg.
226 Transfer complete.
FTP: 3494 byte(s) sent in 5.646 second(s), 618.00 byte(s)/sec.
[ftp] bye
```

Specify **newest.bin** as the main startup file to be used at the next startup for all the member devices.

```
<Sysname> boot-loader file newest.bin slot all main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 1!
  The specified file will be used as the main boot file at the next reboot on slot 2!
```

Reboot the device, and the startup file is updated at the system reboot.

```
<Sysname> reboot
```

Caution

The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Configuring the FTP Server

Configuring FTP Server Operating Parameters

The FTP server uses one of the two modes to update a file when you upload the file (use the **put** command) to the FTP server:

- In fast mode, the FTP server starts writing data to the storage medium after a file is transferred to the memory. This prevents the existing file on the FTP server from being corrupted in the event that anomaly, power failure for example, occurs during a file transfer.
- In normal mode, the FTP server writes data to the storage medium while receiving data. This means that any anomaly, power failure for example, during file transfer might result in file corruption on the FTP server. This mode, however, consumes less memory space than the fast mode.

Follow these steps to configure the FTP server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the FTP server	**ftp server enable**	Required Disabled by default.
Control the access to the device from FTP clients through ACL	**ftp server acl** *acl-number*	Optional By default, the access to the device from FTP clients is not controlled.
Configure the idle-timeout timer	**ftp timeout** *minutes*	Optional 30 minutes by default. Within the idle-timeout time, if there is no information interaction between the FTP server and client, the connection between them is terminated.
Set the file update mode for the FTP server	**ftp update** { **fast** | **normal** }	Optional Normal update is used by default.
Quit to user view	**quit**	—

To do…	Use the command…	Remarks
Manually release the FTP connection established with the specified username	**free ftp user** *username*	Optional Available in user view

Configuring Authentication and Authorization on the FTP Server

To allow an FTP user to access certain directories on the FTP server, you need to create an account for the user, authorizing access to the directories and associating the username and password with the account.

The following configuration is used when the FTP server authenticates and authorizes a local FTP user. If the FTP server needs to authenticate a remote FTP user, you need to configure authentication, authorization and accounting (AAA) policy instead of the local user. For detailed configuration, refer to *AAA Configuration* in the *Security Volume*.

Follow these steps to configure authentication and authorization for FTP server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a local user and enter its view	**local-user** *user-name*	Required No local user exists by default, and the system does not support FTP anonymous user access.
Assign a password to the user	**password** { **simple** | **cipher** } *password*	Required
Assign the FTP service to the user	**service-type ftp**	Required By default, the system does not support anonymous FTP access, and does not assign any service. If the FTP service is assigned, the root directory of the device is used by default.
Configure user properties	**authorization-attribute** { **acl** *acl-number* | **callback-number** *callback-number* | **idle-cut** *minute* | **level** *level* | **user-profile** *profile-name* | **vlan** *vlan-id* | **work-directory** *directory-name* } *	Optional By default, the FTP/SFTP users can access the root directory of the device, and the user level is 0. You can change the default configuration by using this command.

Note

- For more information about the **local-user**, **password**, **service-type ftp**, and **authorization-attribute** commands, refer to *AAA Command* in the *Security Volume*.
- When the device serves as the FTP server, if the client is to perform the write operations (upload, delete, create, and delete for example) on the device's file system, the FTP login users must be level 3 users; if the client is to perform other operations, for example, read operation, the device has no restriction on the user level of the FTP login users, that is, any level from 0 to 3 is allowed.

FTP Server Configuration Example

Single Device Upgrade

Network requirements

- As shown in Figure 2-4, use Device as an FTP server, and the PC as the FTP client. Their IP addresses are 1.2.1.1/16 and 1.1.1.1/16 respectively. An available route exists between Device and PC.
- PC keeps the updated startup file of the device. Use FTP to upgrade the device and back up the configuration file.
- Set the username to **ftp** and the password to **pwd** for the FTP client to log in to the FTP server.

Figure 2-4 Smooth upgrading using the FTP server

Configuration procedure

1) Configure Device (FTP Server)

Create an FTP user account **ftp**, set its password to **pwd** and the user privilege level to level 3 (the manage level).

```
<Sysname> system-view
[Sysname] local-user ftp
[Sysname-luser-ftp] password simple pwd
[Sysname-luser-ftp] authorization-attribute work-directory level 3
```

Authorize ftp's access to the root directory of the flash.

```
[Sysname-luser-ftp] authorization-attribute work-directory flash:/
```

Specify **ftp** to use FTP.

```
[Sysname-luser-ftp] service-type ftp
[Sysname-luser-ftp] quit
```

Enable FTP server.

```
[Sysname] ftp server enable
[Sysname] quit
```

Check files on your device. Remove those redundant to ensure adequate space for the startup file to be uploaded.

```
<Sysname> dir
Directory of flash:/

   0     -rw-  10471471  Sep 18 2008 02:45:15   s4800g -d501.bin
   1     -rw-   9989823  Jul 14 2008 19:30:46   s4800g _b57.bin
   2     -rw-         6  Apr 26 2000 12:04:33   patchstate
   3     -rw-      2337  Apr 26 2000 14:18:45   config.cfg
   4     drw-         -  Apr 26 2000 13:10:56   test
   5     -rw-      2337  Apr 26 2000 13:47:32   archive_1.cfg
```

```
6     -rw-    478164  Apr 26 2000 14:52:35   s4800g1_505.btm
7     -rw-       368  Apr 26 2000 12:04:04   patch_xxx.bin
8     -rw-      2337  Apr 26 2000 14:16:48   sfp.cfg
9     -rw-      2195  Apr 26 2000 14:10:41   s4800g2.cfg

31496 KB total (11004 KB free)
<Sysname> delete /unreserved flash:/sfp.cfg
```

2) Configure the PC (FTP Client)

Log in to the FTP server through FTP.

```
c:\> ftp 1.1.1.1
Connected to 1.1.1.1.
220 FTP service ready.
User(1.1.1.1:(none)):abc
331 Password required for abc.
Password:
230 User logged in.
```

Download the configuration file **config.cfg** of the device to the PC for backup.

```
ftp> get config.cfg back-config.cfg
```

Upload the configuration file **newest.bin** to Device.

```
ftp> put newest.bin
ftp> bye
```

Note

- You can take the same steps to upgrade configuration file with FTP. When upgrading the configuration file with FTP, put the new file under the root directory of the storage medium (For a device that has been partitioned, the configuration file must be saved on the first partition.).
- After you finish upgrading the Boot ROM program through FTP, you must execute the **bootrom update** command to upgrade the Boot ROM.

---

3) Upgrade Device

Specify **newest.bin** as the main startup file to be used at the next startup.

```
<Sysname> boot-loader file newest.bin main
```

Reboot the device and the startup file is updated at the system reboot.

```
<Sysname> reboot
```

Caution

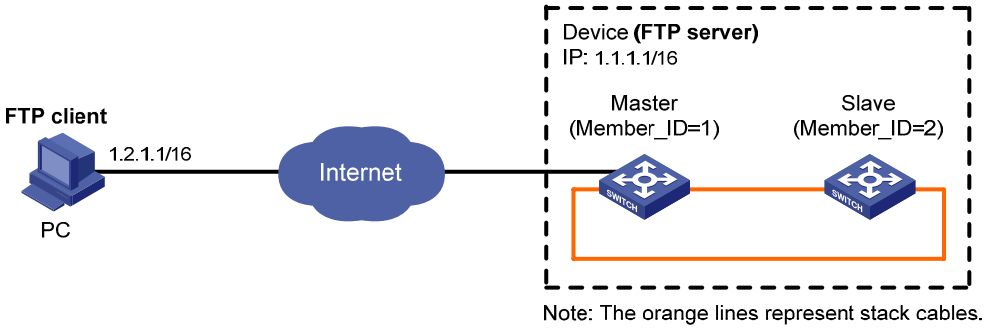
The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Stacking System Upgrade

Network requirements

- As shown in Figure 2-5, use Device as an FTP server, and the PC as the FTP client. An available route exists between Device and PC.
- PC keeps the updated startup file of the device. Use FTP to upgrade the device and back up the configuration file.
- Set the username to **ftp** and the password to **pwd** for the FTP client to log in to the FTP server.

Figure 2-5 Smooth upgrading using the FTP server



Configuration procedure

1) Configure Device (FTP Server)

Create an FTP user account **ftp**, set its password to **pwd** and the user privilege level to level 3 (the manage level).

```
<Sysname> system-view
[Sysname] local-user ftp
[Sysname-luser-ftp] password simple pwd
[Sysname-luser-ftp] authorization-attribute work-directory level 3
```

Authorize ftp's access to the root directory of the flash on the master.

```
[Sysname-luser-ftp] authorization-attribute work-directory flash:/
```

To access the root directory of storage medium of a slave (with the member ID 2):

```
[Sysname-luser-ftp] authorization-attribute work-directory slot2#flash:/
```

Specify **ftp** to use FTP.

```
[Sysname-luser-ftp] service-type ftp
[Sysname-luser-ftp] quit
```

Enable FTP server.

```
[Sysname] ftp server enable
```

```
[Sysname] quit
```

Check files on your device. Remove those redundant to ensure adequate space for the startup file to be uploaded.

```
<Sysname> dir
Directory of flash:/

   0     -rw-  10471471  Sep 18 2008 02:45:15   s4800g-d501.bin
   1     -rw-   9989823  Jul 14 2008 19:30:46   s4800g_b57.bin
   2     -rw-         6  Apr 26 2000 12:04:33   patchstate
   3     -rw-      2337  Apr 26 2000 14:18:45   startup.cfg
   4     drw-         -  Apr 26 2000 13:10:56   test
   5     -rw-      2337  Apr 26 2000 13:47:32   archive_1.cfg
   6     -rw-    478164  Apr 26 2000 14:52:35   s4800g_505.btm
   7     -rw-       368  Apr 26 2000 12:04:04   patch_xxx.bin
   8     -rw-      2337  Apr 26 2000 14:16:48   sfp.cfg
   9     -rw-      2195  Apr 26 2000 14:10:41   s4800g.cfg

31496 KB total (11004 KB free)
```

2) Configure the PC (FTP Client)

Log in to the FTP server through FTP.

```
c:\> ftp 1.1.1.1
Connected to 1.1.1.1.
220 FTP service ready.
User(1.1.1.1:(none)):abc
331 Password required for abc.
Password:
230 User logged in.
```

Download the configuration file **config.cfg** of the device to the PC for backup.

```
ftp> get config.cfg back-config.cfg
```

Upload the configuration file **newest.bin** to the root directory of the storage medium on the master.

```
ftp> put newest.bin
ftp> bye
```

Note

- You can take the same steps to upgrade configuration file with FTP. When upgrading the configuration file with FTP, put the new file under the root directory of the storage medium.
- After you finish upgrading the Boot ROM program through FTP, you must execute the **bootrom update** command to upgrade the Boot ROM.

---

3) Upgrade Device

Copy the startup file **newest.bin** to the root directory of the storage medium on a slave (with the member ID 2).

```
<Sysname> copy newest.bin slot2#flash:/
```

Specify **newest.bin** as the main startup file to be used at the next startup for all the member devices.

```
<Sysname> boot-loader file newest.bin slot all main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 1!
  The specified file will be used as the main boot file at the next reboot on slot 2!
```

Reboot the device and the startup file is updated at the system reboot.

```
<Sysname> reboot
```

Caution

The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Displaying and Maintaining FTP

To do…	Use the command…	Remarks
Display the configuration of the FTP client	**display ftp client configuration**	Available in any view
Display the configuration of the FTP server	**display ftp-server**	Available in any view
Display detailed information about logged-in FTP users	**display ftp-user**	Available in any view

3 TFTP Configuration

When configuring TFTP, go to these sections for information you are interested in:

TFTP Overview

Introduction to TFTP

The Trivial File Transfer Protocol (TFTP) provides functions similar to those provided by FTP, but it is less complex than FTP in interactive access interface and authentication. Therefore, it is more suitable in environments where complex interaction is not needed between client and server.

TFTP uses the UDP port 69 for data transmission. For TFTP basic operation, refer to RFC 1986.

In TFTP, file transfer is initiated by the client.

- In a normal file downloading process, the client sends a read request to the TFTP server, receives data from the server, and then sends the acknowledgement to the server.
- In a normal file uploading process, the client sends a write request to the TFTP server, sends data to the server, and receives the acknowledgement from the server.

TFTP transfers files in two modes:

- Binary mode for program file transmission, like files with the suffixes **.app**, **.bin**, or **.btm**.
- ASCII mode for text file transmission, like files with the suffixes **.txt**, **.bat**, or **.cfg**.

Operation of TFTP

Note

Only the TFTP client service is available with your device at present.

Figure 3-1 TFTP configuration diagram

Before using TFTP, the administrator needs to configure IP addresses for the TFTP client and server, and make sure that there is a reachable route between the TFTP client and server.

When the device serves as the TFTP client, you need to perform the following configuration:

Table 3-1 Configuration when the device serves as the TFTP client

Device	Configuration	Remarks
Device (TFTP client)	• Configure the IP address and routing function, and ensure that the route between the device and the TFTP server is available. • Use the **tftp** command to establish a connection to the remote TFTP server to upload/download files to/from the TFTP server	—
PC (TFTP server)	Enable TFTP server on the PC, and configure the TFTP working directory.	—

Configuring the TFTP Client

When a device acts as a TFTP client, you can upload a file on the device to a TFTP server and download a file from the TFTP server to the local device. You can use either of the following ways to download a file:

- Normal download: The device writes the obtained file to the storage medium directly. In this way, if you use a filename that exists in the directory, the original system file will be overwritten and if file download fails (for example, due to network disconnection), the device cannot start up normally because the original system file has been deleted.
- Secure download: The device saves the obtained file to its memory and does not write it to the storage medium until the whole file is obtained. In this way, if file download fails (for example, due to network disconnection), the device can still start up because the original system file is not overwritten. This mode is more secure but consumes more memory.

You are recommended to use the secure mode or, if you use the normal mode, specify a filename not existing in the current directory as the target filename when downloading the startup file or the startup configuration file.

Source address binding means to configure an IP address on a stable interface such as a loopback interface, and then use this IP address as the source IP address of a TFTP connection. The source address binding function simplifies the configuration of ACL rules and security policies. You just need to specify the source or destination address argument in an ACL rule as this address to filter inbound and outbound packets on the device, ignoring the difference between interface IP addresses as well as the affect of interface statuses. You can configure the source address by configuring the source interface or source IP address. The primary IP address configured on the source interface is the source address of the transmitted packets. The source address of the transmitted packets is selected following these rules:

- If no source address of the TFTP client is specified, a device uses the IP address of the interface determined by the matched route as the source IP address to communicate with a TFTP server.
- If the source address is specified with the **tftp client source** or **tftp** command, this source address is adopted.
- If you use the **tftp client source** command and the **tftp** command to specify a source address respectively, the source address configured with the **tftp** command is used to communicate with a TFTP server.

The source address specified with the **tftp client source** command is valid for all TFTP connections and the source address specified with the **tftp** command is valid only for the current **tftp** connection.

Follow these steps to configure the TFTP client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Control the access to the TFTP servers from the device through ACL	**tftp-server** [**ipv6**] **acl** *acl-number*	Optional By default, the access to the TFTP servers from the device is not controlled.
Configure the source address of the TFTP client	**tftp client source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }	Optional A device uses the source address determined by the matched route to communicate with the TFTP server by default.
Return to user view	**quit**	—
Download or upload a file in an IPv4 network	**tftp** *server-address* { **get** | **put** | **sget** } *source-filename* [*destination-filename*] [**source** { **interface** *interface-type interface-number* | **ip** *source-ip-address* }]	Optional Available in user view
Download or upload a file in an IPv6 network	**tftp ipv6** *tftp-ipv6-server* [**-i** *interface-type interface-number*] { **get** | **put** } *source-file* [*destination-file*]	Optional Available in user view

Note

- If no primary IP address is configured on the source interface, no TFTP connection can be established.
- If you use the **ftp client source** command to first configure the source interface and then the source IP address of the packets of the TFTP client, the new source IP address will overwrite the current one, and vice versa.

Displaying and Maintaining the TFTP Client

To do…	Use the command…	Remarks
Display the configuration of the TFTP client	**display tftp client configuration**	Available in any view

TFTP Client Configuration Example

Single Device Upgrade

Network requirements

- As shown in Figure 3-2, use a PC as the TFTP server and Device as the TFTP client. Their IP addresses are 1.2.1.1/16 and 1.1.1.1/16 respectively. An available route exists between Device and PC.
- Device downloads a startup file from PC for upgrading and uploads a configuration file named **config.cfg** to PC for backup.

Figure 3-2 Smooth upgrading using the TFTP client function

Configuration procedure

1) Configure PC (TFTP Server), the configuration procedure is omitted.
- On the PC, enable the TFTP server
- Configure a TFTP working directory

2) Configure Device (TFTP Client)

Caution

If the available memory space of the device is not enough, use the **fixdisk** command to clear the memory or use the **delete /unreserved** *file-url* command to delete the files not in use and then perform the following operations.

Enter system view.

```
<Sysname> system-view
```

Download application file **newest.bin** from PC.

```
<Sysname> tftp 1.2.1.1 get newest.bin
```

Upload a configuration file **config.cfg** to the TFTP server.

```
<Sysname> tftp 1.2.1.1 put config.cfg configback.cfg
```

Specify **newest.bin** as the main startup file to be used at the next startup.

```
<Sysname> boot-loader file newest.bin main
```

Reboot the device and the software is upgraded.

```
<Sysname> reboot
```

Caution

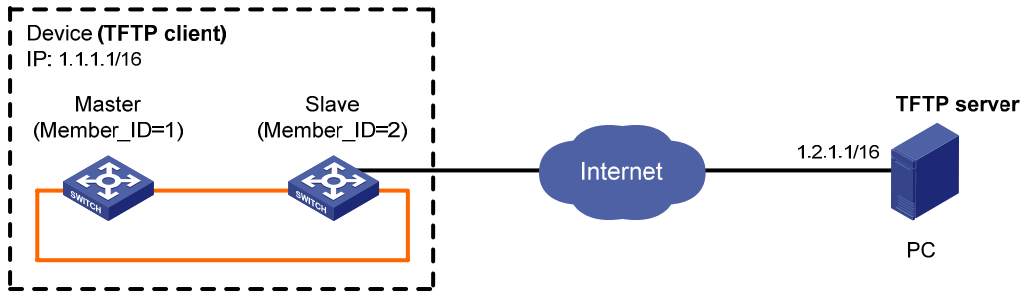
The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Stacking System Upgrade

Network requirements

- As shown in Figure 3-3, use a PC as the TFTP server and Device as the TFTP client. Their IP addresses are 1.2.1.1/16 and 1.1.1.1/16 respectively. An available route exists between Device and PC.
- Device downloads a startup file from PC for upgrading and uploads a configuration file named **config.cfg** to PC for backup.

Figure 3-3 Smooth upgrading using the TFTP client function



Note: The orange lines represent stack cables.

Configuration procedure

1) Configure PC (TFTP Server), the configuration procedure is omitted.
- On the PC, enable the TFTP server
- Configure a TFTP working directory

2) Configure Device (TFTP Client)

Caution

If the available memory space of the device is not enough, use the **fixdisk** command to clear the memory or use the **delete /unreserved** *file-url* command to delete the files not in use and then perform the following operations.

Enter system view.

```
<Sysname> system-view
```

Download application file **newest.bin** from PC to Device.

- Download application file **newest.bin** from PC to the root directory of the storage medium on the master.

```
<Sysname> tftp 1.2.1.1 get newest.bin
```

- Download application file **newest.bin** from PC to the root directory of the storage medium on a slave (with the member ID 2).

```
<Sysname> tftp 1.2.1.1 get newest.bin slot2#flash:/newest.bin
```

Upload a configuration file **config.cfg** to the TFTP server.

```
<Sysname> tftp 1.2.1.1 put config.cfg configback.cfg
```

Specify **newest.bin** as the main startup file to be used at the next startup for all the member devices.

```
<Sysname> boot-loader file newest.bin slot all main
  This command will set the boot file of the specified board. Continue? [Y/N]:y
  The specified file will be used as the main boot file at the next reboot on slot 1!
  The specified file will be used as the main boot file at the next reboot on slot 2!
```

Reboot the device and the software is upgraded.

```
<Sysname> reboot
```

Caution

The startup file used for the next startup must be saved under the root directory of the storage medium. You can copy or move a file to the root directory of the storage medium. For the details of the **boot-loader** command, refer to *Device Management Commands* in the *System Volume*.

Table of Contents

1 HTTP Configuration

When configuring HTTP, go to these sections for information you are interested in:

HTTP Overview

The Hypertext Transfer Protocol (HTTP) is used for transferring web page information across the Internet. It is an application-level protocol in the TCP/IP protocol suite. The connection-oriented Transport Control Protocol (TCP) is adopted on the transport layer.

Currently, HTTP/1.0 is supported on the device.

How HTTP Works

In the HTTP, the client/server mode is used for communication. The client and the server exchange messages following these procedures:

1) A TCP connection is created between the client and the server. Typically, the port number is 80.
2) The client sends a request to the server.
3) The server processes the request and sends back a response.
4) The TCP connection is closed.

Logging In to the Device Through HTTP

You can log onto the device using the HTTP protocol with HTTP service enabled, accessing and controlling the device with Web-based network management.

To implement security management on the device, you can use the following methods to enhance the security of the device.

- Enable HTTP service only when necessary.
- Change the port number of the HTTP service as a port number not commonly used (80 or 8080), thus reducing attacks from illegal users on the HTTP service.
- Associate the HTTP service with an ACL to let pass only the filtered clients.

Protocols and Standards

RFC 1945: Hypertext Transfer Protocol – HTTP/1.0

Enabling the HTTP Service

The device can act as the HTTP server and the users can access and control the device through the Web function only after the HTTP service is enabled.

Follow these steps to enable the HTTP service:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the HTTP service	**ip http enable**	Required

Configuring the Port Number of the HTTP Service

Configuration of the port number of the HTTP service can reduce the attacks from illegal users on the HTTP service.

Follow these steps to configure the port number of the HTTP service:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the port number of the HTTP service	**ip http port** *port-number*	Required By default, the port number of the HTTP service is 80.

Note

If you execute the **ip http port** command for multiple times, the last configured port number is used.

Associating the HTTP Service with an ACL

By associating the HTTP service with an ACL, only the clients that pass ACL filtering are allowed to access the device.

Follow these steps to associate the HTTP service with an ACL:

To do…	Use the command…	Remarks
Enters system view	**system-view**	—
Associate the HTTP service with an ACL	**ip http acl** *acl-number*	Required The HTTP service is not associated with an ACL by default.

Displaying and Maintaining HTTP

To do…	Use the command…	Remarks
Display information about HTTP	**display ip http**	Available in any view

2 HTTPS Configuration

When configuring HTTPS, go to these sections for information you are interested in:

HTTPS Overview

The Secure HTTP (HTTPS) refers to the HTTP protocol that supports the Security Socket Layer (SSL) protocol.

The SSL protocol of HTTPS enhances the security of the device in the following ways:

- Uses the SSL protocol to ensure the legal clients to access the device securely and prohibit the illegal clients;
- Encrypts the data exchanged between the HTTPS client and the device to ensure the data security and integrity, thus realizing the security management of the device;
- Defines certificate attribute-based access control policy for the device to control the access right of the client, in order to further avoid attacks from illegal clients.

Note

- The total number of HTTP connections and HTTPS connections on a device cannot exceed ten.
- For SSL details, refer to *SSL Configuration* in the *Security Volume*.

HTTPS Configuration Task List

Complete these tasks to configure HTTPS:

Configuration task	Remarks
Associating the HTTPS Service with an SSL Server Policy	Required
Enabling the HTTPS Service	Required
Associating the HTTPS Service with a Certificate Attribute Access Control Policy	Optional

Configuration task	Remarks
Configuring the Port Number of the HTTPS Service	Optional
Associating the HTTPS Service with an ACL	Optional

Associating the HTTPS Service with an SSL Server Policy

You need to associate the HTTPS service with a created SSL server policy before enabling the HTTPS service.

Follow these steps to associate the HTTPS service with an SSL server policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Associate the HTTPS service with an SSL server policy	**ip https ssl-server-policy** *policy-name*	Required Not associated by default

Note

- If the **ip https ssl-server-policy** command is executed repeatedly, the HTTPS service is only associated with the last specified SSL server policy.
- When the HTTPS service is disabled, the association between the HTTPS service and the SSL server is automatically removed. To enable it again, you need to re-associate the HTTPS service with an SSL server policy.
- When the HTTPS service is enabled, no modification of its associated SSL server policy takes effect.

Enabling the HTTPS Service

The device can act as the HTTPS server and users can access and control the device through the Web function only when the HTTPS service is enabled.

Follow these steps to enable the HTTPS service:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the HTTPS service	**ip https enable**	Required Disabled by default.

Note

- After the HTTPS service is enabled, you can use the **display ip https** command to view the state of the HTTPS service and verify the configuration.
- Enabling of the HTTPS service will trigger an SSL handshake negotiation process. During the process, if the local certificate of the device already exists, the SSL negotiation is successfully performed, and the HTTPS service can be started normally. If no local certificate exists, a certificate application process will be triggered by the SSL negotiation. Since the application process takes much time, the SSL negotiation may fail and the HTTPS service cannot be started normally. Therefore, the **ip https enable** command must be executed for multiple times to ensure normal startup of the HTTPS service.

Associating the HTTPS Service with a Certificate Attribute Access Control Policy

Associating the HTTPS service with a configured certificate access control policy helps control the access right of the client, thus providing the device with enhanced security.

Follow these steps to associate the HTTPS service with a certificate attribute access control policy:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Associate the HTTPS service with a certificate attribute access control policy	**ip https certificate access-control-policy** *policy-name*	Required Not associated by default.

Note

- If the **ip https certificate access-control-policy** command is executed repeatedly, the HTTPS server is only associated with the last specified certificate attribute access control policy.
- If the HTTPS service is associated with a certificate attribute access control policy, the **client-verify enable** command must be configured in the SSL server policy. Otherwise, the client cannot log onto the device.
- If the HTTPS service is associated with a certificate attribute access control policy, the latter must contain at least one **permit** rule. Otherwise, no HTTPS client can log onto the device.
- For the configuration of an SSL server policy, refer to *PKI Configuration* in the *Security Volume*.

Configuring the Port Number of the HTTPS Service

Configuration of the port number of the HTTPS service can reduce the attacks from illegal users on the HTTPS service.

Follow these steps to configure the port number of the HTTPS service:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the port number of the HTTPS service	**ip https port** *port-number*	Optional By default, the port number of the HTTPS service is 443.

Note

If you execute the **ip https port** command for multiple times, the last configured port number is used.

Associating the HTTPS Service with an ACL

Associating the HTTPS service with an ACL can filter out requests from some clients to let pass only clients that pass the ACL filtering.

Follow these steps to associate the HTTPS service with an ACL:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Associate the HTTPS service with an ACL	**ip https acl** *acl-number*	Required Not associated by default.

Displaying and Maintaining HTTPS

To do…	Use the command…	Remarks
Display information about HTTPS	**display ip https**	Available in any view

HTTPS Configuration Example

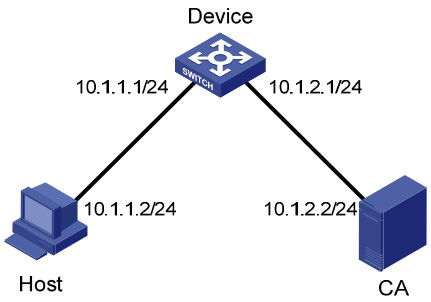
Network requirements

- Host acts as the HTTPS client and Device acts as the HTTPS server.
- Host accesses Device through Web to control Device.
- CA (Certificate Authority) issues certificate to Device. The common name of CA is **new-ca**.

Caution

In this configuration example, Windows Server serves as CA and you need to install Simple Certificate Enrollment Protocol (SCEP) component.

Figure 2-1 Network diagram for HTTPS configuration



Configuration procedure

Perform the following configurations on Device:

1) Apply for a certificate for Device

Configure a PKI entity.

```
<Device> system-view
[Device] pki entity en
[Device-pki-entity-en] common-name http-server1
[Device-pki-entity-en] fqdn ssl.security.com
[Device-pki-entity-en] quit
```

Configure a PKI domain.

```
[Device] pki domain 1
[Device-pki-domain-1] ca identifier new-ca
[Device-pki-domain-1] certificate request url http://10.1.2.2:8080/certsrv/mscep/mscep.dll
[Device-pki-domain-1] certificate request from ra
[Device-pki-domain-1] certificate request entity en
[Device-pki-domain-1] quit
```

Generate a local RSA key pair.

```
[Device] public-key local create rsa
```

Obtain a server certificate from CA.

```
[Device] pki retrieval-certificate ca domain 1
```

Apply for a local certificate.

```
[Device] pki request-certificate domain 1
```

2) Configure an SSL server policy associated with the HTTPS service

Configure an SSL server policy.

```
[Device] ssl server-policy myssl
[Device-ssl-server-policy-myssl] pki-domain 1
[Device-ssl-server-policy-myssl] client-verify enable
[Device-ssl-server-policy-myssl] quit
```

3) Configure a certificate access control policy

Configure a certificate attribute group.

```
[Device] pki certificate attribute-group mygroup1
[Device-pki-cert-attribute-group-mygroup1] attribute 1 issuer-name dn ctn new-ca
[Device-pki-cert-attribute-group-mygroup1] quit
```

Configure certificate access control policy **myacp** and create a control rule.

```
[Device] pki certificate access-control-policy myacp
[Device-pki-cert-acp-myacp] rule 1 permit mygroup1
[Device-pki-cert-acp-myacp] quit
```

4) Reference an SSL server policy

Associate the HTTPS service with the SSL server policy **myssl**.

```
[Device] ip https ssl-server-policy myssl
```

5) Associate the HTTPS service with a certificate attribute access control policy

Associate the HTTPS service with certificate attribute access control policy **myacp**.

```
[Device] ip https certificate access-control-policy myacp
```

6) Enable the HTTPS service

Enable the HTTPS service.

```
[Device] ip https enable
```

7) Verify the configuration

Launch the IE explorer on Host, and enter https://10.1.1.1. You can log in to Device and control it.

Note

- The URL of the HTTPS server starts with https://, and that of the HTTP server starts with http://.
- For details of PKI commands, refer to *PKI Commands* in the *Security Volume*.
- For details of the **public-key local create rsa** command, refer to *Public Key Commands* in the *Security Volume*.
- For details of SSL commands, refer to *SSL Commands* in the *Security Volume*.

Table of Contents

1 SNMP Configuration

When configuring SNMP, go to these sections for information you are interested in:

SNMP Overview

Simple Network Management Protocol (SNMP) offers a framework to monitor network devices through TCP/IP protocol suite. It provides a set of basic operations in monitoring and maintaining the Internet and has the following characteristics:

- Automatic network management: SNMP enables network administrators to search and modify information, find and diagnose network problems, plan for network growth, and generate reports on network nodes.
- SNMP shields the physical differences between various devices and thus realizes automatic management of products from different manufacturers. Offering only the basic set of functions, SNMP makes the management tasks independent of both the physical features of the managed devices and the underlying networking technology. Thus, SNMP achieves effective management of devices from different manufacturers, especially in small, high-speed and low cost network environments.

SNMP Mechanism

An SNMP enabled network comprises a Network Management Station (NMS) and an agent.

- An NMS is a station that runs the SNMP client software. It offers a user friendly interface, making it easier for network administrators to perform most network management tasks.
- An agent is a program on the device. It receives and handles requests sent from the NMS. Only under certain circumstances, such as interface state change, will the agent inform the NMS.

An NMS manages an SNMP enabled network, whereas agents are the managed network device. They exchange management information through the SNMP protocol.

SNMP provides the following four basic operations:

- Get operation: The NMS gets the value of one or more objects of the agent through this operation.
- Set operation: The NMS can reconfigure the value of one or more objects in the agent MIB (Management Information Base) by means of this operation.
- Trap operation: The agent sends traps to the NMS through this operation.
- Inform operation: The NMS sends traps to other NMSs through this operation.

SNMP Protocol Version

Currently, SNMP agents support SNMPv3 and are compatible with SNMPv1 and SNMPv2c.

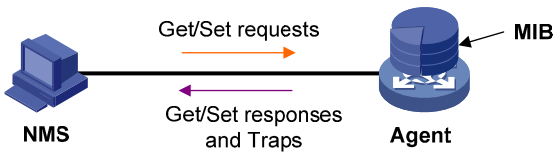
- SNMPv1 uses community name for authentication, which defines the relationship between an SNMP NMS and an SNMP agent. SNMP packets with community names that did not pass the authentication on the device will simply be discarded. A community name performs a similar role as a key word and can be used to regulate access from NMS to agent.
- SNMPv2c uses community name for authentication. Compatible with SNMPv1, it extends the functions of SNMPv1. SNMPv2c provides more operation modes such as GetBulk and InformRequest; it supports more data types such as Counter64; and it provides various error codes, thus being able to distinguish errors in more detail.
- SNMPv3 offers an authentication that is implemented with a User-Based Security Model (USM). You can set the authentication and privacy functions. The former is used to authenticate the validity of the sending end of the authentication packets, preventing access of illegal users; the latter is used to encrypt packets between the NMS and agent, preventing the packets from being intercepted. USM ensures a more secure communication between SNMP NMS and SNMP agent by authentication with privacy, authentication without privacy, or no authentication no privacy.

Successful interaction between NMS and agent requires consistency of SNMP versions configured on them. You can configure multiple SNMP versions for an agent to interact with different NMSs.

MIB Overview

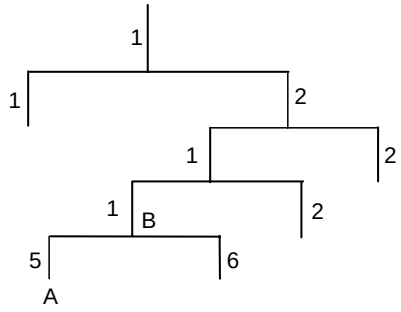
Any managed resource can be identified as an object, which is known as the managed object. Management Information Base (MIB) is a collection of all the managed objects. It defines a set of characteristics associated with the managed objects, such as the object identifier (OID), access right and data type of the objects. Each agent has its own MIB. NMS can read or write the managed objects in the MIB. The relationship between an NMS, agent and MIB is shown in Figure 1-1.

Figure 1-1 Relationship between NMS, agent and MIB



MIB stores data using a tree structure. The node of the tree is the managed object and can be uniquely identified by a path starting from the root node. As illustrated in the following figure, the managed object B can be uniquely identified by a string of numbers {1.2.1.1}. This string of numbers is the OID of the managed object B.

Figure 1-2 MIB tree



SNMP Configuration

As configurations for SNMPv3 differ substantially from those of SNMPv1 and SNMPv2c, their SNMP functionalities is introduced separately below.

Follow these steps to configure SNMPv3:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable SNMP agent	**snmp-agent**	Optional Disabled by default You can enable SNMP agent through this command or any commands that begin with **snmp-agent**.
Configure SNMP agent system information	**snmp-agent sys-info** { **contact** *sys-contact* | **location** *sys-location* | **version** { **all** | { **v1** | **v2c** | **v3** }***** } }	Optional The defaults are as follows: 3Com Corporation. for contact, Marlborough, MA 01752 USA for location, and SNMP v3 for the version.
Configure an SNMP agent group	**snmp-agent group v3** *group-name* [**authentication** | **privacy**] [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*]	Required
Convert the user-defined plain text password to a cipher text password	**snmp-agent calculate-password** *plain-password* **mode** { **md5** | **sha | 3desmd5 | 3dessha** } { **local-engineid** | **specified-engineid** *engineid* }	Optional
Add a new user to an SNMP agent group	**snmp-agent usm-user v3** *user-name group-name* [[**cipher**] **authentication-mode** { **md5** | **sha** } *auth-password* [**privacy-mode** { **3des** | **aes128** | **des56** } *priv-password*]] [**acl** *acl-number*]	Required If the **cipher** keyword is specified, the arguments *auth-password* and *priv-password* are considered as cipher text password.

To do…	Use the command…	Remarks
Configure the maximum size of an SNMP packet that can be received or sent by an SNMP agent	**snmp-agent packet max-size** *byte-count*	Optional 1,500 bytes by default
Configure the engine ID for a local SNMP agent	**snmp-agent local-engineid** *engineid*	Optional Company ID and device ID by default
Create or update the MIB view content for an SNMP agent	**snmp-agent mib-view** { **excluded** | **included** } *view-name oid-tree* [**mask** *mask-value*]	Optional MIB view name is ViewDefault and OID is 1 by default.

Follow these steps to configure SNMPv1 and SNMPv2c:

To do…			Use the command…	Remarks
Enter system view			**system-view**	—
Enable SNMP agent			**snmp-agent**	Optional Disabled by default You can enable SNMP agent through this command or any commands that begin with **snmp-agent**.
Configure SNMP agent system information			**snmp-agent sys-info** { **contact** *sys-contact* | **location** *sys-location* | **version** { { **v1** | **v2c** | **v3** }* | **all** } }	Required The defaults are as follows: 3Com Corporation. for contact, Marlborough, MA 01752 USA for location and SNMP v3 for the version.
Configure SNMP NMS access right	Configure directly	Create an SNMP community	**snmp-agent community** { **read** | **write** } *community-name* [**acl** *acl-number* | **mib-view** *view-name*]*	Use either approach. Both commands can be used to configure SNMP NMS access rights. The second command was introduced to be compatible with SNMPv3. The community name configured on NMS should be consistent with the username configured on the agent.
	Configure indirectly	Configure an SNMP group	**snmp-agent group** { **v1** | **v2c** } *group-name* [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*] [**acl** *acl-number*]	
		Add a new user to an SNMP group	**snmp-agent usm-user** { **v1** | **v2c** } *user-name group-name* [**acl** *acl-number*]	
Configure the maximum size of an SNMP packet that can be received or sent by an SNMP agent			**snmp-agent packet max-size** *byte-count*	Optional 1500 bytes by default
Configure the engine ID for a local SNMP agent			**snmp-agent local-engineid** *engineid*	Optional Company ID and device ID by default

To do…	Use the command…	Remarks
Create or update MIB view content for an SNMP agent	**snmp-agent mib-view** { **excluded** | **included** } *view-name oid-tree* [**mask** *mask-value*]	Optional ViewDefault by default

Caution

The validity of a USM user depends on the engine ID of the SNMP agent. If the engine ID when the USM user is created is not identical to the current engine ID, the USM user is invalid.

Configuring SNMP Logging

Introduction to SNMP Logging

SNMP logs the GET and SET operations that the NMS performs on the SNMP agent. When the GET operation is performed, the agent logs the IP address of the NMS, node name of the GET operation and OID of the node. When the SET operation is performed, the agent logs the IP address of the NMS, node name of the SET operation, OID of the node, the value set and the error code and error index of the SET response. These logs will be sent to the information center, and the level of them is informational, that is, they are taken as the system prompt information. With parameters for the information center set, the output rules for SNMP logs are decided (that is, whether the logs are permitted to output and the output destinations).

SNMP logs GET request, SET request and SET response, but does not log GET response.

Enabling SNMP Logging

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable SNMP logging	**snmp-agent log** { **all** | **get-operation** | **set-operation** }	Required Disabled by default.
Configure SNMP log output rules	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional By default, SNMP logs are output to loghost and logfile only. To output SNMP logs to other destinations such as console or monitor terminal, you need to set the output destinations with this command.

Note

- Logs occupy storage space of the device, thus affecting the performance of the device. Therefore, it is recommended to disable SNMP logging.
- The size of SNMP logs cannot exceed that allowed by the information center, and the total length of the node field and value field of each log record cannot exceed 1K bytes; otherwise, the exceeded part will not be output.
- For the detailed description of system information, the information center and the **info-center source** command, refer to *Information Center Configuration* in the *System Volume*.

SNMP Trap Configuration

Enabling the Trap Function

The SNMP agent sends traps to the NMS to inform the NMS of critical and important events (such as reboot of a managed device). Two types of traps are available: generic traps and self-defined traps. Generic traps supported on the device include: **authentication**, **coldstart**, **linkdown**, **linkup** and **warmstart**. The others are self-defined traps, which are generated by different modules. As traps that occupy large device memory affect device performance, it is recommended not to enable the trap function for all the modules but for the specific modules as needed.

With the trap function enabled on a module, the traps generated by the module will be sent to the information center. The information center has seven information output destinations. By default, traps of all modules are allowed to be output to the console, monitor terminal (monitor), loghost, and logfile; traps of all modules and with level equal to or higher than warnings are allowed to be output to the trapbuffer and SNMP module (snmpagent); and traps cannot be sent to the logbuffer. You can set parameters for the information center based on the levels of the traps generated by each module, and thus decide the output rules of traps (that is, whether traps are allowed to be output and the output destinations). For the configuration of the information center, refer to *Information Center Configuration* in the *System Volume*.

Follow these steps to enable the trap function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the trap function globally	**snmp-agent trap enable** [**bfd** | **bgp** | **configuration** | **flash** | **ospf** [*process-id*] [**ifauthfail** | **ifcfgerror** | **ifrxbadpkt** | **ifstatechange** | **iftxretransmit** | **lsdbapproachoverflow** | **lsdboverflow** | **maxagelsa** | **nbrstatechange** | **originatelsa** | **vifcfgerror** | **virifauthfail** | **virifrxbadpkt** | **virifstatechange** | **viriftxretransmit** | **virnbrstatechange**] * | **standard** [**authentication** | **coldstart** | **linkdown** | **linkup** | **warmstart**] * | **system** | **vrrp** [**authfailure** | **newmaster**]]	Optional By default, the trap function is enabled.
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Enable the trap function of interface state changes	**enable snmp trap updown**	Optional Enabled by default.

Caution

To enable an interface to send linkUp/linkDown traps when its state changes, you need to enable the trap function of interface state changes on an interface and globally. Use the **enable snmp trap updown** command to enable the trap function on an interface, and use the **snmp-agent trap enable** [**standard** [**linkdown** | **linkup**] *] command to enable this function globally.

Configuring Trap Parameters

Configuration prerequisites

To send traps to the NMS, you need to prepare the following:

- Basic SNMP configurations have been completed. These configurations include version configuration: community name is needed when SNMPv1 and v2c are adopted; username and MIB view are needed if SNMPv3 is adopted.
- A connection has bee established between the device and the NMS, and they can operate each other.

Configuration procedure

After traps are sent to the SNMP module, the SNMP module saves the traps in the trap queue. You can set the size of the queue and the holding time of the traps in the queue, and you can also send the traps to the specified destination host (usually the NMS).

Follow these steps to configure trap parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure target host attribute for traps	**snmp-agent target-host trap address udp-domain** { *ip-address* | **ipv6** *ipv6-address* } [**udp-port** *port-number*] [**vpn-instance** *vpn-instance-name*] **params securityname** *security-string* [**v1** | **v2c** | **v3** [**authentication** | **privacy**]]	Optional The **vpn-instance** keyword is applicable in a network supporting IPv4. To send the traps to the NMS, this command is required, and you must specify *ip-address* as the IP address of the NMS.
Configure the source address for traps	**snmp-agent trap source** *interface-type interface-number*	Optional
Extend the standard linkUp/linkDown traps defined in RFC	**snmp-agent trap if-mib link extended**	Optional Standard linkUp/linkDown traps defined in RFC are used by default.

To do…	Use the command…	Remarks
Configure the size of the trap sending queue	**snmp-agent trap queue-size** *size*	Optional 100 by default
Configure the holding time of the traps in the queue	**snmp-agent trap life** *seconds*	Optional 120 seconds by default

Note

- An extended linkUp/linkDown trap is the standard linkUp/linkDown trap (defined in RFC) appended with interface description and interface type information. If the extended messages are not supported on the NMS, disable this function to let the device send standard linkUp/linkDown traps.
- If the sending queue of traps is full, the system will automatically delete some oldest traps to receive new traps.
- The system will automatically delete the traps whose lifetime expires.

Displaying and Maintaining SNMP

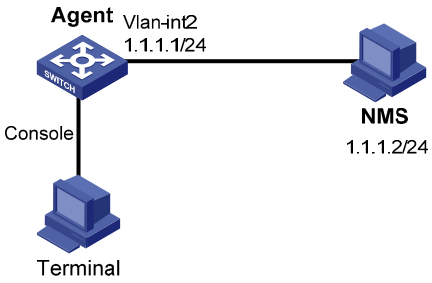
To do…	Use the command…	Remarks
Display SNMP-agent system information, including the contact, location, and version of the SNMP	**display snmp-agent sys-info** [**contact** | **location** | **version**]*	Available in any view
Display SNMP agent statistics	**display snmp-agent statistics**	
Display the SNMP agent engine ID	**display snmp-agent local-engineid**	
Display SNMP agent group information	**display snmp-agent group** [*group-name*]	
Display basic information of the trap queue	**display snmp-agent trap queue**	
Display the modules that can send traps and whether their trap sending is enabled or not	**display snmp-agent trap-list**	
Display SNMP v3 agent user information	**display snmp-agent usm-user** [**engineid** *engineid* | **username** *user-name* | **group** *group-name*] *	
Display SNMP v1 or v2c agent community information	**display snmp-agent community** [**read** | **write**]	
Display MIB view information for an SNMP agent	**display snmp-agent mib-view** [**exclude** | **include** | **viewname** *view-name*]	

SNMP Configuration Example

Network requirements

- The NMS connects to the agent, a switch, through an Ethernet.
- The IP address of the NMS is 1.1.1.2/24.
- The IP address of the VLAN interface on the switch is 1.1.1.1/24.
- The NMS monitors and manages the agent using SNMPv2c. The agent reports errors or faults to the NMS.

Figure 1-3 Network diagram for SNMP



Configuration procedure

1) Configuring the SNMP agent

Configure the SNMP basic information, including version and community name.

```
<Sysname> system-view
[Sysname] snmp-agent sys-info version v2c
[Sysname] snmp-agent community read public
[Sysname] snmp-agent community write private
```

Configure VLAN-interface 2 (with the IP address of 1.1.1.1/24). Add the port GigabitEthernet 1/0/1 to VLAN 2.

```
[Sysname] vlan 2
[Sysname-vlan2] port GigabitEthernet 1/0/1
[Sysname-Vlan2] quit
[Sysname] interface vlan-interface 2
[Sysname-Vlan-interface2] ip address 1.1.1.1 255.255.255.0
[Sysname-Vlan-interface2] quit
```

Configure the contact person and physical location information of the switch.

```
[Sysname] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Sysname] snmp-agent sys-info location telephone-closet,3rd-floor
```

Enable the sending of traps to the NMS with an IP address of 1.1.1.2/24, using **public** as the community name.

```
[Sysname] snmp-agent trap enable
[Sysname] snmp-agent target-host trap address udp-domain 1.1.1.2 udp-port 5000 params securityname public
```

2) Configuring the SNMP NMS

With SNMPv2c, the user needs to specify the read only community, the read and write community, the timeout time, and number of retries. The user can inquire and configure the device through the NMS.

Note

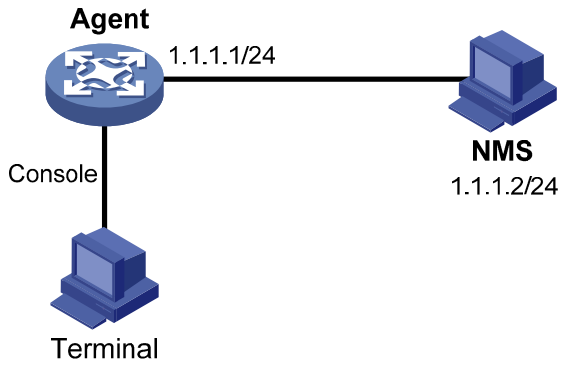
The configurations on the agent and the NMS must match.

SNMP Logging Configuration Example

Network requirements

- The NMS and the agent are connected through an Ethernet
- The IP address of the NMS is 1.1.1.2/24
- The IP address of the VLAN interface on the agent is 1.1.1.1/24
- Configure community name, access right and SNMP version on the agent

Figure 1-4 Network diagram for SNMP logging



Configuration procedure

Note

The configurations for the NMS and agent are omitted.

Enable logging display on the terminal. (This function is enabled by default, so that you can omit this configuration).

```
<Sysname> terminal monitor
<Sysname> terminal logging
```

Enable the information center to output the system information with the severity level equal to or higher than **informational** to the console port.

```
<Sysname> system-view
[Sysname] info-center source snmp channel console log level informational
```

Enable SNMP logging on the agent to log the GET and SET operations of the NMS.

```
[Sysname] snmp-agent log get-operation
[Sysname] snmp-agent log set-operation
```

- The following log information is displayed on the terminal when the NMS performs the GET operation to the agent.

```
%Jan 1 02:49:40:566 2006 Sysname SNMP/6/GET:
seqNO = <10> srcIP = <1.1.1.2> op = <get> node = <sysName(1.3.6.1.2.1.1.5.0)> value=<>
```

- The following log information is displayed on the terminal when the NMS performs the SET operation to the agent.

```
%Jan 1 02:59:42:576 2006 Sysname SNMP/6/SET:
seqNO = <11> srcIP = <1.1.1.2> op = <set> errorIndex = <0> errorStatus =<noError> node =
<sysName(1.3.6.1.2.1.1.5.0)> value = <Sysname>
```

Table 1-1 Description on the output field of SNMP log

Field	Description
Jan 1 02:49:40:566 2006	The time when SNMP log is generated
seqNO	Sequence number of the SNMP log ()
srcIP	IP address of NMS
op	SNMP operation type (GET or SET)
node	Node name of the SNMP operations and OID of the instance
erroIndex	Error index, with 0 meaning no error
errorstatus	Error status, with noError meaning no error
value	Value set when the SET operation is performed (This field is null, meaning the value obtained with the GET operation is not logged.) When the value is a string of characters and the string contains characters not in the range of ASCII 0 to 127 or invisible characters, the string is displayed in hexadecimal. For example, value = <81-43>[hex]

Note

The system information of the information center can be output to the terminal or to the log buffer. In this example, SNMP log is output to the terminal. For configuration of SNMP log output to other destinations, see *Information Center Configuration* in the *System Volume*.

2 MIB Style Configuration

3Com private MIB involves two styles, 3Com compatible MIB and 3Com new MIB. In the 3Com compatible MIB style, the device sysOID is under the 3Com's enterprise ID 25506, and the private MIB is under the enterprise ID 2011. In the 3Com new MIB style, both the device sysOID and the private MIB are under the 3Com's enterprise ID 25506. These two styles of MIBs implement the same management function except for their root nodes. A device is shipped with MIB loaded and the MIB style may vary depending on the device. To implement NMS's flexible management of the device, the device allows you to configure MIB style, that is, you can switch between the two styles of MIBs. However, you need to ensure that the MIB style of the device is the same as that of the NMS.

Setting the MIB Style

Follow these steps to set the MIB style:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the MIB style of the device	**mib-style** [**new** | **compatible**]	Optional **new** by default

Note

The modified MIB style takes effect only after you reboot the device. Therefore, you are recommended to reboot the device after setting the MIB style to ensure that the modification of the MIB style takes effect.

Displaying and Maintaining MIB

To do…	Use the command…	Remarks
Display the MIB style	**display mib-style**	Available in any view

Table of Contents

1 RMON Configuration

When configuring RMON, go to these sections for information you are interested in:

RMON Overview

This section covers these topics:

Introduction

Remote Monitoring (RMON) is implemented based on the Simple Network Management Protocol (SNMP) and is fully compatible with the existing SNMP framework without the need of any modification on SNMP.

RMON provides an efficient means of monitoring subnets and allows SNMP to monitor remote network devices in a more proactive and effective way. It reduces traffic between network management station (NMS) and agent, facilitating large network management.

RMON comprises two parts: NMSs and agents running on network devices.

- Each RMON NMS administers the agents within its administrative domain.
- An RMON agent resides on a network monitor or a network probe. It monitors and collects statistics on traffic over the network segments connected to its interfaces, such as the total number of packets passed through a network segment over a specified period, or the total number of good packets sent to a host.

Working Mechanism

RMON allows multiple monitors. A monitor provides two ways of data gathering:

- Using RMON probes. NMSs can obtain management information from RMON probes directly and control network resources. In this approach, RMON NMSs can obtain all RMON MIB information.
- Embedding RMON agents in network devices such as routers, switches, and hubs to provide the RMON probe function. RMON NMSs exchange data with RMON agents using basic SNMP commands to gather network management information, which, due to system resources limitation, may not cover all MIB information but four groups of information, alarm, event, history, and statistics, in most cases.

The device adopts the second way. By using RMON agents on network monitors, an NMS can obtain information about traffic size, error statistics, and performance statistics for network management.

RMON Groups

Among the ten RMON groups defined by RMON specifications (RFC 1757), the device supports the event group, alarm group, history group and statistics group. Besides, 3Com also defines and implements the private alarm group, which enhances the functions of the alarm group. This section describes the five kinds of groups in general.

Event group

The event group defines event indexes and controls the generation and notifications of the events triggered by the alarms defined in the alarm group and the private alarm group. The events can be handled in one of the following ways:

- Logging event related information in the event log table
- Sending traps to NMSs
- Logging event information in the event log table and sending traps to NMSs
- No action

Alarm group

The RMON alarm group monitors specified alarm variables, such as statistics on a port. If the sampled value of the monitored variable is bigger than or equal to the upper threshold, an upper event is triggered; if the sampled value of the monitored variable is lower than or equal to the lower threshold, a lower event is triggered. The event is then handled as defined in the event group.

The following is how the system handles entries in the RMON alarm table:

1) Samples the alarm variables at the specified interval.
2) Compares the sampled values with the predefined threshold and triggers events if all triggering conditions are met.

Note

If a sampled alarm variable overpasses the same threshold multiple times, only the first one can cause an alarm event. That is, the rising alarm and falling alarm are alternate.

Private alarm group

The private alarm group calculates the sampled values of alarm variables and compares the result with the defined threshold, thereby realizing a more comprehensive alarming function.

System handles the prialarm alarm table entry (as defined by the user) in the following ways:

- Periodically samples the prialarm alarm variables defined in the prialarm formula.
- Calculates the sampled values based on the prialarm formula.
- Compares the result with the defined threshold and generates an appropriate event.

Note

If the count result overpasses the same threshold multiple times, only the first one can cause an alarm event. That is, the rising alarm and falling alarm are alternate.

History group

The history group periodically collects statistics on data at interfaces and saves the statistics in the history record table for query convenience. The statistics data includes bandwidth utilization, number of error packets, and total number of packets.

Once you successfully create a history entry in the specified interface, the history group starts to periodically collect statistics on packet at the specified interface. Each statistical value is a cumulative sum of packets sent/received on the interface during a sampling period.

Ethernet statistics group

The statistics group monitors port utilization. It provides statistics about network collisions, CRC alignment errors, undersize/oversize packets, broadcasts, multicasts, bytes received, packets received, bytes sent, packets sent, and so on.

After the creation of a statistics entry on an interface, the statistics group starts to collect traffic statistics on the current interface. The result of the statistics is a cumulative sum.

Configuring RMON

Configuration Prerequisites

Before configuring RMON, configure the SNMP agent as described in *SNMP Configuration* in the *System Volume*.

Configuration Procedure

Follow these steps to configure RMON:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an event entry in the event table	**rmon event** *entry-number* [**description** *string*] { **log** | **log-trap** *log-trapcommunity* | **none** | **trap** *trap-community* } [**owner** *text*]	Optional
Enter Ethernet interface view	**interface** *interface-type interface-number*	—
Create an entry in the history table	**rmon history** *entry-number* **buckets** *number* **interval** *sampling-interval* [**owner** *text*]	Optional
Create an entry in the statistics table	**rmon statistics** *entry-number* [**owner** *text*]	Optional
Exit Ethernet interface view	**quit**	—

To do…	Use the command…	Remarks
Create an entry in the alarm table	**rmon alarm** *entry-number alarm-variable sampling-interval* { **absolute** | **delta** } **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* [**owner** *text*]	Optional
Create an entry in the private alarm table	**rmon prialarm** *entry-number prialarm-formula prialarm-des sampling-interval* { **absolute** | **changeratio** | **delta** } **rising-threshold** *threshold-value1 event-entry1* **falling-threshold** *threshold-value2 event-entry2* **entrytype** { **forever** | **cycle** *cycle-period* } [**owner** *text*]	Optional

Note

- A new entry cannot be created if its parameters are identical with the corresponding parameters of an existing entry Refer to Table 1-1 for the parameters to be compared for different entries.
- The system limits the total number of each type of entries (Refer to Table 1-1 for the detailed numbers). When the total number of an entry reaches the maximum number of entries that can be created, the creation fails.
- When you create an entry in the history table, if the specified **buckets** *number* argument exceeds the history table size supported by the device, the entry will be created. However, the validated value of the **buckets** *number* argument corresponding to the entry is the history table size supported by the device.

Table 1-1 Restrictions on the configuration of RMON

Entry	Parameters to be compared	Maximum number of entries that can be created
Event	Event description (**description** *string*), event type (**log**, **trap**, **logtrap** or **none**) and community name (*trap-community* or *log-trapcommunity*)	60
History	Sampling interval (**interval** *sampling-interval*)	100
Statistics	Only one statistics entry can be created on an interface.	100
Alarm	Alarm variable (*alarm-variable*), sampling interval (*sampling-interval*), sampling type (**absolute** or **delta**), rising threshold (*threshold-value1*) and falling threshold (*threshold-value2*)	60
Prialarm	Alarm variable formula (*alarm-variable*), sampling interval (*sampling-interval*), sampling type (**absolute, changeratio** or **delta**), rising threshold (*threshold-value1*) and falling threshold (*threshold-value2*)	50

Displaying and Maintaining RMON

To do…	Use the command…	Remarks
Display RMON statistics	**display rmon statistics** [*interface-type interface-number*]	Available in any view
Display the RMON history control entry and history sampling information	**display rmon history** [*interface-type interface-number*]	Available in any view
Display RMON alarm configuration information	**display rmon alarm** [*entry-number*]	Available in any view
Display RMON prialarm configuration information	**display rmon prialarm** [*entry-number*]	Available in any view
Display RMON events configuration information	**display rmon event** [*entry-number*]	Available in any view
Display log information for the specified or all event entries.	**display rmon eventlog** [*entry-number*]	Available in any view

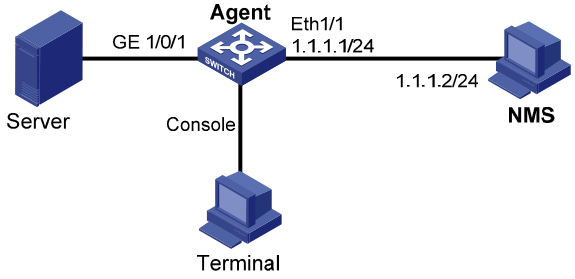
RMON Configuration Example

Network requirements

Agent is connected to a configuration terminal through its console port and to a remote NMS across the Internet.

Create an entry in the RMON Ethernet statistics table to gather statistics on GigabitEthernet 1/0/1, and enable logging after received bytes exceed the specified threshold.

Figure 1-1 Network diagram for RMON



Configuration procedure

Configure RMON to gather statistics for interface GigabitEthernet 1/0/1.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] rmon statistics 1 owner user1-rmon
[Sysname-GigabitEthernet1/0/1] quit
```

Display RMON statistics for interface GigabitEthernet 1/0/1.

```
<Sysname> display rmon statistics GigabitEthernet 1/0/1
Statistics entry 1 owned by user1-rmon is VALID.
  Interface : GigabitEthernet1/0/1<ifIndex.3>
  etherStatsOctets         : 21657     , etherStatsPkts          : 307
```

```
  etherStatsBroadcastPkts  : 56        , etherStatsMulticastPkts : 34
  etherStatsUndersizePkts  : 0         , etherStatsOversizePkts  : 0
  etherStatsFragments      : 0         , etherStatsJabbers       : 0
  etherStatsCRCAlignErrors : 0         , etherStatsCollisions    : 0
  etherStatsDropEvents (insufficient resources): 0
  Packets received according to length:
  64     : 235       ,  65-127  : 67        ,  128-255  : 4
  256-511: 1         ,  512-1023: 0         ,  1024-1518: 0
```

Create an event to start logging after the event is triggered.

```
<Sysname> system-view
[Sysname] rmon event 1 log owner 1-rmon
```

Configure an alarm group to sample received bytes on GigabitEthernet 1/0/1. When the received bytes exceed the upper or below the lower limit, logging is enabled.

```
[Sysname] rmon alarm 1 1.3.6.1.2.1.16.1.1.1.4.1 10 delta rising-threshold 1000 1
falling-threshold 100 1 owner 1-rmon
[Sysname] display rmon alarm 1
Alarm table 1 owned by 1-rmon is VALID.
  Samples type          : delta
  Variable formula      : 1.3.6.1.2.1.16.1.1.1.4.1<etherStatsOctets.1>
  Sampling interval     : 10(sec)
  Rising threshold      : 1000(linked with event 1)
  Falling threshold     : 100(linked with event 1)
  When startup enables  : risingOrFallingAlarm
  Latest value          : 2552
```

Table of Contents

1 MAC Address Table Management Configuration

When configuring MAC address table management, go to these sections for information you are interested in:

Note

- Interfaces that MAC address table management involves can only be Layer 2 Ethernet ports.
- This manual covers only the management of static, dynamic and blackhole MAC address table entries (source and destination). For the management of multicast MAC address table entries, refer to *Multicast Routing and Forwarding Configuration* in the *IP Multicast Volume*.

Introduction to MAC Address Table

A device maintains a MAC address table for frame forwarding. Each entry in this table indicates the MAC address of a connected device, ID of the interface to which this device is connected and ID of the VLAN to which the interface belongs. When forwarding a frame, the device looks up the MAC address table according to the destination MAC address of the frame to rapidly determine the egress port, thus reducing broadcasts.

How a MAC Address Table Entry is Generated

A MAC address table entry can be dynamically learned or manually configured.

Dynamically learn a MAC address table entry

Usually, MAC address tables are automatically generated during the source MAC address learning process of devices.

The following is how a device learns a MAC address after it receives a frame from a port, Port 1 for example:

1) Check the source MAC address (MAC-SOURCE for example) of the frame, that is, the MAC address of the device that sends the frame.
2) Look up the MAC address table for an entry corresponding to the MAC address and do the following:

- If an entry is found for the MAC address, update the entry.

- If no entry is found, add an entry for the MAC address to indicate from which port the frame is received.

When receiving a frame destined for MAC-SOURCE, the device then looks up the MAC address table and forwards it from Port 1.

To adapt to network changes, MAC address table entries need to be constantly updated. Each dynamically learned MAC address table entry has a life period, that is, an aging timer. If an entry is not updated before the aging timer expires, it will be deleted. If yes, the aging timer restarts the timing.

Manually configure a MAC address table entry

When a device dynamically learns MAC address table entries through source MAC address learning, it cannot tell frames of legal users from those of hackers. This brings potential security hazards. For example, if a hacker forges the MAC address of a legal user and uses it as the source MAC address of the attack frames, and accesses the device from a different port than that used by the legal user, the device will learn a forged MAC address entry, and forward frames destined for the legal user to the hacker instead.

To enhance the security of a port, you can manually add MAC address entries into the MAC address table of the device to bind specific user devices to the port, thus preventing hackers from stealing data using forged MAC addresses. Manually configured MAC address table entries have a higher priority than dynamically learned ones.

Types of MAC Address Table Entries

A MAC address table may contain the following types of entries:

- Static entries, which are manually configured and never age out.
- Dynamic entries, which can be manually configured or dynamically learned and may age out.
- Blackhole entries, which are manually configured and never age out. Blackhole entries are configured to filter frames with specific source or destination MAC addresses.

Note

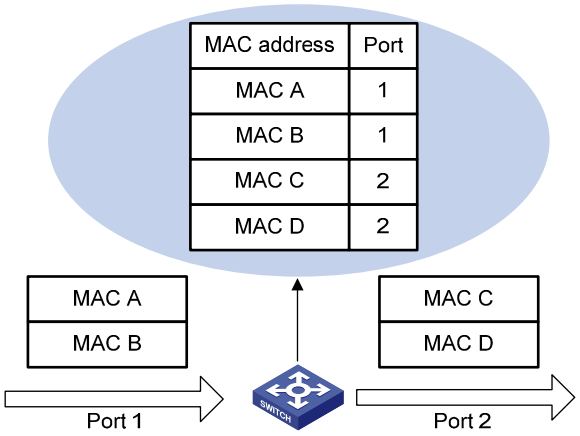
Dynamically-learned MAC addresses cannot overwrite static or blackhole MAC address entries, but the latter can overwrite the former.

MAC Address Table-Based Frame Forwarding

When forwarding a frame, the device adopts the following two forwarding modes based on the MAC address table:

- Unicast mode: If an entry is available for the destination MAC address, the device forwards the frame out the outgoing interface indicated by the MAC address table entry.
- Broadcast mode: If the device receives a frame with the destination address being all ones, or no entry is available for the destination MAC address, the device broadcasts the frame to all the interfaces except the receiving interface.

Figure 1-1 Forward frames using the MAC address table



Configuring MAC Address Table Management

The MAC address table management configuration tasks include:

- Configuring MAC Address Table Entries
- Disabling MAC Address Learning on a VLAN
- Configuring the Aging Timer for Dynamic MAC Address Entries
- Configuring the MAC Learning Limit

These configuration tasks are all optional and randomly sorted. You can choose some of the configuration tasks as required.

Configuring MAC Address Table Entries

Follow these steps to add, modify, or remove entries in the MAC address table globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Add/modify a MAC address entry	**mac-address blackhole** *mac-address* **vlan** *vlan-id*	Required
	mac-address { **dynamic** | **static** } *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id*	

Follow these steps to add, modify, or remove entries in the MAC address table on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Add/modify MAC address entries under the specified interface view	**mac-address** { **dynamic** | **static** } *mac-address* **vlan** *vlan-id*	Required

Disabling MAC Address Learning on a VLAN

You may disable MAC address learning on a per-VLAN basis.

Follow these steps to disable MAC address learning on a VLAN:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable global MAC address learning	**undo mac-address mac-learning disable**	Optional Enabled by default
Enter VLAN view	**vlan** *vlan-id*	—
Disable MAC address learning on the VLAN	**mac-address mac-learning disable**	Required Enabled by default Support for this command depends on the device model.

Note

Once MAC learning is disabled in a VLAN, all MAC address entries learnt in the VLAN are removed.

Configuring the Aging Timer for Dynamic MAC Address Entries

The MAC address table on your device is available with an aging mechanism for dynamic entries to prevent its resources from being exhausted. Set the aging timer appropriately: a long aging interval may cause the MAC address table to retain outdated entries and fail to accommodate the latest network changes; a short interval may result in removal of valid entries and hence unnecessary broadcasts which may affect device performance.

Follow these steps to configure the aging timer for dynamic MAC address entries:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the aging timer for dynamic MAC address entries	**mac-address timer** { **aging** *seconds* | **no-aging** }	Optional 300 by default.

Note

The aging timer for dynamic MAC address entries takes effect globally on dynamic MAC address entries (learned or administratively configured) only.

Configuring the MAC Learning Limit

To prevent a MAC address table from getting so large that it may degrade forwarding performance, you may restrict the number of MAC addresses that can be learned on a per-port, port group basis.

Follow these steps to configure the MAC learning limit:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enter Ethernet interface view, port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Required Use any of these three commands. The configuration you make in Ethernet interface view takes effect on the current interface only; the configuration you make in port group view takes effect on all the member ports in the port group.
	Enter port group view	**port-group manual** *port-group-name*	
Configure the maximum number of MAC addresses that can be learned on an Ethernet port view or port group		**mac-address max-mac-count** *count*	Required The default maximum number of MAC addresses that can be learned is not configured.

Displaying and Maintaining MAC Address Table Management

To do…	Use the command…	Remarks
Display MAC address table information	**display mac-address blackhole** [**vlan** *vlan-id*] [**count**]	Available in any view
	display mac-address [*mac-address* [**vlan** *vlan-id*] | [**dynamic** | **static**] [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]]	
Display the aging timer for dynamic MAC address entries	**display mac-address aging-time**	
Display MAC address statistics	**display mac-address statistics**	

MAC Address Table Management Configuration Example

Network requirements

Log onto your device from the Console port to configure MAC address table management as follows:

- Set the aging timer to 500 seconds for dynamic MAC address entries.
- Add a static entry 000f-e235-dc71 for port GigabitEthernet 1/0/1 in VLAN 1.

Configuration procedure

Add a static MAC address entry.

```
<Sysname> system-view
[Sysname] mac-address static 000f-e235-dc71 interface gigabitethernet 1/0/1 vlan 1
```

Set the aging timer for dynamic MAC address entries to 500 seconds.

```
[Sysname] mac-address timer aging 500
```

Display the MAC address entry for port GigabitEthernet 1/0/1.

```
[Sysname] display mac-address interface gigabitethernet 1/0/1
```

```
MAC ADDR          VLAN ID  STATE           PORT INDEX              AGING TIME(s)

000f-e235-dc71    1        Config static   GigabitEthernet 1/0/1   NOAGED

  ---  1 mac address(es) found  ---
```

2 MAC Information Configuration

When configuring MAC Information, go to these sections for information you are interested in:

Overview

Introduction to MAC Information

To monitor a network, you need to monitor users joining and leaving the network. Because a MAC address uniquely identifies a network user, you can monitor users joining and leaving a network by monitoring their MAC addresses.

With the MAC Information function, Layer-2 Ethernet interfaces send Syslog or Trap messages to the monitor end in the network when they learn or delete MAC addresses. By analyzing these messages, the monitor end can monitor users accessing the network.

How MAC Information Works

When a new MAC address is learned or an existing MAC address is deleted on a device, the device writes related information about the MAC address to the buffer area used to store user information. When the timer set for sending MAC address monitoring Syslog or Trap messages expires, or when the buffer is used up, the device sends the Syslog or Trap messages to the monitor end immediately.

Configuring MAC Information

The MAC Information configuration tasks include:

Enabling MAC Information Globally

Follow these steps to enable MAC Information globally:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable MAC Information globally	**mac-address information enable**	Required Disabled by default

Enabling MAC Information on an Interface

Follow these steps to enable MAC Information on an interface:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Enable MAC Information on the interface	**mac-address information enable** { **added** | **deleted** }	Required Disabled by default

Note

To enable MAC Information on an Ethernet interface, enable MAC Information globally first.

Configuring MAC Information Mode

Follow these steps to configure MAC Information mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure MAC Information mode	**mac-address information mode** { **syslog** | **trap** }	Optional **trap** by default

Configuring the Interval for Sending Syslog or Trap Messages

To prevent Syslog or Trap messages being sent too frequently and thus affecting system performance, you can set the interval for sending Syslog or Trap messages.

Follow these steps to set the interval for sending Syslog or Trap messages:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the interval for sending Syslog or Trap messages	**mac-address information interval** *interval-time*	Optional One second by default.

Configuring the MAC Information Queue Length

To avoid losing user MAC address information, when the buffer storing user MAC address information is used up, the user MAC address information in the buffer is sent to the monitor end in the network, even if the timer set for sending MAC address monitoring Syslog or Trap messages has not expired yet.

Follow these steps to configure the MAC Information queue length:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the MAC Information queue length	**mac-address information queue-length** *value*	Optional 50 by default

Note

Setting the MAC Information queue length to 0 indicates that the device sends a Syslog or Trap message to the network management device as soon as a new MAC address is learned or an existing MAC address is deleted.

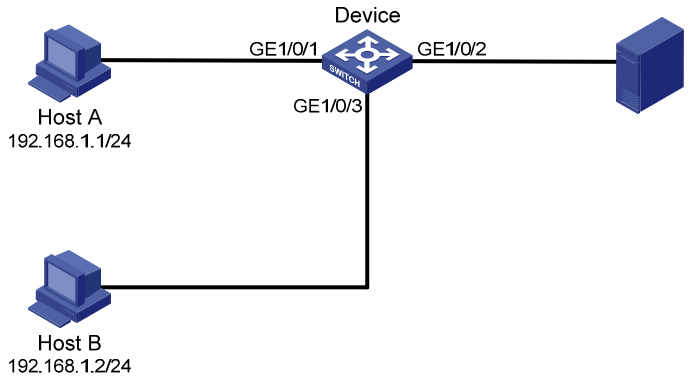
MAC Information Configuration Example

MAC Information Configuration Example

Network requirements

- Host A is connected to a remote server (Server) through Device.
- Enable MAC Information on GigabitEthernet 1/0/1 on Device. Device sends MAC address change information using Syslog messages to Host B through GigabitEthernet 1/0/3. Host B analyzes and displays the Syslog messages.

Figure 2-1 Network diagram for MAC Information configuration



Configuration procedure

1) Configure Device to send Syslog messages to Host B.

Refer to *Information Center Configuration* in the *System Volume* for details.

2) Enable MAC Information.

Enable MAC Information on Device.

```
<Device> system-view
[Device] mac-address information enable
```

Configure MAC Information mode as Syslog.

```
[Device] mac-address information mode syslog
```

Enable MAC Information on GigabitEthernet 1/0/1

```
[Device] interface gigabitethernet 1/0/1
[Device-GigabitEthernet1/0/1] mac-address information enable added
[Device-GigabitEthernet1/0/1] mac-address information enable deleted
[Device-GigabitEthernet1/0/1] quit
```

Set the MAC Information queue length to 100.

```
[Device] mac-address information queue-length 100
```

Set the interval for sending Syslog or Trap messages to 20 seconds.

```
[Device] mac-address information interval 20
```

Table of Contents

1 System Maintaining and Debugging

When maintaining and debugging the system, go to these sections for information you are interested in:

System Maintaining and Debugging Overview

Introduction to System Maintaining

You can use the **ping** command and the **tracert** command to verify the current network connectivity.

The ping command

You can use the **ping** command to verify whether a device with a specified address is reachable, and to examine network connectivity.

The **ping** command involves the following steps in its execution:

1) The source device sends an ICMP echo request to the destination device.
2) If the network is functioning properly, the destination device responds by sending an ICMP echo reply to the source device after receiving the ICMP echo request.
3) If there is network failure, the source device displays timeout or destination unreachable.
4) The source device displays related statistics.

Output of the **ping** command falls into the following:

- The **ping** command can be applied to the destination's name or IP address. If the destination's name is unknown, the prompt information is displayed.
- Information on the destination's responses towards each ICMP echo request. If the source device does not receive an ICMP echo reply within the timeout time, it displays the prompt information. If the source device receives an ICMP echo reply within the timeout time, it displays the number of bytes of the echo reply, the message sequence number, Time to Live (TTL), the response time, and the statistics during the ping operation. The statistics include number of packets sent, number of echo reply messages received, percentage of packets not responded to the total packets sent, and the minimum, average, and maximum response time.

The tracert command

By using the **tracert** command, you can trace the Layer 3 devices involved in delivering a packet from source to destination. This is useful for identification of failed node(s) in the event of network failure.

The **tracert** command involves the following steps in its execution:

1) The source device sends a packet with a TTL value of 1 to the destination device.
2) The first hop (the Layer 3 device that first receives the packet) responds by sending a TTL-expired ICMP message to the source, with its IP address encapsulated. In this way, the source device can get the address of the first Layer 3 device.

3) The source device sends a packet with a TTL value of 2 to the destination device.
4) The second hop responds with a TTL-expired ICMP message, which gives the source device the address of the second Layer 3 device.
5) The above process continues until the ultimate destination device is reached. In this way, the source device can trace the addresses of all the Layer 3 devices involved to get to the destination device.

Introduction to System Debugging

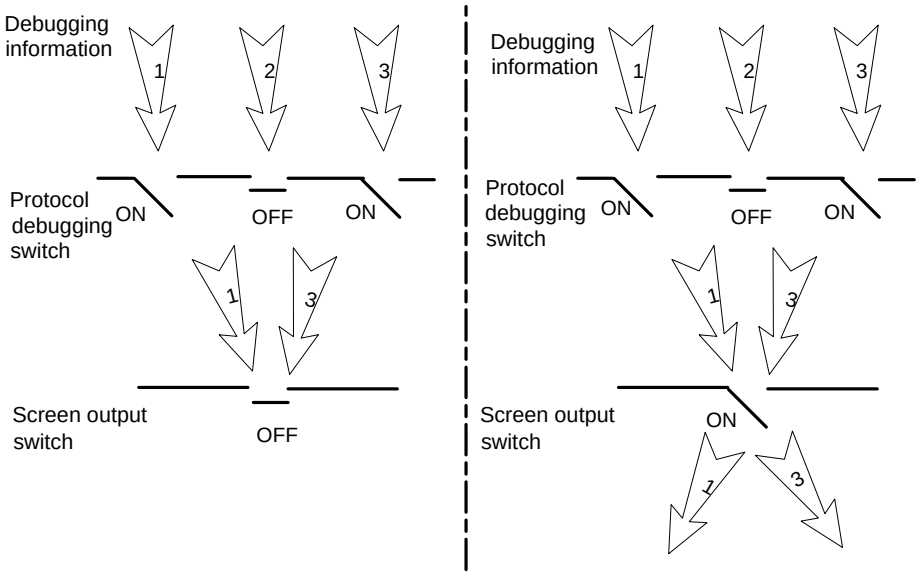
The device provides various debugging functions. For the majority of protocols and features supported, the system provides corresponding debugging information to help users diagnose errors.

The following two switches control the display of debugging information:

- Protocol debugging switch, which controls protocol-specific debugging information.
- Screen output switch, which controls whether to display the debugging information on a certain screen.

As Figure 1-1 illustrates, suppose the device can provide debugging for the three modules 1, 2, and 3. Only when both the protocol debugging switch and the screen output switch are turned on can debugging information be output on a terminal.

Figure 1-1 The relationship between the protocol and screen debugging switch



Note

Outputting debugging information to a terminal is most commonly used. You can also configure to output debugging information to other directions. For detailed configuration, refer to *Information Center Configuration* in the *System Volume*.

System Maintaining and Debugging

System Maintaining

To do…	Use the command…	Remarks
Check whether a specified IP address can be reached	**ping** [**ip**] [**-a** *source-ip* | **-c** *count* | **-f** | **-h** *ttl* | **-i** *interface-type interface-number* | **-m** *interval* | **-n** | **-p** *pad* | **-q** | **-r** | **-s** *packet-size* | **-t** *timeout* | **-tos** *tos* | **-v** | **-vpn-instance** *vpn-instance-name*] * *remote-system*	Optional Used in IPv4 network Available in any view
	ping ipv6 [**-a** *source-ipv6* | **-c** *count* | **-m** *interval* | **-s** *packet-size* | **-t** *timeout*] * *remote-system* [**-i** *interface-type interface-number*]	Optional Used in IPv6 network Available in any view
View the route from the source to the destination	**tracert** [**-a** *source-ip* | **-f** *first-ttl* | **-m** *max-ttl* | **-p** *port* | **-q** *packet-number* | **-vpn-instance** *vpn-instance-name* | **-w** *timeout*] * *remote-system*	Optional Used in IPv4 network Available in any view
	tracert ipv6 [**-f** *first-ttl* | **-m** *max-ttl* | **-p** *port* | **-q** *packet-number* | **-w** *timeout*] * *remote-system*	Optional Used in IPv6 network Available in any view

Note

- For a low-speed network, you are recommended to set a larger value for the timeout timer (indicated by the **-t** parameter in the command) when configuring the **ping** command.
- Only the directly connected segment address can be pinged if the outgoing interface is specified with the **-i** argument.

System Debugging

To do…	Use the command…	Remarks
Enable the terminal monitoring of system information	**terminal monitor**	Optional The terminal monitoring on the console is enabled by default and that on the monitoring terminal is disabled by default. Available in user view
Enable the terminal display of debugging information	**terminal debugging**	Required Disabled by default Available in user view
Enable debugging for a specified module	**debugging** { **all** [**timeout** *time*] | *module-name* [*option*] }	Required Disabled by default Available in user view

To do…	Use the command…	Remarks
Display the enabled debugging functions	**display debugging** [**interface** *interface-type interface-number*] [*module-name*]	Optional Available in any view

Note

- The **debugging** commands are usually used by administrators in diagnosing network failure.
- Output of the debugging information may reduce system efficiency, especially during execution of the **debugging all** command.
- After completing the debugging, you are recommended to use the **undo debugging all** command to disable all the debugging functions.
- You must configure the **debugging**, **terminal debugging** and **terminal monitor** commands first to display the detailed debugging information on the terminal. For the detailed description on the **terminal debugging** and **terminal monitor** commands, refer to *Information Center Commands* in the *System Volume*.

System Maintaining Example

Network requirements

- The IP address of the destination device is 10.1.1.4.
- Display the Layer 3 devices involved while packets are forwarded from the source device to the destination device.

Configuration procedure

```
<Sysname> tracert 10.1.1.4
traceroute to 10.1.1.4 (10.1.1.4) 30 hops max, 40 bytes packet
1  128.3.112.1  19 ms  19 ms  0 ms
2  128.32.216.1  39 ms  39 ms  19 ms
3  128.32.136.23  39 ms  40 ms  39 ms
4  128.32.168.22  39 ms  39 ms  39 ms
5  128.32.197.4  40 ms  59 ms  59 ms
6  131.119.2.5  59 ms  59 ms  59 ms
7  129.140.70.13  99 ms  99 ms  80 ms
8  129.140.71.6  139 ms  239 ms  319 ms
9  129.140.81.7  220 ms  199 ms  199 ms
10  10.1.1.4  239 ms  239 ms  239 ms
```

The above output shows that nine Layer 3 devices are used from the source to the destination device.

Table of Contents

1 Information Center Configuration

When configuring information center, go to these sections for information you are interested in:

Information Center Overview

Introduction to Information Center

Acting as the system information hub, information center classifies and manages system information, offering a powerful support for network administrators and developers in monitoring network performance and diagnosing network problems.

The following describes the working process of information center:

- Receives the log, trap, and debugging information generated by each module.
- Outputs the above information to different information channels according to the user-defined output rules.
- Outputs the information to different destinations based on the information channel-to-destination associations.

To sum up, information center assigns the log, trap and debugging information to the ten information channels according to the eight severity levels and then outputs the information to different destinations. The following describes the working process in details.

Note

By default, the information center is enabled. An enabled information center affects the system performance in some degree due to information classification and output. Such impact becomes more obvious in the event that there is enormous information waiting for processing.

Classification of system information

The system information of the information center falls into three types:

- Log information
- Trap information
- Debugging information

Eight levels of system information

The information is classified into eight levels by severity. The severity levels in the descending order are emergency, alert, critical, error, warning, notice, informational and debug. When the system information is output by level, the information with severity level higher than or equal to the specified level is output. For example, in the output rule, if you configure to output information with severity level being informational, the information with severity level being emergency through informational is all allowed to be output.

Table 1-1 Severity description

Severity	Severity value	Description
Emergency	0	The system is unusable.
Alert	1	Action must be taken immediately
Critical	2	Critical conditions
Error	3	Error conditions
Warning	4	Warning conditions
Notice	5	Normal but significant condition
Informational	6	Informational messages
Debug	7	Debug-level messages

Six output destinations and ten channels of system information

The system supports six information output destinations, including the console, monitor terminal (monitor), log buffer, log host, trap buffer and SNMP module. The specific destinations supported vary with devices.

The system supports ten channels. The six channels 0 through 5 are configured with channel names, output rules, and are associated with output destinations by default. The channel names, output rules and the associations between the channels and output destinations can be changed through commands. Besides, you can configure channels 6, 7, 8, and 9 without changing the default configuration of the six channels.

Table 1-2 Information channels and output destinations

Information channel number	Default channel name	Default output destination
0	console	Console (Receives log, trap and debugging information)
1	monitor	Monitor terminal (Receives log, trap and debugging information, facilitating remote maintenance)
2	loghost	Log host (Receives log, trap and debugging information and information will be stored in files for future retrieval.)
3	trapbuffer	Trap buffer (Receives trap information, a buffer inside the router for recording information.)

Information channel number	Default channel name	Default output destination
4	logbuffer	Log buffer (Receives log and debugging information, a buffer inside the router for recording information.)
5	snmpagent	SNMP module (Receives trap information)
6	channel6	Not specified (Receives log, trap, and debugging information)
7	channel7	Not specified (Receives log, trap, and debugging information)
8	channel8	Not specified (Receives log, trap, and debugging information)

Note

Configurations for the six output destinations function independently and take effect only after the information center is enabled.

Outputting system information by source module

The system is composed of a variety of protocol modules, board drivers, and configuration modules. The system information can be classified, filtered, and output according to source modules. You can use the **info-center source ?** command to view the supported information source modules.

Default output rules of system information

The default output rules define the source modules allowed to output information on each output destination, the output information type, and the output information level as shown in Table 1-3, which indicates that by default and in terms of all modules:

- Log information with severity level equal to or higher than informational is allowed to be output to the log host; log information with severity level equal to or higher than warning is allowed to be output to the console, monitor terminal, and log buffer; log information is not allowed to be output to the trap buffer and the SNMP module.
- All trap information is allowed to be output to the console, monitor terminal and log host; trap information with severity level equal to or higher than warning is allowed to be output to the trap buffer and SNMP module; trap information is not allowed to be output to the log buffer.
- All debugging information is allowed to be output to the console and monitor terminal; debugging information is not allowed to be output to the log host, log buffer, trap buffer and the SNMP module.

Table 1-3 Default output rules for different output destinations

Output destination	Modules allowed	LOG		TRAP		DEBUG	
		Enabled/ disabled	Severity	Enabled/ disabled	Severity	Enabled/ disabled	Severity
Console	default (all modules)	Enabled	Warning	Enabled	Debug	Enabled	Debug

Output destination	Modules allowed	LOG		TRAP		DEBUG	
		Enabled/ disabled	Severity	Enabled/ disabled	Severity	Enabled/ disabled	Severity
Monitor terminal	default (all modules)	Enabled	Warning	Enabled	Debug	Enabled	Debug
Log host	default (all modules)	Enabled	Informational	Enabled	Debug	Disabled	Debug
Trap buffer	default (all modules)	Disabled	Informational	Enabled	Warning	Disabled	Debug
Log buffer	default (all modules)	Enabled	Warning	Disabled	Debug	Disabled	Debug
SNMP module	default (all modules)	Disabled	Debug	Enabled	Warning	Disabled	Debug

System Information Format

The format of system information varies with the output destinations.

- If the output destination is not the log host (such as console, monitor terminal, logbuffer, trapbuffer, SNMP), the system information is in the following format:

```
timestamp sysname module/level/digest:content
```

For example, a monitor terminal connects to the device. When a terminal logs in to the device, the log information in the following format is displayed on the monitor terminal:

```
%Jun 26 17:08:35:809 2008 Sysname SHELL/4/LOGIN: VTY login from 1.1.1.1
```

- If the output destination is the log host, the system information is in the following format according to RFC 3164 (The BSD Syslog Protocol):

```
<Int_16>timestamp sysname %%nnmodule/level/digest: source content
```

Note

- The closing set of angel brackets < >, the space, the forward slash /, and the colon are all required in the above format.
- The format in the previous part is the original format of system information, so you may see the information in a different format. The displayed format depends on the tools you use to view the logs.

What follows is a detailed explanation of the fields involved:

Int_16 (priority)

The priority is calculated using the following formula: facility*8+severity, in which facility represents the logging facility name and can be configured when you set the log host parameters. The facility ranges from local0 to local7 (16 to 23 in decimal integers) and defaults to local7. The facility is mainly used to mark different log sources on the log host, query and filter the logs of the corresponding log source. Severity ranges from 0 to 7. Table 1-1 details the value and meaning associated with each severity.

Note that the priority field takes effect only when the information has been sent to the log host.

timestamp

Timestamp records the time when system information is generated to allow users to check and identify system events. You can use the **info-center timestamp** command to configure whether to include a timestamp in the system information as well as the timestamp format if it is included. The time stamp of the system information sent from the information center to the log host is with a precision of seconds, whereas that of the system information sent from the information center to the other destinations is with a precision of milliseconds.

sysname

Sysname is the system name of the current host. You can use the **sysname** command to modify the system name. (Refer to *Basic System Configuration Commands* in the *System Volume* for details)

%%

This field is a preamble used to identify a vendor. It is displayed only when the output destination is log host.

nn

This field is a version identifier of syslog. It is displayed only when the output destination is log host.

module

The module field represents the name of the module that generates system information. You can enter the **info-center source ?** command in system view to view the module list.

level (severity)

System information can be divided into eight levels based on its severity, from 0 to 7. Refer to Table 1-1 for definition and description of these severity levels. The levels of system information generated by modules are predefined by developers, and you cannot change the system information levels. However, you can configure to output information of the specified level using the **info-center source** command.

digest

The digest field is a string of up to 32 characters, outlining the system information.

For system information destined to the log host:

- If the character string ends with (l), the information is log information
- If the character string ends with (t), the information is trap information
- If the character string ends with (d), the information is debugging information

For system information destined to other destinations:

- If the timestamp starts with a %, the information is log information
- If the timestamp starts with a #, the information is trap information

- If the timestamp starts with a *, the information is debugging information

source

This field indicates the source of the information, such as the IRF member ID, or the source IP address of the log sender. This field is optional and is displayed only when the output destination is the log host.

content

This field provides the content of the system information.

Configuring Information Center

Information Center Configuration Task List

Complete the following tasks to configure information center:

Task	Remarks
Outputting System Information to the Console	Optional
Outputting System Information to a Monitor Terminal	Optional
Outputting System Information to a Log Host	Optional
Outputting System Information to the Trap Buffer	Optional
Outputting System Information to the Log Buffer	Optional
Outputting System Information to the SNMP Module	Optional
Configuring Synchronous Information Output	Optional

Outputting System Information to the Console

Outputting system information to the console

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.
Configure the channel through which system information can be output to the console	**info-center console channel** { *channel-number* | *channel-name* }	Optional By default, system information is output to the console through channel 0 (known as console).
Configure the output rules of system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.

To do…	Use the command…	Remarks
Configure the format of the time stamp	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	Optional The time stamp format for log, trap and debugging information is **date** by default.

Enabling the display of system information on the console

After setting to output system information to the console, you need to enable the associated display function to display the output information on the console.

Follow these steps in user view to enable the display of system information on the console:

To do…	Use the command…	Remarks
Enable the monitoring of system information on the console	**terminal monitor**	Optional Enabled on the console and disabled on the monitor terminal by default.
Enable the display of debugging information on the console	**terminal debugging**	Required Disabled by default
Enable the display of log information on the console	**terminal logging**	Optional Enabled by default
Enable the display of trap information on the console	**terminal trapping**	Optional Enabled by default

Outputting System Information to a Monitor Terminal

System information can also be output to a monitor terminal, which is a user terminal that has login connections through the AUX, VTY user interface.

Outputting system information to a monitor terminal

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.
Configure the channel through which system information can be output to a monitor terminal	**info-center monitor channel** { *channel-number* | *channel-name* }	Optional By default, system information is output to the monitor terminal through channel 1 (known as monitor).

To do…	Use the command…	Remarks
Configure the output rules of the system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.
Configure the format of the time stamp	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	Optional By default, the time stamp format for log, trap and debugging information is **date**.

Enabling the display of system information on a monitor terminal

After setting to output system information to a monitor terminal, you need to enable the associated display function in order to display the output information on the monitor terminal.

Follow these steps to enable the display of system information on a monitor terminal:

To do…	Use the command…	Remarks
Enable the monitoring of system information on a monitor terminal	**terminal monitor**	Required Enabled on the console and disabled on the monitor terminal by default.
Enable the display of debugging information on a monitor terminal	**terminal debugging**	Required Disabled by default
Enable the display of log information on a monitor terminal	**terminal logging**	Optional Enabled by default
Enable the display of trap information on a monitor terminal	**terminal trapping**	Optional Enabled by default

Outputting System Information to a Log Host

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.

To do…	Use the command…	Remarks
Specify a log host and configure the parameters when system information is output to the log host	**info-center loghost** *host-ip* [**channel** { *channel-number* | *channel-name* } | **facility** *local-number*] *	Required By default, the system does not output information to a log host. If you specify to output system information to a log host, the system uses channel 2 (loghost) by default.
Configure the output rules of the system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.
Specify the source IP address for the log information	**info-center loghost source** *interface-type interface-number*	Optional By default, the source interface is determined by the matched route, and the primary IP address of this interface is the source IP address of the log information.
Configure the format of the time stamp for system information output to the log host	**info-center timestamp loghost** { **date** | **no-year-date** | **none** }	Optional **date** by default.

Outputting System Information to the Trap Buffer

Note

The trap buffer receives the trap information only, and discards the log and debugging information even if you have configured to output them to the trap buffer.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.
Configure the channel through which system information can be output to the trap buffer and specify the buffer size	**info-center trapbuffer** [**channel** { *channel-number* | *channel-name* } | **size** *buffersize*] *	Optional By default, system information is output to the trap buffer through channel 3 (known as trapbuffer) and the default buffer size is 256.

To do…	Use the command…	Remarks
Configure the output rules of the system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.
Configure the format of the time stamp	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	Optional The time stamp format for log, trap and debugging information is **date** by default.

Outputting System Information to the Log Buffer

Note

You can configure to output log, trap, and debugging information to the log buffer, but the log buffer receives the log and debugging information only, and discards the trap information.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default.
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.
Configure the channel through which system information can be output to the log buffer and specify the buffer size	**info-center logbuffer** [**channel** { *channel-number* | *channel-name* } | **size** *buffersize*] *	Optional By default, system information is output to the log buffer through channel 4 (known as logbuffer) and the default buffer size is 512.
Configure the output rules of the system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.
Configure the format of the timestamp	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	Optional The time stamp format for log, trap and debugging information is **date** by default.

Outputting System Information to the SNMP Module

Note

The SNMP module receives the trap information only, and discards the log and debugging information even if you have configured to output them to the SNMP module.

To monitor the device running status, trap information is usually sent to the SNMP network management station (NMS). In this case, you need to configure to send traps to the SNMP module, and then set the trap sending parameters for the SNMP module to further process traps. For details, refer to *SNMP Configuration* in the *System Volume*.

Follow these steps to configure to output system information to the SNMP module:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable information center	**info-center enable**	Optional Enabled by default
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 1-2 for default channel names.
Configure the channel through which system information can be output to the SNMP module	**info-center snmp channel** { *channel-number* | *channel-name* }	Optional By default, system information is output to the SNMP module through channel 5 (known as snmpagent).
Configure the output rules of the system information	**info-center source** { *module-name* | **default** } **channel** { *channel-number* | *channel-name* } [**debug** { **level** *severity* | **state** *state* } * | **log** { **level** *severity* | **state** *state* } * | **trap** { **level** *severity* | **state** *state* } *] *	Optional Refer to Default output rules of system information.
Configure the format of the timestamp	**info-center timestamp** { **debugging** | **log** | **trap** } { **boot** | **date** | **none** }	Optional The time stamp format for log, trap and debugging information is **date** by default.

Configuring Synchronous Information Output

Synchronous information output refers to the feature that if the user's input is interrupted by system output such as log, trap, or debugging information, then after the completion of system output the system will display a command line prompt (a prompt in command editing mode, or a [Y/N] string in interaction mode) and your input so far.

This command is used in the case that your input is interrupted by a large amount of system output. With this feature enabled, you can continue your operations from where you were stopped.

Follow these steps to enable synchronous information output:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable synchronous information output	**info-center synchronous**	Required Disabled by default

Note

- If system information, such as log information, is output before you input any information under the current command line prompt, the system will not display the command line prompt after the system information output.
- If system information is output when you are inputting some interactive information (non Y/N confirmation information), then after the system information output, the system will not display the command line prompt but your previous input in a new line.

Disabling a Port from Generating Link Up/Down Logging Information

By default, all the ports of the device generate link up/down logging information when the port state changes. Therefore, you may need to use this function in some cases, for example:

- You only concern the states of some of the ports. In this case, you can use this function to disable the other ports from generating link up/down logging information.
- The state of a port is not stable, and therefore redundant logging information will be generated. In this case, you can use this function to disable the port from generating link up/down logging information.

Follow the steps below to disable a port from generating link up/down logging information:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Disable the port from generating link up/down logging information	**undo enable log updown**	Required By default, all ports are allowed to generate link up/down logging information when the port state changes.

Note

With this feature applied to a port, when the state of the port changes, the system does not generate port link up/down logging information. In this case, you cannot monitor the port state changes conveniently. Therefore, it is recommended to use the default configuration in normal cases.

Displaying and Maintaining Information Center

To do…	Use the command…	Remarks
Display information about information channels	**display channel** [*channel-number* | *channel-name*]	Available in any view
Display the information of each output destination	**display info-center**	Available in any view
Display the state of the log buffer and the log information recorded	**display logbuffer** [**reverse**] [**level** *severity* | **size** *buffersize* | **slot** *slot-number*] * [**|** { **begin** | **exclude** | **include** } *regular-expression*]	Available in any view
Display a summary of the log buffer	**display logbuffer summary** [**level** *severity* **| slot** *slot-number*] *	Available in any view
Display the content of the log file buffer	**display logfile buffer**	Available in any view
Display the configuration of the log file	**display logfile summary**	Available in any view
Display the state of the trap buffer and the trap information recorded	**display trapbuffer** [**reverse**] [**size** *buffersize*]	Available in any view
Reset the log buffer	**reset logbuffer**	Available in user view
Reset the trap buffer	**reset trapbuffer**	Available in user view

Information Center Configuration Examples

Outputting Log Information to a Unix Log Host

Network requirements

- Send log information to a Unix log host with an IP address of 1.2.0.1/16;
- Log information with severity higher than informational will be output to the log host;
- The source modules are ARP and IP.

Figure 1-1 Network diagram for outputting log information to a Unix log host

Configuration procedure

Before the configuration, make sure that there is a route between Device and PC.

1) Configure the device

Enable information center.

```
<Sysname> system-view
```

```
[Sysname] info-center enable
```

Specify the host with IP address 1.2.0.1/16 as the log host, use channel **loghost** to output log information (optional, **loghost** by default), and use **local4** as the logging facility.

```
[Sysname] info-center loghost 1.2.0.1 channel loghost facility local4
```

Disable the output of log, trap, and debugging information of all modules on channel **loghost**.

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```

Caution

As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (**loghost** in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of ARP and IP modules with severity equal to or higher than **informational** to be output to the log host. (Note that the source modules allowed to output information depend on the device model.)

```
[Sysname] info-center source arp channel loghost log level informational state on
[Sysname] info-center source ip channel loghost log level informational state on
```

2) Configure the log host

The following configurations were performed on SunOS 4.0 which has similar configurations to the Unix operating systems implemented by other vendors.

Step 1: Log in to the log host as a root user.

Step 2: Create a subdirectory named **Device** under directory **/var/log/**, and create file **info.log** under the **Device** directory to save logs of **Device**.

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

Step 3: Edit file **/etc/syslog.conf** and add the following contents.

```
# Device configuration messages
local4.info    /var/log/Device/info.log
```

In the above configuration, **local4** is the name of the logging facility used by the log host to receive logs. **info** is the information level. The Unix system will record the log information with severity level equal to or higher than **informational** to file **/var/log/Device/info.log**.

Note

Be aware of the following issues while editing file **/etc/syslog.conf**:

- Comments must be on a separate line and begin with the # sign.
- No redundant spaces are allowed after the file name.
- The logging facility name and the information level specified in the **/etc/syslog.conf** file must be identical to those configured on the device using the **info-center loghost** and **info-center source** commands; otherwise the log information may not be output properly to the log host.

Step 4: After log file **info.log** is created and file **/etc/syslog.conf** is modified, you need to issue the following commands to display the process ID of **syslogd**, kill the **syslogd** process and then restart **syslogd** using the **–r** option to make the modified configuration take effect.

```
# ps -ae | grep syslogd
147
# kill -HUP 147
# syslogd -r &
```

After the above configurations, the system will be able to record log information into the log file.

Outputting Log Information to a Linux Log Host

Network requirements

- Send log information to a Linux log host with an IP address of 1.2.0.1/16;
- Log information with severity higher than informational will be output to the log host;
- All modules can output log information.

Figure 1-2 Network diagram for outputting log information to a Linux log host

Configuration procedure

Before the configuration, make sure that there is a route between Device and PC.

1) Configure the device

Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

Specify the host with IP address 1.2.0.1/16 as the log host, use channel **loghost** to output log information (optional, **loghost** by default), and use **local5** as the logging facility.

```
[Sysname] info-center loghost 1.2.0.1 channel loghost facility local5
```

Disable the output of log, trap, and debugging information of all modules on channel **loghost**.

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```

Caution

As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (**loghost** in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of all modules with severity equal to or higher than **informational** to be output to the log host.

```
[Sysname] info-center source default channel loghost log level informational state on
```

2) Configure the log host

Step 1: Log in to the log host as a root user.

Step 2: Create a subdirectory named **Device** under directory **/var/log/**, and create file **info.log** under the **Device** directory to save logs of **Device**.

```
# mkdir /var/log/Device
# touch /var/log/Device/info.log
```

Step 3: Edit file **/etc/syslog.conf** and add the following contents.

```
# Device configuration messages
local5.info    /var/log/Device/info.log
```

In the above configuration, **local5** is the name of the logging facility used by the log host to receive logs. **info** is the information level. The Linux system will record the log information with severity level equal to or higher than **informational** to file **/var/log/Device/info.log**.

Note

Be aware of the following issues while editing file **/etc/syslog.conf**:

- Comments must be on a separate line and begin with the # sign.
- No redundant spaces are allowed after the file name.
- The logging facility name and the information level specified in the **/etc/syslog.conf** file must be identical to those configured on the device using the **info-center loghost** and **info-center source** commands; otherwise the log information may not be output properly to the log host.

Step 4: After log file **info.log** is created and file **/etc/syslog.conf** is modified, you need to issue the following commands to display the process ID of **syslogd**, kill the **syslogd** process, and restart **syslogd** using the **-r** option to make the modified configuration take effect.

```
# ps -ae | grep syslogd
147
# kill -9 147
```

```
# syslogd -r &
```

Note

Ensure that the **syslogd** process is started with the -r option on a Linux log host.

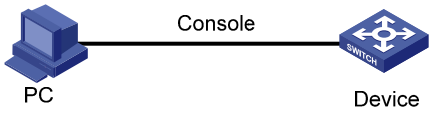
After the above configurations, the system will be able to record log information into the log file.

Outputting Log Information to the Console

Network requirements

- Log information with a severity higher than informational will be output to the console;
- The source modules are ARP and IP.

Figure 1-3 Network diagram for sending log information to the console



Configuration procedure

Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

Use channel **console** to output log information to the console (optional, **console** by default).

```
[Sysname] info-center console channel console
```

Disable the output of log, trap, and debugging information of all modules on channel **console**.

```
[Sysname] info-center source default channel console debug state off log state off trap state off
```

Caution

As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (**console** in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of ARP and IP modules with severity equal to or higher than **informational** to be output to the console. (Note that the source modules allowed to output information depend on the device model.)

```
[Sysname] info-center source arp channel console log level informational state on
[Sysname] info-center source ip channel console log level informational state on
[Sysname] quit
```

Enable the display of log information on a terminal. (Optional, this function is enabled by default.)

```
<Sysname> terminal monitor
% Current terminal monitor is on
<Sysname> terminal logging
% Current terminal logging is on
```

After the above configuration takes effect, if the specified module generates log information, the information center automatically sends the log information to the console, which then displays the information.

Table of Contents

1 PoE Configuration

When configuring PoE, go to these sections for information you are interested in:

PoE Overview

Introduction to PoE

Power over Ethernet (PoE) means that power sourcing equipment (PSE) supplies power to powered devices (PD) such as IP telephone, wireless LAN access point, and web camera from Ethernet interfaces through twisted pair cables.

Advantages

- Reliable: Power is supplied in a centralized way so that it is very convenient to provide a backup power supply.
- Easy to connect: A network terminal requires only one Ethernet cable, but no external power supply.
- Standard: In compliance with IEEE 802.3af, and a globally uniform power interface is adopted.
- Promising: It can be applied to IP telephones, wireless LAN access points, portable chargers, card readers, web cameras, and data collectors.

Composition

A PoE system consists of PoE power, PSE, and PD.

- PoE power

The whole PoE system is powered by the PoE power, which includes external PoE power and internal PoE power.

- PSE

PSE detecting that a PD is unplugged, the PSE stops supplying power to the PD.

An Ethernet interface with the PoE capability is called PoE interface. Currently, a PoE interface can be an FE or GE interface.

- PD

A PD is a device accepting power from the PSE. There are standard PDs and nonstandard PDs. A standard PD refers to the one that complies with IEEE 802.3af. The PD that is being powered by the PSE can be connected to other power supply units for redundancy backup.

Protocol Specification

The protocol specification related to PoE is IEEE 802.3af.

PoE Configuration Task List

Complete these tasks to configure PoE:

Task	Remarks
Configuring the PoE Interface	Required
Configuring PoE Power Management	Optional
Configuring the PoE Monitoring Function	Optional
Upgrading PSE Processing Software Online	Optional
Configuring a PD Disconnection Detection Mode	Optional
Enabling the PSE to Detect Nonstandard PDs	Optional

Caution

- When the PoE power or PSE fails, you cannot configure PoE.
- Turning off of the PoE power during the startup of the device might result in the failure to restore the PoE configuration.

Configuring the PoE Interface

You can configure a PoE interface in either of the following two ways:

- Adopting the command line.
- Configuring a PoE configuration file and applying the file to the specified PoE interface(s).

Usually, you can adopt the command line to configure a single PoE interface, and adopt a PoE configuration file to configure multiple PoE interfaces at the same time.

Caution

You can adopt either mode to configure, modify, or delete a PoE configuration parameter under the same PoE interface.

The PSE supplies power for a PoE interface in the following two modes:

- Signal cables modes: For a device with only signal cables, power is supplied over signal cables.
- Spare cables modes: For a device with spare cables and signal cables, power can be supplied over spare cables or signal cables.

Note

Switch 4800G only support for signal mode.

Configuring a PoE Interface through the Command Line

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter PoE interface view	**interface** *interface-type interface-number*	—
Enable PoE	**poe enable**	Required Disabled by default.
Configure the maximum power for the PoE interface	**poe max-power** *max-power*	Optional 15,400 milliwatts by default.
Configure the PoE mode for the PoE interface	**poe mode signal**	Optional **signal** (power over signal cables) by default.
Configure a description for the PD connected to the PoE interface	**poe pd-description** *string*	Optional By default, no description for the PD connected to the PoE interface is available.

Configuring PoE Interfaces Through a PoE Configuration File

A PoE configuration file is used to configure at the same time multiple PoE interfaces with the same attributes to simplify operations. This configuration method is a supplement to the command line configuration.

Commands in a PoE configuration file are called configurations.

Follow these steps to configure PoE interfaces through a PoE configuration file:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a PoE configuration file and enter PoE configuration file view	**poe-profile** *profile-name* [*index*]	Required

To do…		Use the command…	Remarks
Enable PoE for the PoE interface		**poe enable**	Required Disabled by default.
Configure the maximum power for the PoE interface		**poe max-power** *max-power*	Optional 15,400 milliwatts by default.
Configure the PoE mode for the PoE interface		**poe mode signal**	Optional **signal** (power over signal cables) by default.
Return to system view		**quit**	—
Apply the PoE configuration file to the PoE interface (s)	Apply the PoE configuration file to one or more PoE interfaces	**apply poe-profile** { **index** *index* | **name** *profile-name* } **interface** *interface-range*	Use either approach
	Apply the PoE configuration file to the current PoE interface in PoE interface view	**interface** *interface-type interface-number*	
		apply poe-profile { **index** *index* | **name** *profile-name* }	

Caution

- After a PoE configuration file is applied to a PoE interface, other PoE configuration files can not take effect on this PoE interface.
- If a PoE configuration file is already applied to a PoE interface, you must execute the **undo apply poe-profile** command to remove the application to the interface before deleting or modifying the PoE configuration file.
- If you have configured a PoE interface through the command line, you cannot configure it through a PoE configuration file again. If you want to reconfigure the interface through a PoE configuration file, you must first remove the command line configuration on the PoE interface.
- You must use the same mode (command line or PoE configuration file) to configure the **poe max-power** *max-power* and **poe priority** { **critical** | **high** | **low** } commands.

Configuring PoE Power Management

Configuring PD Power Management

The power priority of a PD depends on the priority of the PoE interface. The priority levels of PoE interfaces include critical, high and low in descending order. Power supply to a PD is subject to PD power management policies.

All PSEs implement the same PD power management policies. When the PSE supplies power to a PD,

- By default, no power will be supplied to a new PD if the PSE power is overloaded.
- Under the control of a priority policy, the PD with a lower priority is first powered off to guarantee the power supply to the new PD with a higher priority when the PSE power is overloaded.

Note

- 19 watts guard band is reserved for each PoE interface on the device to prevent a PD from being powered off because of sudden increase of the power of the PD. When the remaining power of the interface is lower than 19 watts and no priority is configured for a PoE interface, the PSE does not supply power to the new PD; when the remaining power of the interface is lower than 19 watts, but priority is configured for a PoE interface, the interface with a higher priority can preempt the power of the interface with a lower priority to ensure the normal working of the higher priority interface.
- If the sudden increase of the power of the PD results in PSE power overload, power supply to the PD on the PoE interface with a lower priority will be stopped.

If the guaranteed remaining PSE power (maximum PSE power – power allocated to the critical PoE interface, regardless of whether PoE is enabled for the PoE interface) is lower than the maximum power of the PoE interface, you will fail to set the priority of the PoE interface to **critical**. Otherwise, you can succeed in setting the priority to **critical**, and this PoE interface will preempt the power of other PoE interfaces with a lower priority level. In the latter case, the PoE interfaces whose power is preempted will be powered off, but their configurations will remain unchanged. When you change the priority of a PoE interface from critical to a lower level, the PDs connecting to other PoE interfaces will have an opportunity of being powered.

Configuration prerequisites

Enable PoE for PoE interfaces.

Configuration procedure

Follow these steps to configure PD power management:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Configure the power priority for a PoE interface	Configure the power priority for the PoE interface in PoE interface view	**interface** *interface-type interface-number*	Use either command. By default, the power priority of a PoE interface is **low**.
		poe priority { **critical** | **high** | **low** }	
	Configure the power priority for the PoE interface in PoE configuration file view	**poe-profile** *profile-name* [*index*]	
		poe priority { **critical** | **high** | **low** }	

To do…	Use the command…	Remarks
Configure a PD power management priority policy	**poe pd-policy priority**	Optional By default, no PD power management priority policy is configured.

Configuring the PoE Monitoring Function

The PoE monitoring function involves monitoring of PoE power, PSE and PD.

- Monitoring PoE power means monitoring the voltage of the PoE power.
- When the current power utilization of the PSE is above or below the alarm threshold for the first time, the system will send a Trap message.
- When the PSE starts or stops supplying power to a PD, the system will send a Trap message, too.

Configuring a Power Alarm Threshold for the PSE

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a power alarm threshold for the PSE	**poe utilization-threshold** *utilization-threshold-value* **pse** *pse-id*	Optional 80% by default.

Upgrading PSE Processing Software Online

You can upgrade the PSE processing software online in either of the following two modes:

- refresh mode

This mode enables you to update the PSE processing software without deleting it.

- full mode

This mode deletes the PSE processing software and reloads it. When the PSE processing software is damaged (in this case, you can execute none of PoE commands successfully), you can upgrade the PSE software processing software in full mode to restore the PSE function.

Online PSE processing software upgrade may be unexpectedly interrupted (for example, an error results in device reboot). If you fail to upgrade the PSE processing software in full mode after reboot, you can power off the device and restart it before upgrading it again. After upgrade, restart the device manually to make the original PoE configurations take effect.

Follow these steps to upgrade the PSE processing software online:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Upgrade the PSE processing software online	**poe update** { **full** | **refresh** } *filename* **pse** *pse-id*	Optional

Configuring a PD Disconnection Detection Mode

To detect the PD connection with PSE, PoE provides two detection modes: AC detection and DC detection. The AC detection mode is energy saving relative to the DC detection mode.

Follow these steps to configure a PD disconnection detection mode:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure a PD disconnection detection mode	**poe disconnect** { **ac** | **dc** }	Optional The default PD disconnection detection mode varies with devices.

Caution

If you adjust the PD disconnection detection mode when the device is running, the connected PDs will be powered off. Therefore, be cautious to do so.

Enabling the PSE to Detect Nonstandard PDs

There are standard PDs and nonstandard PDs. Usually, the PSE can detect only standard PDs and supply power to them. The PSE can detect nonstandard PDs and supply power to them only after the PSE is enabled to detect nonstandard PDs.

Follow these steps to enable the PSE to detect nonstandard PDs:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the PSE to detect nonstandard PDs	**poe legacy enable pse** *pse-id*	Optional Disabled by default.

Displaying and Maintaining PoE

To do…	Use the command…	Remarks
Display the mapping between ID, module, and member ID of all PSEs.	**display poe device**	Available in any view
Display the power state and information of the specified PoE interface	**display poe interface** [*interface-type interface-number*]	
Display the power information of a PoE interface(s)	**display poe interface power** [*interface-type interface-number*]	
Display the information of PSE	**display poe pse** [*pse-id*]	
Display the power state and information of all PoE interfaces connected with the PSE	**display poe pse** *pse-id* **interface**	
Display the power of all PoE interfaces connected with the PSE	**display poe pse** *pse-id* **interface power**	
Display all information of the configurations and applications of the PoE configuration file	**display poe-profile** [**index** *index* | **name** *profile-name*]	
Display all information of the configurations and applications of the PoE configuration file applied to the specified PoE interface	**display poe-profile interface** *interface-type interface-number*	

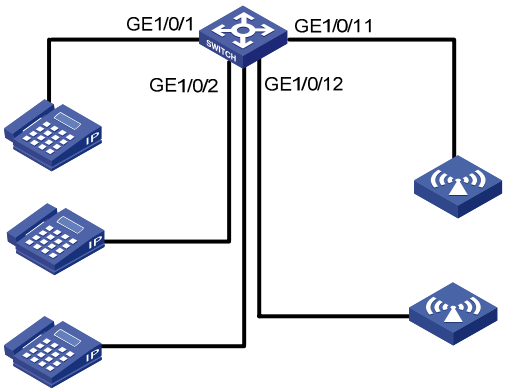
PoE Configuration Example

Network requirements

The device provides power supply for PDs through PoE interfaces.

- GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are connected to IP telephones.
- GigabitEthernet 1/0/11 and GigabitEthernet 1/0/12 are connected to access point (AP) devices.
- The power priority of GigabitEthernet 1/0/2 is critical.
- The power of the AP device connected to GigabitEthernet 1/0/11 does not exceed 9,000 milliwatts.

Figure 1-1 Network diagram for PoE



Configuration procedure

Enable PoE on GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, GigabitEthernet 1/0/11, and GigabitEthernet 1/0/12.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] poe enable
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface GigabitEthernet 1/0/2
[Sysname-GigabitEthernet1/0/2] poe enable
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] interface GigabitEthernet 1/0/11
[Sysname-GigabitEthernet1/0/11] poe enable
[Sysname-GigabitEthernet1/0/11] quit
[Sysname] interface GigabitEthernet 1/0/12
[Sysname-GigabitEthernet1/0/12] poe enable
[Sysname-GigabitEthernet1/0/12] quit
```

Set the power priority level of GigabitEthernet 1/0/2 to **critical**.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/2
[Sysname-GigabitEthernet1/0/2] poe priority critical
[Sysname-GigabitEthernet1/0/2] quit
# Set the maximum power of GigabitEthernet 1/0/11 to 9,000 milliwatts.
[Sysname] interface GigabitEthernet 1/0/11
[Sysname-GigabitEthernet1/0/11] poe max-power 9000
[Sysname-GigabitEthernet1/0/11] quit
```

After the configuration takes effect, the IP phone and AR device are powered and can work normally.

Troubleshooting PoE

Symptom 1: Setting of the priority of a PoE interface to **critical** fails.

Analysis:

- The guaranteed remaining power of the PSE is lower than the maximum power of the PoE interface.

- The priority of the PoE interface is already set.

Solution:

- In the first case, you can solve the problem by increasing the maximum PSE power, or by reducing the maximum power of the PoE interface when the guaranteed remaining power of the PSE cannot be modified.
- In the second case, you should first remove the priority already configured.

Symptom 2: Applying a PoE configuration file to a PoE interface fails.

Analysis:

- Some configurations in the PoE configuration file are already configured.
- Some configurations in the PoE configuration file do not meet the configuration requirements of the PoE interface.
- Another PoE configuration file is already applied to the PoE interface.

Solution:

- In the first case, you can solve the problem by removing the original configurations of those configurations.
- In the second case, you need to modify some configurations in the PoE configuration file.
- In the third case, you need to remove the application of the undesired PoE configuration file to the PoE interface.

Symptom 3: Provided that parameters are valid, configuring an AC input under-voltage threshold fails.

Analysis:

The AC input under-voltage threshold is greater than or equal to the AC input over-voltage threshold.

Solution:

You can drop the AC input under-voltage threshold below the AC input over-voltage threshold.

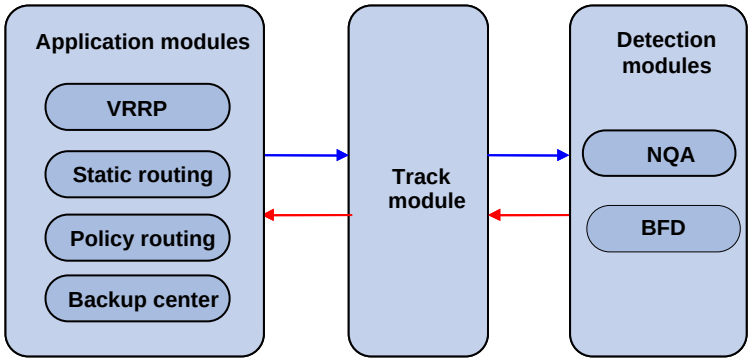
Table of Contents

1 Track Configuration

When configuring Track, go to these sections for information you are interested in:

Track Overview

Figure 1-1 Collaboration through the Track module



The Track module is used to implement collaboration between different modules.

The collaboration here involves three parts: the application modules, the Track module, and the detection modules. These modules collaborate with one another through collaboration objects. That is, the detection modules trigger the application modules to perform certain operations through the Track module. More specifically, the detection modules probe the link status, network performance and so on, and inform the application modules of the detection result through the Track module. After the application modules are aware of the changes of network status, they deal with the changes accordingly to avoid communication interruption and network performance degradation.

The Track module works between the application modules and the detection modules and is mainly used to obscure the difference of various detection modules to provide a unified interface for application modules.

Collaboration Between the Track Module and the Detection Modules

You can establish the collaboration between the Track module and the detection modules through configuration. A detection module probes the link status and informs the Track module of the probe result. The Track module then changes the status of the Track object accordingly:

- If the probe succeeds, the status of the corresponding Track object is **Positive**;
- If the probe fails, the status of the corresponding Track object is **Negative**.

At present, the detection modules that can collaborate with the Track module include the Network Quality Analyzer (NQA) and the Bidirectional Forwarding Detection (BFD) module. Refer to *NQA Configuration* in the *System Volume* for details of NQA and *BFD Configuration* in the *IP Routing Volume* for details of BFD.

Collaboration Between the Track Module and the Application Modules

You can establish the collaboration between the Track module and the application modules through configuration. If the status of the Track object changes, the Track module tells the application modules to deal with the change accordingly.

At present, the application modules that can collaborate with the Track module include:

- VRRP
- Static routing

Track Configuration Task List

To implement the collaboration function, you need to establish collaboration between the Track module and the detection modules, and between the Track module and the application modules.

Complete these tasks to configure Track module:

Task		Remarks
Configuring Collaboration Between the Track Module and the Detection Modules	Configuring Track-NQA Collaboration	Use either approach
	Configuring Track-BFD Collaboration	
Configuring Collaboration Between the Track Module and the Application Modules	Configuring Track-VRRP Collaboration	Use at least one of the two approaches
	Configuring Track-Static Routing Collaboration	

Configuring Collaboration Between the Track Module and the Detection Modules

Configuring Track-NQA Collaboration

Through the following configuration, you can establish the collaboration between the Track module and the NQA, which probes the link status and informs the Track module of the probe result.

Follow these steps to configure Track-NQA collaboration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a Track object and associate it with the specified Reaction entry of the NQA test group	**track** *track-entry-number* **nqa entry** *admin-name operation-tag* **reaction** *item-num*	Required No Track object is created by default.

Caution

When you configure a Track object, the specified NQA test group and Reaction entry can be nonexistent. In this case, the status of the configured Track object is **Invalid**.

Configuring Track-BFD Collaboration

Through the following configuration, you can establish the collaboration between the Track module and BFD, which probes the link status and informs the Track module of the probe result.

Follow these steps to configure Track-BFD collaboration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create a Track object and associate it with the BFD session	**track** *track-entry-number* **bfd echo interface** *interface-type interface-number* **remote ip** *remote-ip* **local ip** *local-ip*	Required No Track object is created by default.

Configuring Collaboration Between the Track Module and the Application Modules

Configuring Track-VRRP Collaboration

Through the Track-VRRP collaboration, you can:

- Monitor the upper link. If there is a fault on the upper link of the master of a VRRP group, hosts in the LAN cannot access the external network through the master. In this case, the status of the monitored Track object changes to Negative, and the priority of the master thus decreases by a specified value, allowing a higher priority router in the VRRP group to become the master to maintain proper communication between the hosts in the LAN and the external network.
- Monitor the master on a backup. If there is a fault on the master, the backup working in the switchover mode will switch to the master immediately to maintain normal communication.

Configuration prerequisites

Before configuring VRRP to monitor a Track object, you need to create a VRRP group on an interface and configure the virtual IP address of the VRRP group.

Configuration procedure

Follow these steps to configure Track-VRRP collaboration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—

To do…	Use the command…	Remarks
Create a VRRP group and configure its virtual IP address	**vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*	Required No VRRP group is created by default.
Specify a Track object to be monitored by VRRP	**vrrp vrid** *virtual-router-id* **track** *track-entry-number* [**reduced** *priority-reduced* | **switchover**]	Required No Track object is specified for VRRP by default.

Note

- Do not perform Track object monitoring on the IP address owner.
- When the status of the monitored Track object turns from Negative to Positive, the corresponding router restores its priority automatically.
- The monitored Track object can be nonexistent, so that you can first specify the Track object to be monitored using the **vrrp vrid track** command, and then create the Track object using the **track** command.
- Refer to *VRRP Configuration* in the *System Volume* for details of VRRP.

Configuring Track-Static Routing Collaboration

You can check the validity of a static route in real time by establishing collaboration between Track and static routing.

If you specify the next hop but not the egress interface when configuring a static route, you can associate the static route with a Track object and thus check the validity of the static route according to the status of the Track object.

- If the status of the Track object is **Positive**, then the next hop of the static route is reachable, and the configured static route is valid.
- If the status of the Track object is **Negative**, then the next hop of the static route is unreachable, and the configured static route is invalid.

Follow these steps to configure the Track-Static Routing collaboration:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the Track-Static Routing collaboration, so as to check the reachability of the next hop of the static route	**ip route-static** *dest-address* { *mask* | *mask-length* } { *next-hop-address* | **vpn-instance** *d-vpn-instance-name next-hop-address* } **track** *track-entry-number* [**preference** *preference-value*] [**tag** *tag-value*] [**description** *description-text*]	Use either command. Not configured by default.
	ip route-static vpn-instance s-vpn-instance-name&<1-6> dest-address { mask | mask-length } { next-hop-address **track** track-entry-number [**public**] | **vpn-instance** d-vpn-instance-name next-hop-address **track** track-entry-number } [**preference** preference-value] [**tag** tag-value] [**description** description-text]	

Note

- For the configuration of Track-Static Routing collaboration, the specified static route can be an existent or nonexistent one. For an existent static route, the static route and the specified Track object are associated directly; for a nonexistent static route, the system creates the static route and then associates it with the specified Track object.
- The Track object to be associated with the static route can be a nonexistent one. After you use the **track** command to create the Track object, the association takes effect.
- If the Track module detects the next hop reachability of the static route in a private network through NQA, the VPN instance name of the next hop of the static route must be consistent with that configured for the NQA test group. Otherwise, the reachability detection cannot function properly.
- If a static route needs route recursion, the associated Track object must monitor the next hop of the recursive route instead of that of the static route; otherwise, a valid route may be considered invalid.
- For details of static route configuration, refer to *Static Routing Configuration* in the *IP Routing Volume*.

Displaying and Maintaining Track Object(s)

To do…	Use the command…	Remarks
Display information about the specified Track object or all Track objects	**display track** { *track-entry-number* | **all** }	Available in any view

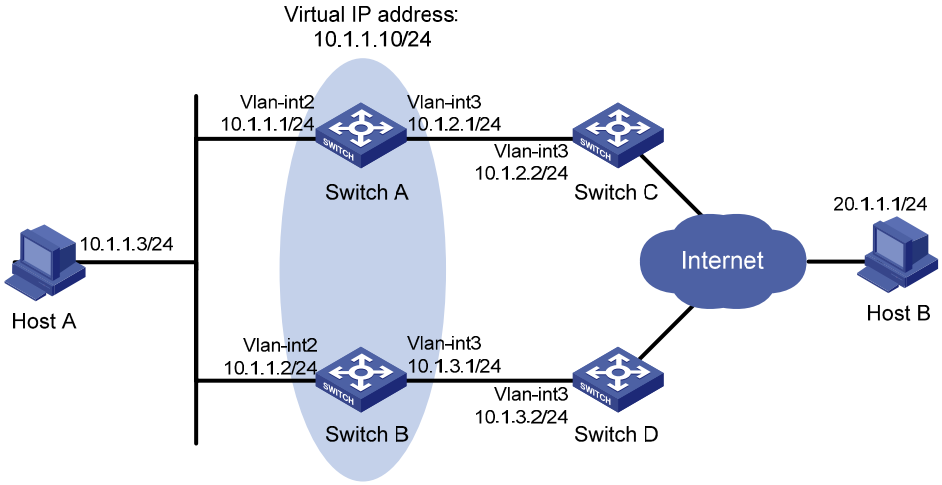
Track Configuration Examples

VRRP-Track-NQA Collaboration Configuration Example

Network requirements

- Host A needs to access Host B on the Internet. The default gateway of Host A is 10.1.1.10/24.
- Switch A and Switch B belong to VRRP group 1, whose virtual IP address is 10.1.1.10.
- When Switch A works normally, packets from Host A to Host B are forwarded through Switch A. When VRRP finds that there is a fault on the upper link of Switch A through NQA, packets from Host A to Host B are forwarded through Switch B.

Figure 1-2 Network diagram for VRRP-Track-NQA collaboration configuration



Configuration procedure

1) Configure the IP address of each interface as shown in Figure 1-2.
2) Configure an NQA test group on Switch A.

```
<SwitchA> system-view
```

Create an NQA test group with the administrator name **admin** and the operation tag **test**.

```
[SwitchA] nqa entry admin test
```

Configure the test type as ICMP-echo.

```
[SwitchA-nqa-admin-test] type icmp-echo
```

Configure the destination address as 10.1.2.2.

```
[SwitchA-nqa-admin-test-icmp-echo] destination ip 10.1.2.2
```

Set the test frequency to 100 ms.

```
[SwitchA-nqa-admin-test-icmp-echo] frequency 100
```

Configure Reaction entry 1, specifying that five consecutive probe failures trigger the Track-NQA collaboration.

```
[SwitchA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail threshold-type
consecutive 5 action-type trigger-only
[SwitchA-nqa-admin-test-icmp-echo] quit
```

Start NQA probes.

```
[SwitchA] nqa schedule admin test start-time now lifetime forever
```

3) Configure a Track object on Switch A.

Configure Track object 1, and associate it with Reaction entry 1 of the NQA test group (with the administrator **admin**, and the operation tag **test)**.

```
[SwitchA] track 1 nqa entry admin test reaction 1
```

4) Configure VRRP on Switch A.

Create VRRP group 1, and configure the virtual IP address 10.1.1.10 for the group.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Set the authentication mode of VRRP group 1 to **simple**, and the authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Configure the master to send VRRP packets at an interval of five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

Configure Switch A to work in preemptive mode, and set the preemption delay to five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

Configure to monitor Track object 1 and specify the priority decrement to 30.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track 1 reduced 30
```

5) Configure VRRP on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 2
```

Create VRRP group 1, and configure the virtual IP address 10.1.1.10 for the group.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Set the authentication mode of VRRP group 1 to **simple**, and the authentication key to **hello**.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Configure the master to send VRRP packets at an interval of five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

Configure Switch B to work in preemptive mode, and set the preemption delay to five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

6) Verify the configuration

After configuration, ping Host B on Host A, and you can see that Host B is reachable. Use the **display vrrp** command to view the configuration result.

Display detailed information about VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Track Object    : 1                   Pri Reduced     : 0
 Virtual IP      : 10.1.1.10
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 10.1.1.1
```

Display detailed information about VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
```

```
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Virtual IP      : 10.1.1.10
 Master IP       : 10.1.1.1
```

The above output information indicates that in VRRP group 1, Switch A is the master and Switch B is a backup. Packets from Host A to Host B are forwarded through Switch A.

When there is a fault on the link between Switch A and Switch C, you can still successfully ping Host B on Host A. Use the **display vrrp** command to view information about VRRP group 1.

Display detailed information about VRRP group 1 on Switch A when there is a fault on the link between Switch A and Switch C.

```
[SwitchA-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Backup
 Config Pri      : 110                 Run Pri         : 80
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Track Object    : 1                   Pri Reduced     : 30
 Virtual IP      : 10.1.1.10
 Master IP       : 10.1.1.2
```

Display detailed information about VRRP group 1 on Switch B when there is a fault on the link between Switch A and Switch C.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Master
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Virtual IP      : 10.1.1.10
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 10.1.1.2
```

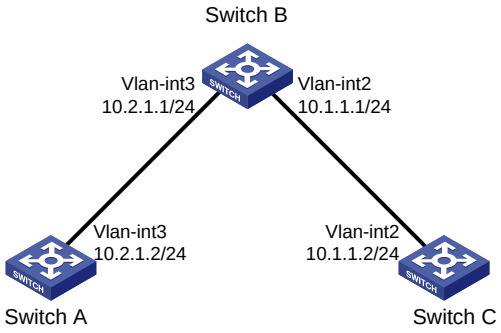
The output information indicates that when there is a fault on the link between Switch A and Switch C, the priority of Switch A decreases to 80. Switch A becomes the backup, and Switch B becomes the master. Packets from Host A to Host B are forwarded through Switch B.

Static Routing-Track-NQA Collaboration Configuration Example

Network requirements

- The next hop of the static route from Switch A to Switch C is Switch B.
- Configure Static Routing-Track-NQA collaboration on Switch A to implement real-time monitoring of the validity of the static route to Switch C.

Figure 1-3 Network diagram for Static Routing-Track-NQA collaboration configuration



Configuration procedure

1) Configure the IP address of each interface as shown in Figure 1-3.
2) Configure a static route on Switch A and associate it with the Track object.

Configure the address of the next hop of the static route to Switch C as 10.2.1.1, and configure the static route to associate with Track object 1.

```
<SwitchA> system-view
[SwitchA] ip route-static 10.1.1.2 24 10.2.1.1 track 1
```

3) Configure an NQA test group on Switch A.

Create an NQA test group with the administrator **admin** and the operation tag **test**.

```
[SwitchA] nqa entry admin test
```

Configure the test type as ICMP-echo.

```
[SwitchA-nqa-admin-test] type icmp-echo
```

Configure the destination address as 10.2.1.1

```
[SwitchA-nqa-admin-test-icmp-echo] destination ip 10.2.1.1
```

Configure the test frequency as 100 ms.

```
[SwitchA-nqa-admin-test-icmp-echo] frequency 100
```

Configure Reaction entry 1, specifying that five consecutive probe failures trigger the Static Routing-Track-NQA collaboration.

```
[SwitchA-nqa-admin-test-icmp-echo] reaction 1 checked-element probe-fail threshold-type
consecutive 5 action-type trigger-only
[SwitchA-nqa-admin-test-icmp-echo] quit
```

Start NQA probes.

```
[SwitchA] nqa schedule admin test start-time now lifetime forever
```

4) Configure a Track object on Switch A.

Configure Track object 1, and associate it with Reaction entry 1 of the NQA test group (with the administrator **admin**, and the operation tag **test)**.

```
[SwitchA] track 1 nqa entry admin test reaction 1
```

5) Verify the configuration

Display information of the Track object on Switch A.

```
[SwitchA] display track all
Track ID: 1
  Status: Positive
  Reference object:
    NQA entry: admin test
    Reaction: 1
```

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 5        Routes : 5

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.1.1.0/24         Static 60   0            10.2.1.1        Vlan3
10.2.1.0/24         Direct 0    0            10.2.1.2        Vlan3
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output information above indicates the NQA test result, that is, the next hop 10.2.1.1 is reachable (the status of the Track object is **Positive**), and the configured static route is valid.

Remove the IP address of interface VLAN-interface 3 on Switch B.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] undo ip address
```

Display information of the Track object on Switch A.

```
[SwitchA] display track all
Track ID: 1
  Status: Negative
  Reference object:
    NQA entry: admin test
    Reaction: 1
```

Display the routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 4        Routes : 4

Destination/Mask    Proto  Pre  Cost         NextHop         Interface

10.2.1.0/24         Direct 0    0            10.2.1.2        Vlan3
10.2.1.2/32         Direct 0    0            127.0.0.1       InLoop0
```

```
127.0.0.0/8         Direct 0    0            127.0.0.1       InLoop0
127.0.0.1/32        Direct 0    0            127.0.0.1       InLoop0
```

The output information above indicates the NQA test result, that is, the next hop 10.2.1.1 is unreachable (the status of the Track object is **Negative**), and the configured static route is invalid.

Table of Contents

1 NQA Configuration

When configuring NQA, go to these sections for information you are interested in:

NQA Overview

Introduction to NQA

Network Quality Analyzer (NQA) analyzes network performance, services and service quality through sending test packets, and provides you with network performance and service quality parameters such as jitter, TCP connection delay, FTP connection delay and file transfer rate.

With the NQA test results, you can:

1) Know network performance in time and then take corresponding measures.
2) Diagnose and locate network faults.

Features of NQA

Supporting multiple test types

Ping can use only the Internet Control Message Protocol (ICMP) to test the reachability of the destination host and the roundtrip time of a packet to the destination. As an enhancement to the Ping tool, NQA provides multiple test types and more functions.

At present, NQA supports ten test types: ICMP echo, DHCP, FTP, HTTP, UDP jitter, SNMP, TCP, UDP echo, voice and DLSw.

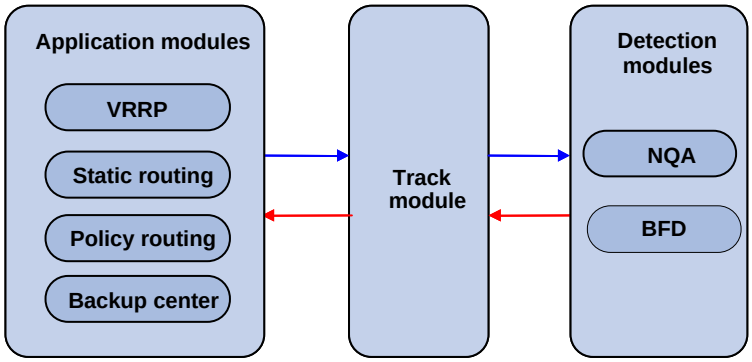
In an NQA test, the client sends different types of test packets to the peer to detect the availability and the response time of the peer, helping you know protocol availability and network performance based on the test results.

Supporting the collaboration function

Collaboration is implemented by establishing collaboration objects to monitor the detection results of the current test group. If the number of consecutive probe failures reaches a certain limit, NQA's

collaboration with other modules is triggered. The implementation of collaboration is shown in Figure 1-1.

Figure 1-1 Implementation of collaboration



The collaboration here involves three parts: the application modules, the Track module, and the detection modules.

- The detection modules monitor the link status, network performance and so on, and inform the Track module of the detection result.
- Upon receiving the detection result, the Track module changes the status of the Track object accordingly and informs the application modules. The Track module works between the application modules and the detection modules and is mainly used to obscure the difference of various detection modules to provide a unified interface for application modules.
- The application modules then deal with the changes accordingly based on the status of the Track object, and thus collaboration is implemented.

Take static routing as an example. You have configured a static route with the next hop 192.168.0.88. If 192.168.0.88 is reachable, the static route is valid; if 192.168.0.88 is unreachable, the static route is invalid. With the collaboration between NQA, Track module and application modules, real time monitoring of reachability of the static route can be implemented:

1) Monitor reachability of the destination 192.168.0.88 through NQA.
2) If 192.168.0.88 is detected to be unreachable, NQA will inform the static routing module through Track module.
3) The static routing module then can know that the static route is invalid.

Note

- At present, policy routing and backup center are not supported.
- For the detailed description of the Track module, see *Track Configuration* in the *System Volume*.

Supporting delivery of traps

You can set whether to send traps to the network management server when an NQA test is performed. When a probe fails or a test is completed, the network management server can be notified, and the network administrator can know the network running status and performance in time through the traps sent.

Basic Concepts of NQA

Test group

Before performing an NQA test, you need to create an NQA test group, and configure NQA test parameters such as test type, destination address and destination port.

Each test group has an administrator name and operation tag, which can uniquely define a test group.

Test and probe

After an NQA test is started, one test is performed at a regular interval and you can set the interval as needed.

One NQA test involves multiple consecutive probes and you can set the number of the probes.

Note

Only one probe can be made in one voice test.

In different test types, probe has different meanings:

- For a TCP or DLSw test, one probe means one connection;
- For a UDP jitter or a voice test, multiple packets are sent successively in one probe, and the number of packets sent in one probe depends on the configuration of the **probe packet-number** command;
- For an FTP, HTTP or DHCP test, one probe means to carry out a corresponding function;
- For an ICMP echo or UDP echo test, one packet is sent in one probe;
- For an SNMP test, three packets are sent in one probe.

NQA client and server

NQA client is the device initiating an NQA test and the NQA test group is created on the NQA client.

NQA server processes the test packets sent from the NQA client, as shown in Figure 1-2. The NQA server makes a response to the request originated by the NQA client by listening to the specified destination address and port number.

Figure 1-2 Relationship between the NQA client and NQA server

In most NQA tests, you only need to configure the NQA client; while in TCP, UDP echo, UDP jitter, and voice tests, you must configure the NQA server.

You can create multiple TCP or UDP listening services on the NQA server, with each listening service corresponding to a specified destination address and port number. The IP address and port number specified for a listening service on the server must be consistent with those on the client and must be different from those of an existing listening service.

NQA Test Operation

An NQA test operation is as follows:

1) The NQA client constructs packets with the specified type, and sends them to the peer device;
2) Upon receiving the packet, the peer device replies with a response with a timestamp.
3) The NQA client computes the packet loss rate and RTT based on whether it has received the response and the timestamp in the response.

NQA Configuration Task List

For TCP, UDP jitter, UDP echo or voice tests, you need to configure the NQA server on the peer device.

Follow these steps to enable the NQA server:

Task	Remarks
Configuring the NQA Server	Required for TCP, UDP echo, UDP jitter and voice tests

To perform an NQA test successfully, make the following configurations on the NQA client:

1) Enable the NQA client;
2) Create a test group and configure test parameters according to the test type. The test parameters may vary with test types;
3) Start the NQA test;

After the test, you can view test results using the **display** or **debug** commands.

Complete these tasks to configure NQA client:

Task		Remarks
Enabling the NQA Client		Required
Creating an NQA Test Group		Required
Configuring an NQA Test Group	Configuring an ICMP Echo Test	Required Use any of the approaches
	Configuring a DHCP Test	
	Configuring an FTP Test	
	Configuring an HTTP Test	
	Configuring a UDP Jitter Test	
	Configuring an SNMP Test	
	Configuring a TCP Test	
	Configuring a UDP Echo Test	
	Configuring a Voice Test	
	Configuring a DLSw Test	
Configuring the Collaboration Function		Optional
Configuring Trap Delivery		Optional
Configuring the NQA Statistics Function		Optional

Task	Remarks
Configuring Optional Parameters Common to an NQA Test Group	Optional
Scheduling an NQA Test Group	Required

Configuring the NQA Server

Before performing TCP, UDP echo, UDP jitter or voice tests, you need to configure the NQA server on the peer device. The NQA server makes a response to the request originated by the NQA client by listening to the specified destination address and port number.

Follow these steps to configure the NQA server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the NQA server	**nqa server enable**	Required Disabled by default.
Configure the UDP or TCP listening function on the NQA server	**nqa server** { **tcp-connect** | **udp-echo** } *ip-address port-number*	Required The IP address and port number must be consistent with those configured on the NQA client and must be different from those of an existing listening service.

Enabling the NQA Client

Configurations on the NQA client take effect only when the NQA client is enabled.

Follow these steps to enable the NQA client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the NQA client	**nqa agent enable**	Optional Enabled by default.

Creating an NQA Test Group

One test corresponds to one test group. You can configure test types after you create a test group and enter the test group view.

Follow theses steps to create an NQA test group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Create an NQA test group and enter the NQA test group view	**nqa entry** *admin-name operation-tag*	Required

Note

If you execute the **nqa entry** command to enter the test group view with test type configured, you will enter the test type view of the test group directly.

Configuring an NQA Test Group

Configuring an ICMP Echo Test

An ICMP echo test is used to test reachability of the destination host according to the ICMP echo reply or timeout information. An ICMP echo test has the same function with the **ping** command but has more abundant output information. You can use the ICMP echo test to locate connectivity problems in a network.

Follow these steps to configure an ICMP echo test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as ICMP echo and enter test type view	**type icmp-echo**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation.
Configure the size of probe packets sent	**data-size** *size*	Optional 100 bytes by default.
Configure the filler string of a probe packet sent	**data-fill** *string*	Optional By default, the filler string of a probe packet is the hexadecimal number 00010203040506070809.
Specify a VPN instance	**vpn-instance** *instance*	Optional Not specified by default.
Specify the IP address of an interface as the source IP address of an ICMP echo request	**source interface** *interface-type interface-number*	Optional By default, no interface address is specified as the source IP address of ICMP probe requests. If you use the **source ip** command to configure the source IP address of ICMP echo probe requests, the **source interface** command is invalid. The interface specified by this command must be up. Otherwise, the probe will fail.

To do…	Use the command…	Remarks
Configure the source IP address of a probe request	**source ip** *ip-address*	Optional By default, no source IP address is specified. If no source IP address is specified, but the source interface is specified, the IP address of the source interface is taken as the source IP address of ICMP probe requests. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the probe will fail.
Configure the next hop IP address for an ICMP echo request	**next-hop** *ip-address*	Optional By default, no next hop IP address is configured.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Configuring a DHCP Test

A DHCP test is mainly used to test the existence of a DHCP server on the network as well as the time necessary for the DHCP server to respond to a client request and assign an IP address to the client.

Configuration prerequisites

Before performing a DHCP test, you need to configure the DHCP server. If the NQA (DHCP client) and the DHCP server are not in the same network segment, you need to configure a DHCP relay. For the configuration of DHCP server and DHCP relay, see *DHCP Configuration* in the *IP Services Volume*.

Configuring a DHCP test

Follow these steps to configure a DHCP test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as DHCP and enter test type view	**type dhcp**	Required
Specify an interface for a DHCP test	**operation interface** *interface-type interface-number*	Required By default, no interface is specified to perform a DHCP test. The interface specified by the **source interface** command must be up; otherwise, the test fails.

To do…	Use the command…	Remarks
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Note

- As DHCP test is a process to simulate address allocation in DHCP, the IP address of the interface performing the DHCP test will not be changed.
- After the DHCP test is completed, the NQA client will send a DHCP-RELEASE packet to release the obtained IP address.

Configuring an FTP Test

An FTP test is mainly used to test the connection between the NQA client and a specified FTP server and the time necessary for the FTP client to transfer a file to or download a file from the FTP server.

Configuration prerequisites

Before an FTP test, you need to perform some configurations on the FTP server. For example, you need to configure the username and password used to log onto the FTP server. For the FTP server configuration, see *File System Management Configuration* in the *System Volume*.

Configuring an FTP test

Follow these steps to configure an FTP test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as FTP and enter test type view	**type ftp**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination IP address for a test operation is the IP address of the FTP server.
Configure the source IP address of a probe request	**source ip** *ip-address*	Required By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.

To do…	Use the command…	Remarks
Configure the operation type	**operation** { **get** | **put** }	Optional By default, the operation type for the FTP is **get**, that is, obtaining files from the FTP server.
Configure a login username	**username** *name*	Required By default, no login username is configured.
Configure a login password	**password** *password*	Required By default, no login password is configured.
Specify a file to be transferred between the FTP server and the FTP client	**filename** *file-name*	Required By default, no file is specified.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Note

- When you execute the **put** command, a file *file-name* with fixed size and content is created on the FTP server; when you execute the **get** command, the device does not save the files obtained from the FTP server.
- When you execute the **get** command, the FTP test cannot succeed if a file named *file-name* does not exist on the FTP server.
- When you execute the **get** command, please use a file with a smaller size as a big file may result in test failure because of timeout, or may affect other services because of occupying too much network bandwidth.

Configuring an HTTP Test

An HTTP test is used to test the connection between the NQA client and a specified HTTP server and the time required to obtain data from the HTTP server, thus detecting the connectivity and performance of the HTTP server.

Configuration prerequisites

Before performing an HTTP test, you need to configure the HTTP server.

Configuring an HTTP test

Follow these steps to configure an HTTP test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—

To do…	Use the command…	Remarks
Configure the test type as HTTP and enter test type view	**type http**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination IP address for a test operation is the IP address of the HTTP server.
Configure the source IP address of a probe request	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure the operation type	**operation** { **get** | **post** }	Optional By default, the operation type for the HTTP is **get**, that is, obtaining data from the HTTP server.
Configure the website that an HTTP test visits	**url** *url*	Required
Configure the HTTP version used in the HTTP test	**http-version v1.0**	Optional By default, HTTP 1.0 is used in an HTTP test.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Note

The TCP port number for the HTTP server must be 80 in an HTTP test. Otherwise, the test will fail.

Configuring a UDP Jitter Test

Note

It is recommended not to perform an NQA UDP jitter test on known ports, namely, ports from 1 to 1023. Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.

Real-time services such as voice and video have high requirements on delay jitters. With the UDP jitter test, uni/bi-directional delay jitters can be obtained to judge whether a network can carry real-time services.

Delay jitter refers to the difference between the interval of receiving two packets consecutively and the interval of sending these two packets. The procedure of a UDP jitter test is as follows:

- The source sends packets at regular intervals to the destination port.
- The destination affixes a time stamp to each packet that it receives and then sends it back to the source.
- Upon receiving the packet, the source calculates the delay jitter, and the network status can be analyzed.

Configuration prerequisites

A UDP jitter test requires cooperation between the NQA server and the NQA client. Before the UDP jitter test, make sure that the UDP listening function is configured on the NQA server. For the configuration of the UDP listening function, see Configuring the NQA Server.

Configuring a UDP jitter test

Follow these steps to configure a UDP jitter test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as UDP jitter and enter test type view	**type udp-jitter**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination IP address must be consistent with that of the existing listening service on the NQA server.
Configure the destination port for a test operation	**destination port** *port-number*	Required By default, no destination port number is configured for a test operation. The destination port must be consistent with that of the existing listening service on the NQA server.
Specify the source port number for a request	**source port** *port-number*	Optional By default, no source port number is specified.
Configure the size of a probe packet sent	**data-size** *size*	Optional 100 bytes by default.
Configure the filler string of a probe packet sent	**data-fill** *string*	Optional By default, the filler string of a probe packet is the hexadecimal number 00010203040506070809.

To do…	Use the command…	Remarks
Configure the number of packets sent in a UDP jitter probe	**probe packet-number** *packet-number*	Optional 10 by default.
Configure the interval for sending packets in a UDP jitter probe	**probe packet-interval** *packet-interval*	Optional 20 milliseconds by default.
Configure the time for waiting for a response in a UDP jitter test	**probe packet-timeout** *packet-timeout*	Optional 3000 milliseconds by default.
Configure the source IP address of a probe request in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Note

The number of probes made in a UDP jitter test depends on the **probe count** command, while the number of probe packets sent in each probe depends on the configuration of the **probe packet-number** command.

Configuring an SNMP Test

An SNMP query test is used to test the time the NQA client takes to send an SNMP query packet to the SNMP agent and then receive a response packet.

Configuration prerequisites

The SNMP agent function must be enabled on the device serving as an SNMP agent before an SNMP test. For the configuration of SNMP agent, see *SNMP Configuration* in the *System Volume*.

Configuring an SNMP test

Follow these steps to configure an SNMP test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as SNMP and enter test type view	**type snmp**	Required

To do…	Use the command…	Remarks
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation.
Specify the source port number for a probe request in a test operation	**source port** *port-number*	Optional By default, no source port number is specified.
Configure the source IP address of a probe request in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Configuring a TCP Test

A TCP test is used to test the TCP connection between the client and the specified port on the NQA server and the setup time for the connection, thus judge the availability and performance of the services provided on the specified port on the server.

Configuration prerequisites

A TCP test requires cooperation between the NQA server and the NQA client. The TCP listening function needs to be configured on the NQA server before the TCP test. For the configuration of the TCP listening function, see Configuring the NQA Server.

Configuring a TCP test

Follow these steps to configure a TCP test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as TCP and enter test type view	**type tcp**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination address must be the IP address of the listening service configured on the NQA server.

To do…	Use the command…	Remarks
Configure the destination port	**destination port** *port-number*	Required By default, no destination port number is configured for a test operation. The destination port number must be consistent with port number of the listening service configured on the NQA server.
Configure the source IP address of a probe request in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Configuring a UDP Echo Test

A UDP echo test is used to test the connectivity and roundtrip time of a UDP echo packet from the client to the specified UDP port on the NQA server.

Configuration prerequisites

A UDP echo test requires cooperation between the NQA server and the NQA client. The UDP listening function needs to be configured on the NQA server before the UDP echo test. For the configuration of the UDP listening function, see Configuring the NQA Server.

Configuring a UDP echo test

Follow these steps to configure a UDP echo test

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as UDP echo and enter test type view	**type udp-echo**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination address must be the IP address of the listening service configured on the NQA server.

To do…	Use the command…	Remarks
Configure the destination port	**destination port** *port-number*	Required By default, no destination port number is configured for a test operation. The destination port number must be the port number of the listening service configured on the NQA server.
Configure the size of probe packets sent	**data-size** *size*	Optional 100 bytes by default.
Configure the filler string of a probe packet sent	**data-fill** *string*	Optional By default, the filler string of a probe packet is the hexadecimal number 00010203040506070809.
Specify a source port number for a probe request in a test operation	**source port** *port-number*	Optional By default, no source port number is specified.
Configure the source IP address of a probe request in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Configuring a Voice Test

Note

It is recommended not to perform an NQA UDP jitter test on known ports, namely, ports from 1 to 1023. Otherwise, the NQA test will fail or the corresponding services of these ports will be unavailable.

A voice test is used to test voice over IP (VoIP) network status, and collect VoIP network parameters so that users can adjust the network according the network status. The procedure of a voice test is as follows:

1) The source (NQA client) sends voice packets of G.711 A-law, G.711 μ-law or G.729 A-law codec type at regular intervals to the destination (NQA server).
2) The destination affixes a time stamp to each packet that it receives and then sends it back to the source.
3) Upon receiving the packets, the source calculates the delay jitter and delay by calculating the difference between the interval for the destination to receive two successive packets and the

interval for the source to send these two successive packets, and thus the network status can be analyzed.

The voice parameter values that indicate VoIP network status can also be calculated in a voice test, including:

- Calculated Planning Impairment Factor (ICPIF): Measures attenuation of voice data in a network, depending on packet loss and delay. A higher value represents a lower network quality.
- Mean Opinion Scores (MOS): Measures quality of a VoIP network. A MOS value can be evaluated by using the ICPIF value, in the range 1 to 15. A higher value represents a higher quality of a VoIP network.

The evaluation of voice quality depends on users' tolerance to voice quality, and this factor should be taken into consideration. For users with higher tolerance to voice quality, you can use the **advantage-factor** command to configure the advantage factor. When the system calculates the ICPIF value, this advantage factor is subtracted to modify ICPIF and MOS values and thus both the objective and subjective factors are considered when you evaluate the voice quality.

Configuration prerequisites

A voice test requires cooperation between the NQA server and the NQA client. Before a voice test, make sure that the UDP listening function is configured on the NQA server. For the configuration of UDP listening function, see Configuring the NQA Server.

Configuring a voice test

Follow these steps to configure a voice test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as voice and enter test type view	**type voice**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation. The destination IP address must be consistent with that of the existing listening service on the NQA server.
Configure the destination port for a test operation	**destination port** *port-number*	Required By default, no destination port number is configured for a test operation. The destination port must be consistent with that of the existing listening service on the NQA server.
Configure the codec type	**codec-type** { **g711a** | **g711u** | **g729a** }	Optional By default, the codec type is G.711 A-law.

To do…	Use the command…	Remarks
Configure the advantage factor for calculating MOS and ICPIF values	**advantage-factor** *factor*	Optional By default, the advantage factor is 0.
Specify the source IP address for the requests in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Specify the source port number for the requests in a test operation	**source port** *port-number*	Optional By default, no source port number is specified.
Configure the size of a probe packet to be sent	**data-size** *size*	Optional By default, the probe packet size depends on the codec type. The default packet size is 172 bytes for G.711A-law and G.711 μ-law codec type, and is 32 bytes for G.729 A-law codec type.
Configure the filler string of a probe packet sent	**data-fill** *string*	Optional By default, the filler string of a probe packet is the hexadecimal number 00010203040506070809.
Configure the number of packets sent in a voice probe	**probe packet-number** *packet-number*	Optional 1000 by default.
Configure the interval for sending packets in a voice probe	**probe packet-interval** *packet-interval*	Optional 20 milliseconds by default.
Configure the timeout for waiting for a response in a voice test	**probe packet-timeout** *packet-timeout*	Optional 5000 milliseconds by default.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Note

Only one probe can be made in one voice test, and the number of probe packets sent in each probe depends on the configuration of the **probe packet-number** command.

Configuring a DLSw Test

A DLSw test is used to test the response time of the DLSw device.

Configuration prerequisites

Enable the DLSw function on the peer device before DLSw test.

Configuring a DLSw test

Follow these steps to configure a DLSw test:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Configure the test type as DLSw and enter test type view	**type dlsw**	Required
Configure the destination address for a test operation	**destination ip** *ip-address*	Required By default, no destination IP address is configured for a test operation.
Configure the source IP address of a probe request in a test operation	**source ip** *ip-address*	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	See Configuring Optional Parameters Common to an NQA Test Group	Optional

Configuring the Collaboration Function

Collaboration is implemented by establishing collaboration objects to monitor the detection results of the current test group. If the number of consecutive probe failures reaches the threshold, the configured action is triggered.

Follow these steps to configure the collaboration function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Enter test type view of the test group	**type** { **dhcp** | **dlsw** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** }	The collaboration function is not supported in UDP jitter or voice tests.
Create a collaboration object	**reaction** *item-num* **checked-element probe-fail threshold-type consecutive** *occurrences* [**action-type** { **none** | **trigger-only** }]	Required Not created by default.
Exit to system view	**quit**	—

To do…	Use the command…	Remarks
Create a Track object and associate it with the specified collaboration object of the NQA test group	**track** *entry-number* **nqa entry** *admin-name operation-tag* **reaction** *item-num*	Required Not created by default.

Note

- You cannot modify the content of a reaction entry using the **reaction** command after the collaboration object is created.
- The collaboration function is not supported in a UDP jitter or voice test.

Configuring Trap Delivery

Traps can be sent to the network management server when test is completed, test fails or probe fails.

Configuration prerequisites

Before configuring trap delivery, you need to configure the destination address of the trap message with the **snmp-agent target-host** command, create an NQA test group, and configure related parameters. For the introduction to the **snmp-agent target-host** command, see *SNMP Commands* in the *System Volume*.

Configuring trap delivery

Follow these steps to configure trap delivery:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Enter test type view of the test group	**type** { **dhcp** | **dlsw** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	—
Configure to send traps to network management server under specified conditions	**reaction trap** { **probe-failure** *consecutive-probe-failures* | **test-complete** | **test-failure** *cumulate-probe-failures* }	Optional No traps are sent to the network management server by default.

Note

Only the **reaction trap test-complete** command is supported in a voice test, namely, in a voice test, traps are sent to the NMS only if the test succeeds.

Configuring the NQA Statistics Function

NQA puts the NQA tests completed in a specified interval into one group, and calculates the statistics of the test results of the group. These statistics form a statistics group. You can use the **display nqa statistics** command to view information of the statistics group, and use the **statistics interval** command to set the interval for collecting statistics.

When the number of statistics groups kept reaches the upper limit, if a new statistics group is generated, the statistics group that is kept for the longest time is deleted. You can use the **statistics max-group** command to set the maximum number of statistics groups that can be kept.

A statistics group is formed after the last test is completed within the specified interval. A statistics group has the aging mechanism. A statistics group will be deleted after it is kept for a period of time. You can use the **statistics hold-time** command to set the hold time of a statistics group.

Follow these steps to configure the NQA statistics function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Enter test type view of the test group	**type** { **dlsw** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	—
Configure the interval for collecting the statistics of the test results	**statistics interval** *interval*	Optional 60 minutes by default.
Configure the maximum number of statistics groups that can be kept	**statistics max-group** *number*	Optional 2 by default. If the maximum number is 0, it indicates that no statistics is performed.
Configure the hold time of a statistics group	**statistics hold-time** *hold-time*	Optional 120 minutes by default.

Note

- The NQA statistics function is not supported in a DHCP test.
- If you specify the *interval* argument in the **frequency** *interval* command as 0, no statistics group information is generated.

Configuring Optional Parameters Common to an NQA Test Group

Optional parameters common to an NQA test group are valid only for tests in this test group.

Unless otherwise specified, the following parameters are applicable to all test types and they can be configured according to the actual conditions.

Follow these steps to configure optional parameters common to an NQA test group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter NQA test group view	**nqa entry** *admin-name operation-tag*	—
Enter test type view of a test group	**type** { **dhcp** | **dlsw** | **ftp** | **http** | **icmp-echo** | **snmp** | **tcp** | **udp-echo** | **udp-jitter** | **voice** }	—
Configure the descriptive string for a test group	**description** *text*	Optional By default, no descriptive string is available for a test group.
Configure the interval between two consecutive tests for a test group	**frequency** *interval*	Optional By default, the interval between two consecutive tests for a test group is 0 milliseconds, that is, only one test is performed. If the last test is not completed when the interval specified by the **frequency** command is reached, a new test is not started.
Configure the number of probes in an NQA test	**probe count** *times*	Optional By default, one probe is performed in a test. Only one probe can be made in one voice test. Therefore, this command is not available in a voice test.
Configure the NQA probe timeout time	**probe timeout** *timeout*	Optional By default, the timeout time is 3000 milliseconds. This parameter is not available for a UDP jitter test.
Configure the maximum number of history records that can be saved in a test group	**history-records** *number*	Optional 50 by default.
Configure the maximum number of hops a probe packet traverses in the network	**ttl** *value*	Optional 20 by default. This parameter is not available for a DHCP test.
Configure the ToS field in an IP packet header in an NQA probe packet	**tos** *value*	Optional 0 by default. This parameter is not available for a DHCP test.
Enable the routing table bypass function	**route-option bypass-route**	Optional Disabled by default. This parameter is not available for a DHCP test.

Scheduling an NQA Test Group

With this configuration, you can set the start time and test duration for a test group to perform NQA tests. The start time can take a specific value or can be **now**, which indicates that a test is started immediately; the test duration can take a specific value or can be **forever**, which indicates that a test will not stop until you use the **undo nqa schedule** command to stop the test.

A test group performs tests when the system time is between the start time and the end time (the start time plus test duration). If the system time is behind the start time when you execute the **nqa schedule** command, a test is started when the system time reaches the start time; if the system time is between the start time and the end time, a test is started at once; if the system time is ahead of the end time, no test is started. You can use the **display clock** command to view the current system time.

Configuration prerequisites

Before scheduling an NQA test group, make sure:

- Required test parameters corresponding to a test type have been configured;
- For the test which needs the cooperation with the NQA server, configuration on the NQA server has been completed.

Scheduling an NQA test group

Follow these steps to schedule an NQA test group:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Schedule an NQA test group	**nqa schedule** *admin-name operation-tag* **start-time** { *hh:mm:ss* [*yyyy/mm/dd*] | **now** } **lifetime** { *lifetime* | **forever** }	Required
Configure the maximum number of the tests that the NQA client can simultaneously perform	**nqa agent max-concurrent** *number*	Optional 2 by default.

Caution

- After an NQA test group is scheduled, you cannot enter the test group view or test type view.
- A started test group or a test group that has completed tests will not be influenced by the system time change; only a test group that is waiting to perform tests will be influenced by the system time change.

Displaying and Maintaining NQA

To do…	Use the command…	Remarks
Display history records of NQA test operation information	**display nqa history** [*admin-name operation-tag*]	Available in any view
Display the results of the last NQA test	**display nqa result** [*admin-name operation-tag*]	
Display the statistics of a type of NQA test	**display nqa statistics** [*admin-name operation-tag*]	
Display NQA server status	**display nqa server status**	

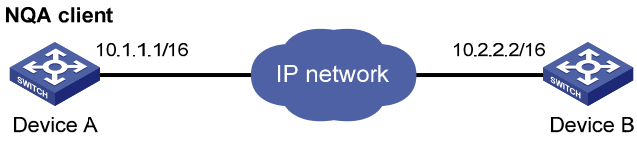
NQA Configuration Examples

ICMP Echo Test Configuration Example

Network requirements

Use the NQA ICMP function to test whether the NQA client (Device A) can send packets to the specified destination (Device B) and test the roundtrip time of packets.

Figure 1-3 Network diagram for ICMP echo tests



Configuration procedure

Create an ICMP echo test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type icmp-echo
[DeviceA-nqa-admin-test-icmp-echo] destination ip 10.2.2.2
```

Configure optional parameters.

```
[DeviceA-nqa-admin-test-icmp-echo] probe count 10
[DeviceA-nqa-admin-test-icmp-echo] probe timeout 500
[DeviceA-nqa-admin-test-icmp-echo] frequency 5000
[DeviceA-nqa-admin-test-icmp-echo] history-records 10
[DeviceA-nqa-admin-test-icmp-echo] quit
```

Enable ICMP echo test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable ICMP echo test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last ICMP echo test.

```
[DeviceA] display nqa result admin test
```

```
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 10              Receive response times: 10
      Min/Max/Average round trip time: 2/5/3
      Square-Sum of round trip time: 96
      Last succeeded probe time: 2007-08-23 15:00:01.2
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of ICMP echo tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    370        3            Succeeded        2007-08-23 15:00:01.2
    369        3            Succeeded        2007-08-23 15:00:01.2
    368        3            Succeeded        2007-08-23 15:00:01.2
    367        5            Succeeded        2007-08-23 15:00:01.2
    366        3            Succeeded        2007-08-23 15:00:01.2
    365        3            Succeeded        2007-08-23 15:00:01.2
    364        3            Succeeded        2007-08-23 15:00:01.1
    363        2            Succeeded        2007-08-23 15:00:01.1
    362        3            Succeeded        2007-08-23 15:00:01.1
    361        2            Succeeded        2007-08-23 15:00:01.1
```

DHCP Test Configuration Example

Network requirements

Use the NQA DHCP function to test the time necessary for Switch A to obtain an IP address from the DHCP server Switch B.

Figure 1-4 Network diagram for DHCP

Configuration procedure

Create a DHCP test group and configure related test parameters.

```
<SwitchA> system-view
[SwitchA] nqa entry admin test
```

```
[SwitchA-nqa-admin-test] type dhcp
[SwitchA-nqa-admin-test-dhcp] operation interface vlan-interface 2
[SwitchA-nqa-admin-test-dhcp] quit
```

Enable DHCP test.

```
[SwitchA] nqa schedule admin test start-time now lifetime forever
```

Disable DHCP test after the test begins for a period of time.

```
[SwitchA] undo nqa schedule admin test
```

Display the result of the last DHCP test.

```
[SwitchA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 624/624/624
      Square-Sum of round trip time: 389376
      Last succeeded probe time: 2007-11-22 09:56:03.2
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of DHCP tests.

```
[SwitchA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          624          Succeeded        2007-11-22 09:56:03.2
```

FTP Test Configuration Example

Network requirements

Use the NQA FTP function to test the connection with a specified FTP server and the time necessary for Device A to upload a file to the FTP server. The login username is **admin**, the login password is **systemtest**, and the file to be transferred to the FTP server is **config.txt**.

Figure 1-5 Network diagram for FTP tests

Configuration procedure

Create an FTP test group and configure related test parameters.

```
<DeviceA> system-view
```

```
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type ftp
[DeviceA-nqa-admin-test-ftp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-ftp] source ip 10.1.1.1
[DeviceA-nqa-admin-test-ftp] operation put
[DeviceA-nqa-admin-test-ftp] username admin
[DeviceA-nqa-admin-test-ftp] password systemtest
[DeviceA-nqa-admin-test-ftp] filename config.txt
[DeviceA-nqa-admin-test-ftp] quit
```

Enable FTP test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable FTP test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last FTP test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 173/173/173
      Square-Sum of round trip time: 29929
      Last succeeded probe time: 2007-11-22 10:07:28.6
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of FTP tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          173          Succeeded        2007-11-22 10:07:28.6
```

HTTP Test Configuration Example

Network requirements

Use the HTTP function to test the connection with a specified HTTP server and the time required to obtain data from the HTTP server.

Figure 1-6 Network diagram for the HTTP tests

Configuration procedure

Create an HTTP test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type http
[DeviceA-nqa-admin-test-http] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-http] operation get
[DeviceA-nqa-admin-test-http] url /index.htm
[DeviceA-nqa-admin-test-http] http-version v1.0
[DeviceA-nqa-admin-test-http] quit
```

Enable HTTP test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable HTTP test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last HTTP test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 64/64/64
      Square-Sum of round trip time: 4096
      Last succeeded probe time: 2007-11-22 10:12:47.9
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors:
      Packet(s) arrived late: 0
```

Display the history of HTTP tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          64           Succeeded        2007-11-22 10:12:47.9
```

UDP Jitter Test Configuration Example

Network requirements

Use the NQA UDP jitter function to test the delay jitter of packet transmission between Device A and Device B.

Figure 1-7 Network diagram for UDP jitter tests

Configuration procedure

1) Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 9000.

```
<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server udp-echo 10.2.2.2 9000
```

2) Configure Device A.

Create a UDP jitter test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type udp-jitter
[DeviceA-nqa-admin-test-udp-jitter] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-udp-jitter] destination port 9000
[DeviceA-nqa-admin-test-udp-jitter] frequency 1000
[DeviceA-nqa-admin-test-udp-jitter] quit
```

Enable UDP jitter test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable UDP jitter test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display the result of the last UDP jitter test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 10              Receive response times: 10
      Min/Max/Average round trip time: 15/32/17
      Square-Sum of round trip time: 3235
      Last succeeded probe time: 2008-05-29 13:56:17.6
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
```

```
    Failures due to sequence error: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packet(s) arrived late: 0
  UDP-jitter results:
   RTT number: 10
    Min positive SD: 4                    Min positive DS: 1
    Max positive SD: 21                   Max positive DS: 28
    Positive SD number: 5                 Positive DS number: 4
    Positive SD sum: 52                   Positive DS sum: 38
    Positive SD average: 10               Positive DS average: 10
    Positive SD square sum: 754           Positive DS square sum: 460
    Min negative SD: 1                    Min negative DS: 6
    Max negative SD: 13                   Max negative DS: 22
    Negative SD number: 4                 Negative DS number: 5
    Negative SD sum: 38                   Negative DS sum: 52
    Negative SD average: 10               Negative DS average: 10
    Negative SD square sum: 460           Negative DS square sum: 754
  One way results:
    Max SD delay: 15                      Max DS delay: 16
    Min SD delay: 7                       Min DS delay: 7
    Number of SD delay: 10                Number of DS delay: 10
    Sum of SD delay: 78                   Sum of DS delay: 85
    Square sum of SD delay: 666           Square sum of DS delay: 787
    SD lost packet(s): 0                  DS lost packet(s): 0
    Lost packet(s) for unknown reason: 0
```

Display the statistics of UDP jitter tests.

```
[DeviceA] display nqa statistics admin test
  NQA entry(admin admin, tag test) test statistics:
    NO. : 1
    Destination IP address: 10.2.2.2
      Start time: 2008-05-29 13:56:14.0
      Life time: 47
      Send operation times: 410           Receive response times: 410
      Min/Max/Average round trip time: 1/93/19
      Square-Sum of round trip time: 206176
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
    UDP-jitter results:
     RTT number: 410
```

```
  Min positive SD: 3                  Min positive DS: 1
  Max positive SD: 30                 Max positive DS: 79
  Positive SD number: 186             Positive DS number: 158
  Positive SD sum: 2602               Positive DS sum: 1928
  Positive SD average: 13             Positive DS average: 12
  Positive SD square sum: 45304       Positive DS square sum: 31682
  Min negative SD: 1                  Min negative DS: 1
  Max negative SD: 30                 Max negative DS: 78
  Negative SD number: 181             Negative DS number: 209
  Negative SD sum: 181                Negative DS sum: 209
  Negative SD average: 13             Negative DS average: 14
  Negative SD square sum: 46994       Negative DS square sum: 3030
One way results:
  Max SD delay: 46                    Max DS delay: 46
  Min SD delay: 7                     Min DS delay: 7
  Number of SD delay: 410             Number of DS delay: 410
  Sum of SD delay: 3705               Sum of DS delay: 3891
  Square sum of SD delay: 45987       Square sum of DS delay: 49393
  SD lost packet(s): 0                DS lost packet(s): 0
  Lost packet(s) for unknown reason: 0
```

Note

The **display nqa history** command cannot show you the results of UDP jitter tests. Therefore, to know the result of a UDP jitter test, you are recommended to use the **display nqa result** command to view the probe results of the latest NQA test, or use the **display nqa statistics** command to view the statistics of NQA tests.

SNMP Test Configuration Example

Network requirements

Use the NQA SNMP query function to test the time it takes for Device A to send an SNMP query packet to the SNMP agent and receive a response packet.

Figure 1-8 Network diagram for SNMP tests

Configuration procedure

1) Configurations on SNMP agent.

Enable the SNMP agent service and set the SNMP version to **all**, the read community to **public**, and the write community to **private**.

```
<DeviceB> system-view
[DeviceB] snmp-agent sys-info version all
[DeviceB] snmp-agent community read public
[DeviceB] snmp-agent community write private
```

2) Configurations on Device A.

Create an SNMP query test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type snmp
[DeviceA-nqa-admin-test-snmp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-snmp] quit
```

Enable SNMP query test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable SNMP query test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last SNMP test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 50/50/50
      Square-Sum of round trip time: 2500
      Last succeeded probe time: 2007-11-22 10:24:41.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of SNMP tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          50           Timeout          2007-11-22 10:24:41.1
```

TCP Test Configuration Example

Network requirements

Use the NQA TCP function to test the time for establishing a TCP connection between Device A and Device B. The port number used is 9000.

Figure 1-9 Network diagram for TCP tests

Configuration procedure

1) Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 9000.

```
<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server tcp-connect 10.2.2.2 9000
```

2) Configure Device A.

Create a TCP test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type tcp
[DeviceA-nqa-admin-test-tcp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-tcp] destination port 9000
[DeviceA-nqa-admin-test-tcp] quit
```

Enable TCP test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable TCP test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last TCP test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 13/13/13
      Square-Sum of round trip time: 169
      Last succeeded probe time: 2007-11-22 10:27:25.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of TCP tests.

```
[DeviceA] display nqa history admin test
```

```
NQA entry(admin admin, tag test) history record(s):
  Index      Response     Status           Time
  1          13           Succeeded        2007-11-22 10:27:25.1
```

UDP Echo Test Configuration Example

Network requirements

Use the NQA UDP echo function to test the round trip time between Device A and Device B. The port number is 8000.

Figure 1-10 Network diagram for the UDP echo tests

Configuration procedure

1) Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 8000.

```
<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server udp-echo 10.2.2.2 8000
```

2) Configure Device A.

Create a UDP echo test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type udp-echo
[DeviceA-nqa-admin-test-udp-echo] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-udp-echo] destination port 8000
[DeviceA-nqa-admin-test-udp-echo] quit
```

Enable UDP echo test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable UDP echo test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display results of the last UDP echo test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 25/25/25
      Square-Sum of round trip time: 625
      Last succeeded probe time: 2007-11-22 10:36:17.9
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
```

```
    Failures due to disconnect: 0
    Failures due to no connection: 0
    Failures due to sequence error: 0
    Failures due to internal error: 0
    Failures due to other errors: 0
    Packet(s) arrived late: 0
```

Display the history of UDP echo tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          25           Succeeded        2007-11-22 10:36:17.9
```

Voice Test Configuration Example

Network requirements

Use the NQA voice function to test the delay jitter of voice packet transmission and voice quality between Device A and Device B.

Figure 1-11 Network diagram for voice tests

Configuration procedure

1) Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 9000.

```
<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server udp-echo 10.2.2.2 9000
```

2) Configure Device A.

Create a voice test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type voice
[DeviceA-nqa-admin-test-voice] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-voice] destination port 9000
[DeviceA-nqa-admin-test-voice] quit
```

Enable voice test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable the voice test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display the result of the last voice test.

```
[DeviceA] display nqa result admin test
```

```
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1000           Receive response times: 1000
      Min/Max/Average round trip time: 31/1328/33
      Square-Sum of round trip time: 2844813
      Last succeeded probe time: 2008-06-13 09:49:31.1
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
    Voice results:
     RTT number: 1000
      Min positive SD: 1                   Min positive DS: 1
      Max positive SD: 204                 Max positive DS: 1297
      Positive SD number: 257              Positive DS number: 259
      Positive SD sum: 759                 Positive DS sum: 1797
      Positive SD average: 2               Positive DS average: 6
      Positive SD square sum: 54127        Positive DS square sum: 1691967
      Min negative SD: 1                   Min negative DS: 1
      Max negative SD: 203                 Max negative DS: 1297
      Negative SD number: 255              Negative DS number: 259
      Negative SD sum: 759                 Negative DS sum: 1796
      Negative SD average: 2               Negative DS average: 6
      Negative SD square sum: 53655        Negative DS square sum: 1691776
    One way results:
      Max SD delay: 343                    Max DS delay: 985
      Min SD delay: 343                    Min DS delay: 985
      Number of SD delay: 1                Number of DS delay: 1
      Sum of SD delay: 343                 Sum of DS delay: 985
      Square sum of SD delay: 117649       Square sum of DS delay: 970225
      SD lost packet(s): 0                 DS lost packet(s): 0
      Lost packet(s) for unknown reason: 0
    Voice scores:
      MOS value: 4.38                      ICPIF value: 0
```

Display the statistics of voice tests.

```
[DeviceA] display nqa statistics admin test
  NQA entry(admin admin, tag test) test statistics:
    NO. : 1
    Destination IP address: 10.2.2.2
      Start time: 2008-06-13 09:45:37.8
      Life time: 331
      Send operation times: 4000           Receive response times: 4000
```

```
  Min/Max/Average round trip time: 15/1328/32
  Square-Sum of round trip time: 7160528
Extended results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
  Packet(s) arrived late: 0
Voice results:
 RTT number: 4000
  Min positive SD: 1                    Min positive DS: 1
  Max positive SD: 360                  Max positive DS: 1297
  Positive SD number: 1030              Positive DS number: 1024
  Positive SD sum: 4363                 Positive DS sum: 5423
  Positive SD average: 4                Positive DS average: 5
  Positive SD square sum: 497725        Positive DS square sum: 2254957
  Min negative SD: 1                    Min negative DS: 1
  Max negative SD: 360                  Max negative DS: 1297
  Negative SD number: 1028              Negative DS number: 1022
  Negative SD sum: 1028                 Negative DS sum: 1022
  Negative SD average: 4                Negative DS average: 5
  Negative SD square sum: 495901        Negative DS square sum: 5419
One way results:
  Max SD delay: 359                     Max DS delay: 985
  Min SD delay: 0                       Min DS delay: 0
  Number of SD delay: 4                 Number of DS delay: 4
  Sum of SD delay: 1390                 Sum of DS delay: 1079
  Square sum of SD delay: 483202        Square sum of DS delay: 973651
  SD lost packet(s): 0                  DS lost packet(s): 0
  Lost packet(s) for unknown reason: 0
Voice scores:
  Max MOS value: 4.38                   Min MOS value: 4.38
  Max ICPIF value: 0                    Min ICPIF value: 0
```

Note

The **display nqa history** command cannot show you the results of voice tests. Therefore, to know the result of a voice test, you are recommended to use the **display nqa result** command to view the probe results of the latest NQA test, or use the **display nqa statistics** command to view the statistics of NQA tests.

DLSw Test Configuration Example

Network requirements

Use the NQA DLSw function to test the response time of the DLSw device.

Figure 1-12 Network diagram for the DLSw tests

Configuration procedure

Create a DLSw test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type dlsw
[DeviceA-nqa-admin-test-dlsw] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-dlsw] quit
```

Enable DLSw test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Disable DLSw test after the test begins for a period of time.

```
[DeviceA] undo nqa schedule admin test
```

Display the result of the last DLSw test.

```
[DeviceA] display nqa result admin test
  NQA entry(admin admin, tag test) test results:
    Destination IP address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average round trip time: 19/19/19
      Square-Sum of round trip time: 361
      Last succeeded probe time: 2007-11-22 10:40:27.7
    Extended results:
      Packet lost in test: 0%
      Failures due to timeout: 0
      Failures due to disconnect: 0
      Failures due to no connection: 0
      Failures due to sequence error: 0
      Failures due to internal error: 0
      Failures due to other errors: 0
      Packet(s) arrived late: 0
```

Display the history of DLSw tests.

```
[DeviceA] display nqa history admin test
  NQA entry(admin admin, tag test) history record(s):
    Index      Response     Status           Time
    1          19           Succeeded        2007-11-22 10:40:27.7
```

Table of Contents

1 NTP Configuration

When configuring NTP, go to these sections for information you are interested in:

NTP Overview

Defined in RFC 1305, the Network Time Protocol (NTP) synchronizes timekeeping among distributed time servers and clients. NTP runs over the User Datagram Protocol (UDP), using UDP port 123.

The purpose of using NTP is to keep consistent timekeeping among all clock-dependent devices within the network so that the devices can provide diverse applications based on the consistent time.

For a local system running NTP, its time can be synchronized by other reference sources and can be used as a reference source to synchronize other clocks.

Applications of NTP

An administrator can by no means keep time synchronized among all the devices within a network by changing the system clock on each station, because this is a huge amount of workload and cannot guarantee the clock precision. NTP, however, allows quick clock synchronization within the entire network while it ensures a high clock precision.

NTP is used when all devices within the network must be consistent in timekeeping, for example:

- In analysis of the log information and debugging information collected from different devices in network management, time must be used as reference basis.
- All devices must use the same reference clock in a charging system.
- To implement certain functions, such as scheduled restart of all devices within the network, all devices must be consistent in timekeeping.
- When multiple systems process a complex event in cooperation, these systems must use that same reference clock to ensure the correct execution sequence.
- For incremental backup between a backup server and clients, timekeeping must be synchronized between the backup server and all the clients.

Advantages of NTP

- NTP uses a stratum to describe the clock precision, and is able to synchronize time among all devices within the network.
- NTP supports access control and MD5 authentication.

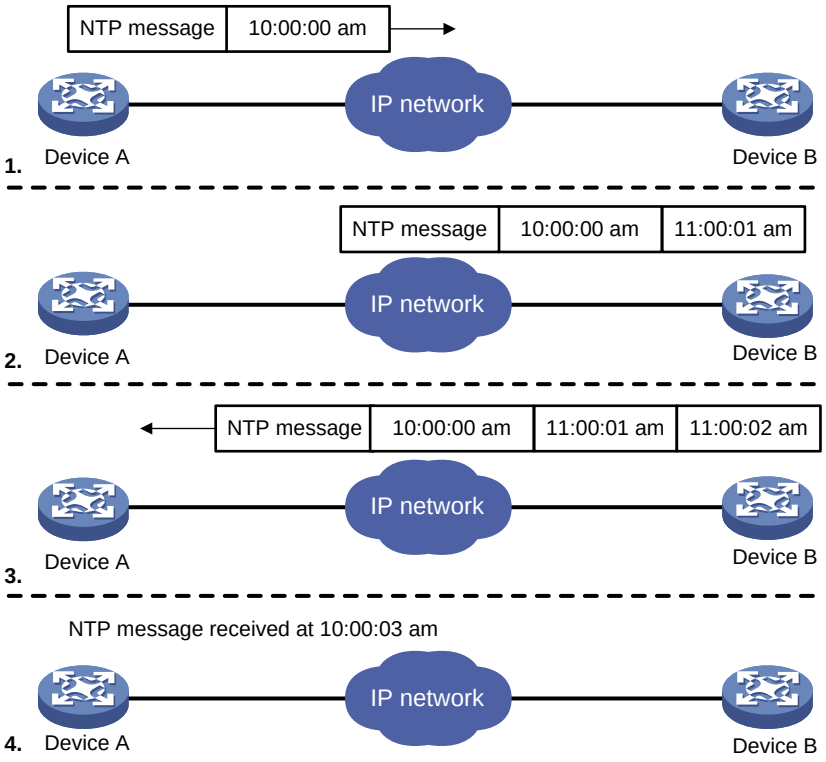
- NTP can unicast, multicast or broadcast protocol messages.

How NTP Works

Figure 1-1 shows the basic workflow of NTP. Device A and Device B are interconnected over a network. They have their own independent system clocks, which need to be automatically synchronized through NTP. For an easy understanding, we assume that:

- Prior to system clock synchronization between Device A and Device B, the clock of Device A is set to 10:00:00 am while that of Device B is set to 11:00:00 am.
- Device B is used as the NTP time server, namely, Device A synchronizes its clock to that of Device B.
- It takes 1 second for an NTP message to travel from one device to the other.

Figure 1-1 Basic work flow of NTP



The process of system clock synchronization is as follows:

- Device A sends Device B an NTP message, which is timestamped when it leaves Device A. The time stamp is 10:00:00 am (T1).
- When this NTP message arrives at Device B, it is timestamped by Device B. The timestamp is 11:00:01 am (T2).
- When the NTP message leaves Device B, Device B timestamps it. The timestamp is 11:00:02 am (T3).
- When Device A receives the NTP message, the local time of Device A is 10:00:03 am (T4).

Up to now, Device A has sufficient information to calculate the following two important parameters:

- The roundtrip delay of NTP message: Delay = (T4–T1) – (T3-T2) = 2 seconds.
- Time difference between Device A and Device B: Offset = ((T2-T1) + (T3-T4))/2 = 1 hour.

Based on these parameters, Device A can synchronize its own clock to the clock of Device B.

This is only a rough description of the work mechanism of NTP. For details, refer to RFC 1305.

NTP Message Format

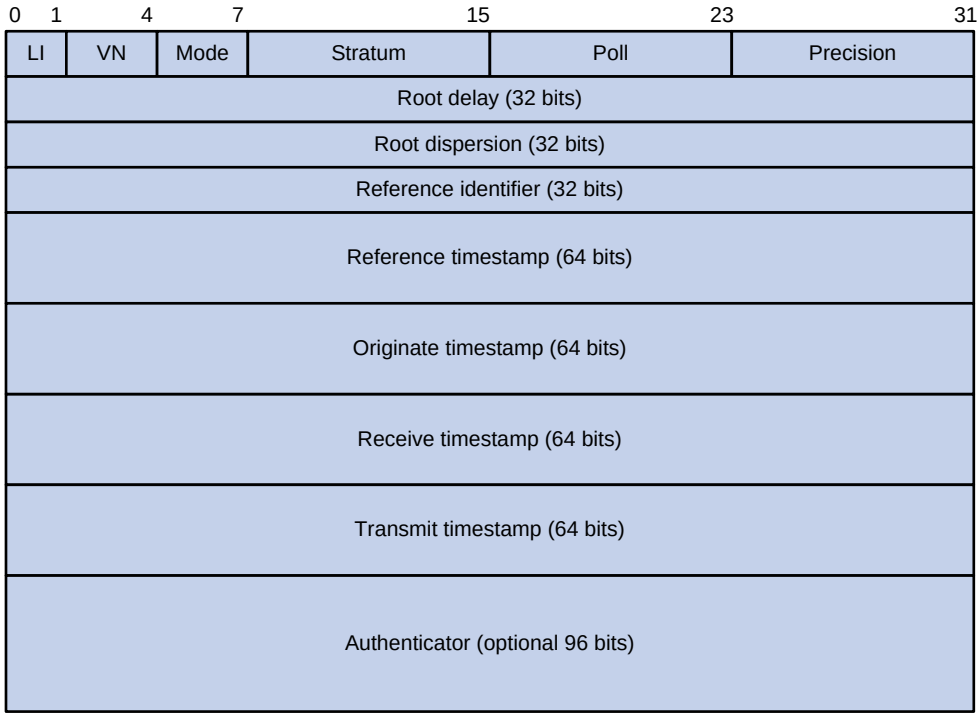
NTP uses two types of messages, clock synchronization message and NTP control message. An NTP control message is used in environments where network management is needed. As it is not a must for clock synchronization, it will not be discussed in this document.

Note

All NTP messages mentioned in this document refer to NTP clock synchronization messages.

A clock synchronization message is encapsulated in a UDP message, in the format shown in Figure 1-2.

Figure 1-2 Clock synchronization message format



Main fields are described as follows:

- LI: 2-bit leap indicator. When set to 11, it warns of an alarm condition (clock unsynchronized); when set to any other value, it is not to be processed by NTP.
- VN: 3-bit version number, indicating the version of NTP. The latest version is version 3.
- Mode: a 3-bit code indicating the work mode of NTP. This field can be set to these values: 0 – reserved; 1 – symmetric active; 2 – symmetric passive; 3 – client; 4 – server; 5 – broadcast or multicast; 6 – NTP control message; 7 – reserved for private use.
- Stratum: an 8-bit integer indicating the stratum level of the local clock, with the value ranging from 1 to 16. The clock precision decreases from stratum 1 through stratum 16. A stratum 1 clock has the highest precision, and a stratum 16 clock is not synchronized and cannot be used as a reference clock.
- Poll: 8-bit signed integer indicating the poll interval, namely the maximum interval between successive messages.

- Precision: an 8-bit signed integer indicating the precision of the local clock.
- Root Delay: roundtrip delay to the primary reference source.
- Root Dispersion: the maximum error of the local clock relative to the primary reference source.
- Reference Identifier: Identifier of the particular reference source.
- Reference Timestamp: the local time at which the local clock was last set or corrected.
- Originate Timestamp: the local time at which the request departed from the client for the service host.
- Receive Timestamp: the local time at which the request arrived at the service host.
- Transmit Timestamp: the local time at which the reply departed from the service host for the client.
- Authenticator: authentication information.

Operation Modes of NTP

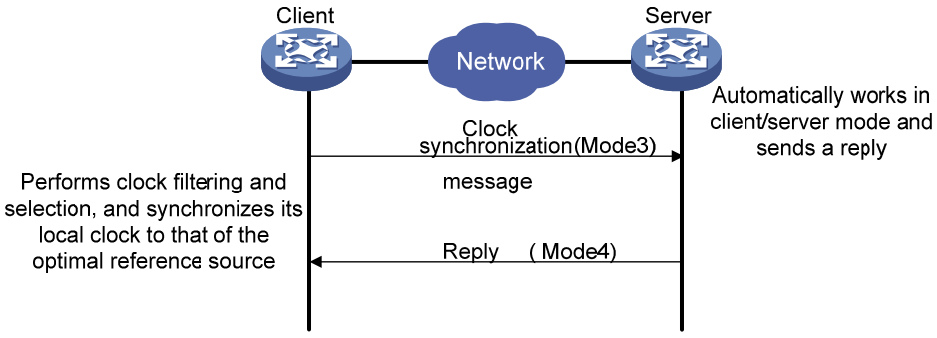
Devices running NTP can implement clock synchronization in one of the following modes:

- Client/server mode
- Symmetric peers mode
- Broadcast mode
- Multicast mode

You can select operation modes of NTP as needed. In case that the IP address of the NTP server or peer is unknown and many devices in the network need to be synchronized, you can adopt the broadcast or multicast mode; while in the client/server and symmetric peers modes, a device is synchronized from the specified server or peer, and thus clock reliability is enhanced.

Client/server mode

Figure 1-3 Client/server mode

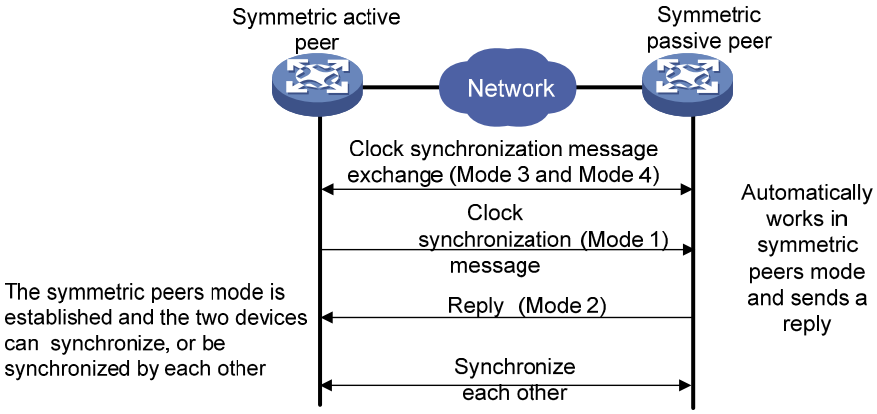


When working in the client/server mode, a client sends a clock synchronization message to servers, with the Mode field in the message set to 3 (client mode). Upon receiving the message, the servers automatically work in the server mode and send a reply, with the Mode field in the messages set to 4 (server mode). Upon receiving the replies from the servers, the client performs clock filtering and selection, and synchronizes its local clock to that of the optimal reference source.

In this mode, a client can be synchronized to a server, but not vice versa.

Symmetric peers mode

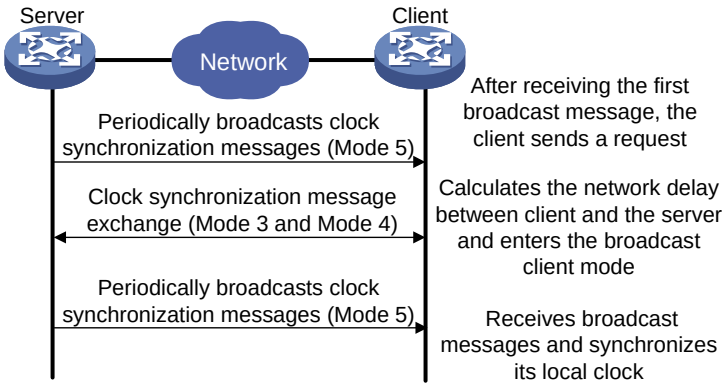
Figure 1-4 Symmetric peers mode



A device working in the symmetric active mode periodically sends clock synchronization messages, with the Mode field in the message set to 1 (symmetric active); the device that receives this message automatically enters the symmetric passive mode and sends a reply, with the Mode field in the message set to 2 (symmetric passive). By exchanging messages, the symmetric peers mode is established between the two devices. Then, the two devices can synchronize, or be synchronized by each other. If the clocks of both devices have been already synchronized, the device whose local clock has a lower stratum level will synchronize the clock of the other device.

Broadcast mode

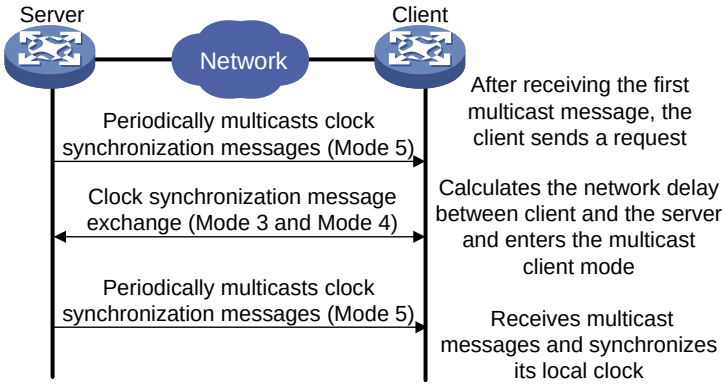
Figure 1-5 Broadcast mode



In the broadcast mode, a server periodically sends clock synchronization messages to the broadcast address 255.255.255.255, with the Mode field in the messages set to 5 (broadcast mode). Clients listen to the broadcast messages from servers. After a client receives the first broadcast message, the client and the server start to exchange messages, with the Mode field set to 3 (client mode) and 4 (server mode) to calculate the network delay between client and the server. Then, the client enters the broadcast client mode and continues listening to broadcast messages, and synchronizes its local clock based on the received broadcast messages.

Multicast mode

Figure 1-6 Multicast mode



In the multicast mode, a server periodically sends clock synchronization messages to the user-configured multicast address, or, if no multicast address is configured, to the default NTP multicast address 224.0.1.1, with the Mode field in the messages set to 5 (multicast mode). Clients listen to the multicast messages from servers. After a client receives the first multicast message, the client and the server start to exchange messages, with the Mode field set to 3 (client mode) and 4 (server mode) to calculate the network delay between client and the server. Then, the client enters the multicast client mode and continues listening to multicast messages, and synchronizes its local clock based on the received multicast messages.

Note

In symmetric peers mode, broadcast mode and multicast mode, the client (or the symmetric active peer) and the server (the symmetric passive peer) can work in the specified NTP working mode only after they exchange NTP messages with the Mode field being 3 (client mode) and the Mode field being 4 (server mode). During this message exchange process, NTP clock synchronization can be implemented.

Multiple Instances of NTP

The client/server mode and symmetric mode support multiple instances of NTP and thus support clock synchronization within more than one VPN network. Namely, network devices at different physical location can get their clocks synchronized through NTP, as long as they are in the same VPN.

NTP Configuration Task List

Complete the following tasks to configure NTP:

Task	Remarks
Configuring the Operation Modes of NTP	Required
Configuring Optional Parameters of NTP	Optional

Task	Remarks
Configuring Access-Control Rights	Optional
Configuring NTP Authentication	Optional

Configuring the Operation Modes of NTP

Devices can implement clock synchronization in one of the following modes:

- Client/server mode
- Symmetric mode
- Broadcast mode
- Multicast mode

For the client/server mode or symmetric mode, you need to configure only clients or symmetric-active peers; for the broadcast or multicast mode, you need to configure both servers and clients.

Note

A single device can have a maximum of 128 associations at the same time, including static associations and dynamic associations. A static association refers to an association that a user has manually created by using an NTP command, while a dynamic association is a temporary association created by the system during operation. A dynamic association will be removed if the system fails to receive messages from it over a specific long time. In the client/server mode, for example, when you carry out a command to synchronize the time to a server, the system will create a static association, and the server will just respond passively upon the receipt of a message, rather than creating an association (static or dynamic). In the symmetric mode, static associations will be created at the symmetric-active peer side, and dynamic associations will be created at the symmetric-passive peer side; in the broadcast or multicast mode, static associations will be created at the server side, and dynamic associations will be created at the client side.

Configuring NTP Client/Server Mode

For devices working in the client/server mode, you only need to make configurations on the clients, but not on the servers.

Follow these steps to configure an NTP client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify an NTP server for the device	**ntp-service unicast-server** [**vpn-instance** *vpn-instance-name*] { *ip-address* | *server-name* } [**authentication-keyid** *keyid* | **priority** | **source-interface** *interface-type interface-number* | **version** *number*] *	Required No NTP server is specified by default.

Note

- In the **ntp-service unicast-server** command, *ip-address* must be a unicast address, rather than a broadcast address, a multicast address or the IP address of the local clock.
- When the source interface for NTP messages is specified by the **source-interface** argument, the source IP address of the NTP messages will be configured as the primary IP address of the specified interface.
- A device can act as a server to synchronize the clock of other devices only after its clock has been synchronized. If the clock of a server has a stratum level higher than or equal to that of a client's clock, the client will not synchronize its clock to the server's.
- You can configure multiple servers by repeating the **ntp-service unicast-server** command. The clients will choose the optimal reference source.

Configuring the NTP Symmetric Peers Mode

For devices working in the symmetric mode, you need to specify a symmetric-passive peer on a symmetric-active peer.

Following these steps to configure a symmetric-active device:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify a symmetric-passive peer for the device	**ntp-service unicast-peer** [**vpn-instance** *vpn-instance-name*] { *ip-address* | *peer-name* } [**authentication-keyid** *keyid* | **priority** | **source-interface** *interface-type interface-number* | **version** *number*] *	Required No symmetric-passive peer is specified by default.

Note

- In the symmetric mode, you should use any NTP configuration command in Configuring the Operation Modes of NTP to enable NTP; otherwise, a symmetric-passive peer will not process NTP messages from a symmetric-active peer.
- In the **ntp-service unicast-peer** command, *ip-address* must be a unicast address, rather than a broadcast address, a multicast address or the IP address of the local clock.
- When the source interface for NTP messages is specified by the **source-interface** argument, the source IP address of the NTP messages will be configured as the primary IP address of the specified interface.
- Typically, at least one of the symmetric-active and symmetric-passive peers has been synchronized; otherwise the clock synchronization will not proceed.
- You can configure multiple symmetric-passive peers by repeating the **ntp-service unicast-peer** command.

Configuring NTP Broadcast Mode

The broadcast server periodically sends NTP broadcast messages to the broadcast address 255.255.255.255. After receiving the messages, the device working in NTP broadcast client mode sends a reply and synchronizes its local clock.

For devices working in the broadcast mode, you need to configure both the server and clients. Because an interface needs to be specified on the broadcast server for sending NTP broadcast messages and an interface also needs to be specified on each broadcast client for receiving broadcast messages, the NTP broadcast mode can be configured only in the specific interface view.

Configuring a broadcast client

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Required Enter the interface used to receive NTP broadcast messages.
Configure the device to work in the NTP broadcast client mode	**ntp-service broadcast-client**	Required

Configuring the broadcast server

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Enter the interface used to send NTP broadcast messages.
Configure the device to work in the NTP broadcast server mode	**ntp-service broadcast-server** [**authentication-keyid** *keyid* | **version** *number*] *	Required

Note

A broadcast server can synchronize broadcast clients only after its clock has been synchronized.

Configuring NTP Multicast Mode

The multicast server periodically sends NTP multicast messages to multicast clients, which send replies after receiving the messages and synchronize their local clocks.

For devices working in the multicast mode, you need to configure both the server and clients. The NTP multicast mode must be configured in the specific interface view.

Configuring a multicast client

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Enter the interface used to receive NTP multicast messages.
Configure the device to work in the NTP multicast client mode	**ntp-service multicast-client** [*ip-address*]	Required

Configuring the multicast server

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	Enter the interface used to send NTP multicast message.
Configure the device to work in the NTP multicast server mode	**ntp-service multicast-server** [*ip-address*] [**authentication-keyid** *keyid* | **ttl** *ttl-number* | **version** *number*] *	Required

Note

- A multicast server can synchronize broadcast clients only after its clock has been synchronized.
- You can configure up to 1024 multicast clients, among which 128 can take effect at the same time.

Configuring Optional Parameters of NTP

Specifying the Source Interface for NTP Messages

If you specify the source interface for NTP messages, the device sets the source IP address of the NTP messages as the primary IP address of the specified interface when sending the NTP messages.

When the device responds to an NTP request received, the source IP address of the NTP response is always the IP address of the interface that received the NTP request.

Following these steps to specify the source interface for NTP messages:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify the source interface for NTP messages	**ntp-service source-interface** *interface-type interface-number*	Required By default, no source interface is specified for NTP messages, and the system uses the IP address of the interface determined by the matching route as the source IP address of NTP messages.

Caution

- If you have specified the source interface for NTP messages in the **ntp-service unicast-server** or **ntp-service unicast-peer** command, the interface specified in the **ntp-service unicast-server** or **ntp-service unicast-peer** command serves as the source interface of NTP messages.
- If you have configured the **ntp-service broadcast-server** or **ntp-service multicast-server** command, the source interface of the broadcast or multicast NTP messages is the interface configured with the respective command.

Disabling an Interface from Receiving NTP Messages

When NTP is enabled, NTP messages can be received from all the interfaces by default, and you can disable an interface from receiving NTP messages through the following configuration.

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Disable the interface from receiving NTP messages	**ntp-service in-interface disable**	Required An interface is enabled to receive NTP messages by default.

Configuring the Maximum Number of Dynamic Sessions Allowed

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the maximum number of dynamic sessions allowed to be established locally	**ntp-service max-dynamic-sessions** *number*	Required 100 by default

Configuring Access-Control Rights

With the following command, you can configure the NTP service access-control right to the local device. There are four access-control rights, as follows:

- **query**: control query permitted. This level of right permits the peer devices to perform control query to the NTP service on the local device but does not permit a peer device to synchronize its clock to that of the local device. The so-called “control query” refers to query of some states of the NTP service, including alarm information, authentication status, clock source information, and so on.
- **synchronization**: server access only. This level of right permits a peer device to synchronize its clock to that of the local device but does not permit the peer devices to perform control query.
- **server**: server access and query permitted. This level of right permits the peer devices to perform synchronization and control query to the local device but does not permit the local device to synchronize its clock to that of a peer device.

- **peer**: full access. This level of right permits the peer devices to perform synchronization and control query to the local device and also permits the local device to synchronize its clock to that of a peer device.

From the highest NTP service access-control right to the lowest one are **peer**, **server**, **synchronization**, and **query**. When a device receives an NTP request, it will perform an access-control right match and will use the first matched right.

Configuration Prerequisites

Prior to configuring the NTP service access-control right to the local device, you need to create and configure an ACL associated with the access-control right. For the configuration of ACL, refer to *ACL Configuration* in the *Security Volume*.

Configuration Procedure

Follow these steps to configure the NTP service access-control right to the local device:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the NTP service access-control right for a peer device to access the local device	**ntp-service access** { **peer** | **query** | **server** | **synchronization** } *acl-number*	Required **peer** by default

Note

The access-control right mechanism provides only a minimum degree of security protection for the system running NTP. A more secure method is identity authentication.

Configuring NTP Authentication

The NTP authentication feature should be enabled for a system running NTP in a network where there is a high security demand. This feature enhances the network security by means of client-server key authentication, which prohibits a client from synchronizing with a device that has failed authentication.

Configuration Prerequisites

The configuration of NTP authentication involves configuration tasks to be implemented on the client and on the server.

When configuring the NTP authentication feature, pay attention to the following principles:

- For all synchronization modes, when you enable the NTP authentication feature, you should configure an authentication key and specify it as a trusted key. Namely, the **ntp-service authentication enable** command must work together with the **ntp-service authentication-keyid** command and the **ntp-service reliable authentication-keyid** command. Otherwise, the NTP authentication function cannot be normally enabled.
- For the client/server mode or symmetric mode, you need to associate the specified authentication key on the client (symmetric-active peer if in the symmetric peer mode) with the corresponding

NTP server (symmetric-passive peer if in the symmetric peer mode). Otherwise, the NTP authentication feature cannot be normally enabled.

- For the broadcast server mode or multicast server mode, you need to associate the specified authentication key on the broadcast server or multicast server with the corresponding NTP server. Otherwise, the NTP authentication feature cannot be normally enabled.
- For the client/server mode, if the NTP authentication feature has not been enabled for the client, the client can synchronize with the server regardless of whether the NTP authentication feature has been enabled for the server or not. If the NTP authentication is enabled on a client, the client can be synchronized only to a server that can provide a trusted authentication key.
- For all synchronization modes, the server side and the client side must be consistently configured.

Configuration Procedure

Configuring NTP authentication for a client

Follow these steps to configure NTP authentication for a client:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable NTP authentication	**ntp-service authentication enable**	Required Disabled by default
Configure an NTP authentication key	**ntp-service authentication-keyid** *keyid* **authentication-mode md5** *value*	Required No NTP authentication key by default
Configure the key as a trusted key	**ntp-service reliable authentication-keyid** *keyid*	Required No authentication key is configured to be trusted by default.
Associate the specified key with an NTP server	Client/server mode: **ntp-service unicast-server** { *ip-address* | *server-name* } **authentication-keyid** *keyid*	Required You can associate a non-existing key with an NTP server. To enable NTP authentication, you must configure the key and specify it as a trusted key after associating the key with the NTP server.
	Symmetric peers mode: **ntp-service unicast-peer** { *ip-address* | *peer-name* } **authentication-keyid** *keyid*	

Note

After you enable the NTP authentication feature for the client, make sure that you configure for the client an authentication key that is the same as on the server and specify that the authentication key is trusted; otherwise, the client cannot be synchronized to the server.

Configuring NTP authentication for a server

Follow these steps to configure NTP authentication for a server:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable NTP authentication	**ntp-service authentication enable**	Required Disabled by default
Configure an NTP authentication key	**ntp-service authentication-keyid** *keyid* **authentication-mode md5** *value*	Required No NTP authentication key by default
Configure the key as a trusted key	**ntp-service reliable authentication-keyid** *keyid*	Required No authentication key is configured to be trusted by default.
Enter interface view	**interface** *interface-type interface-number*	—
Associate the specified key with an NTP server	Broadcast server mode: **ntp-service broadcast-server authentication-keyid** *keyid*	Required You can associate a non-existing key with an NTP server. To enable NTP authentication, you must configure the key and specify it as a trusted key after associating the key with the NTP server.
	Multicast server mode: **ntp-service multicast-server authentication-keyid** *keyid*	

Note

The procedure of configuring NTP authentication on a server is the same as that on a client, and the same authentication key must be configured on both the server and client sides.

Displaying and Maintaining NTP

To do…	Use the command…	Remarks
View the information of NTP service status	**display ntp-service status**	Available in any view
View the information of NTP sessions	**display ntp-service sessions** [**verbose**]	Available in any view
View the brief information of the NTP servers from the local device back to the primary reference source	**display ntp-service trace**	Available in any view

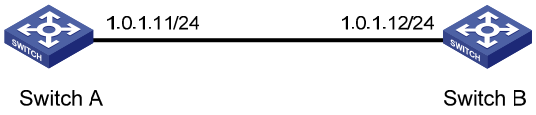
NTP Configuration Examples

Configuring NTP Client/Server Mode

Network requirements

- The local clock of Switch A is to be used as a master clock, with the stratum level of 2.
- Switch B works in the client/server mode and Switch A is to be used as the NTP server of Switch B.

Figure 1-7 Network diagram for NTP client/server mode configuration



Configuration procedure

View the NTP status of Switch B before clock synchronization.

```
<SwitchB> display ntp-service status
 Clock status: unsynchronized
 Clock stratum: 16
 Reference clock ID: none
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 0.00 ms
 Root dispersion: 0.00 ms
 Peer dispersion: 0.00 ms
 Reference time: 00:00:00.000 UTC Jan 1 1900 (00000000.00000000)
```

Specify Switch A as the NTP server of Switch B so that Switch B is synchronized to Switch A.

```
<SwitchB> system-view
[SwitchB] ntp-service unicast-server 1.0.1.11
```

View the NTP status of Switch B after clock synchronization.

```
[SwitchB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 1.0.1.11
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 1.05 ms
 Peer dispersion: 7.81 ms
 Reference time: 14:53:27.371 UTC Sep 19 2005 (C6D94F67.5EF9DB22)
```

As shown above, Switch B has been synchronized to Switch A, and the clock stratum level of Switch B is 3, while that of Switch A is 2.

View the NTP session information of Switch B, which shows that an association has been set up between Switch B and Switch A.

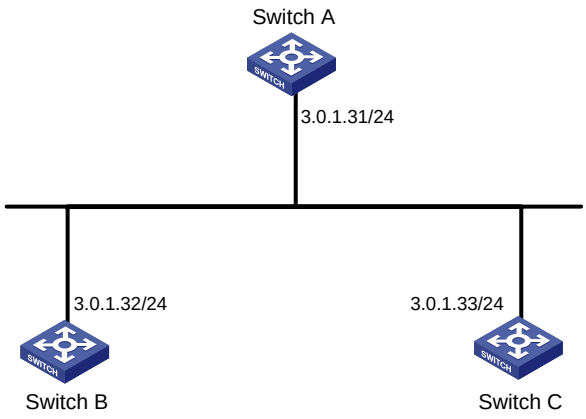
```
[SwitchB] display ntp-service sessions
      source      reference   stra  reach  poll  now  offset  delay  disper
**************************************************************************
[12345] 1.0.1.11  127.127.1.0    2    63    64    3    -75.5    31.0  16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

Configuring the NTP Symmetric Mode

Network requirements

- The local clock of Switch A is to be used as the master clock, with a stratum level of 2.
- Switch B works in the client mode and Switch A is to be used as the NTP server of Switch B.
- Switch C works in the symmetric-active mode and Switch B will act as peer of Switch C. Switch C is the symmetric-active peer while Switch B is the symmetric-passive peer.

Figure 1-8 Network diagram for NTP symmetric peers mode configuration



Configuration procedure

1) Configuration on Switch B:

Specify Switch A as the NTP server of Switch B.

```
<SwitchB> system-view
[SwitchB] ntp-service unicast-server 3.0.1.31
```

2) Configuration on Switch C (after Switch B is synchronized to Switch A):

Specify the local clock as the reference source, with the stratum level of 1.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 1
```

Configure Switch B as a symmetric peer after local synchronization.

```
[SwitchC] ntp-service unicast-peer 3.0.1.32
```

In the step above, Switch B and Switch C are configured as symmetric peers, with Switch C in the symmetric-active mode and Switch B in the symmetric-passive mode. Because the stratus level of Switch C is 1 while that of Switch B is 3, Switch B is synchronized to Switch C.

View the NTP status of Switch B after clock synchronization.

```
[SwitchB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 3.0.1.33
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: -21.1982 ms
 Root delay: 15.00 ms
 Root dispersion: 775.15 ms
 Peer dispersion: 34.29 ms
 Reference time: 15:22:47.083 UTC Sep 19 2005 (C6D95647.153F7CED)
```

As shown above, Switch B has been synchronized to Switch C, and the clock stratum level of Switch B is 2, while that of Switch C is 1.

View the NTP session information of Switch B, which shows that an association has been set up between Switch B and Switch C.

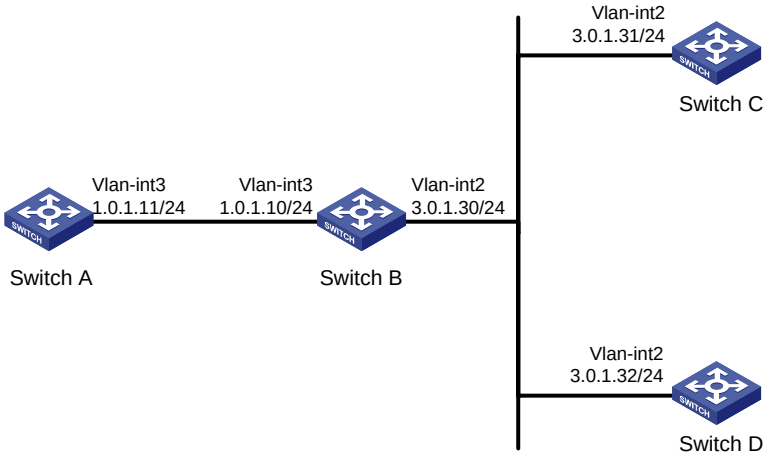
```
[SwitchB] display ntp-service sessions
       source     reference   stra  reach  poll  now   offset delay  disper
**************************************************************************
[245] 3.0.1.31  127.127.1.0    2    15    64   24   10535.0  19.6   14.5
[1234] 3.0.1.33   LOCL          1    14    64   27    -77.0   16.0   14.8
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  2
```

Configuring NTP Broadcast Mode

Network requirements

- The local clock of Switch C is to be used as the master clock, with a stratum level of 2.
- Switch C works in the broadcast server mode and sends out broadcast messages from VLAN-interface 2.
- Switch A and Switch D work in the broadcast client mode. Switch A listens to broadcast messages through its VLAN-interface 3 and Switch D from its VLAN-interface 2.

Figure 1-9 Network diagram for NTP broadcast mode configuration



Configuration procedure

1) Configuration on Switch C:

Configure Switch C to work in the broadcast server mode and send broadcast messages through VLAN-interface 2.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server
```

2) Configuration on Switch D:

Configure Switch D to work in the broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service broadcast-client
```

3) Configuration on Switch A:

Configure Switch A to work in the broadcast client mode and receive broadcast messages on VLAN-interface 3.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ntp-service broadcast-client
```

Because Switch A and Switch C are on different subnets, Switch A cannot receive the broadcast messages from Switch C. Switch D gets synchronized upon receiving a broadcast message from Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 3.0.1.31
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 8.31 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch C, and the clock stratum level of Switch D is 3, while that of Switch C is 2.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD-Vlan-interface2] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset   delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0   2   254     64    62   -16.0    32.0   16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
```

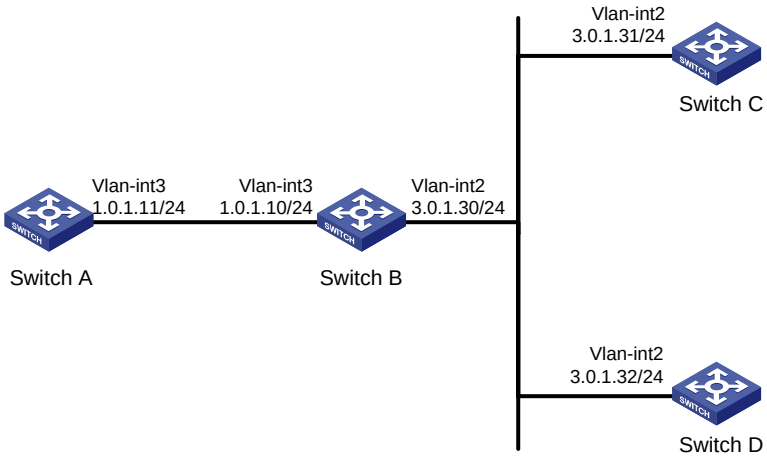
```
Total associations :  1
```

Configuring NTP Multicast Mode

Network requirements

- The local clock of Switch C is to be used as the master clock, with a stratum level of 2.
- Switch C works in the multicast server mode and sends out multicast messages from VLAN-interface 2.
- Switch A and Switch D work in the multicast client mode and receive multicast messages through VLAN-interface 3 and VLAN-interface 2 respectively.

Figure 1-10 Network diagram for NTP multicast mode configuration



Configuration procedure

1) Configuration on Switch C:

Configure Switch C to work in the multicast server mode and send multicast messages through VLAN-interface 2.

```
<SwitchC> system-view
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service multicast-server
```

2) Configuration on Switch D:

Configure Switch D to work in the multicast client mode and receive multicast messages on VLAN-interface 2.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service multicast-client
```

Because Switch D and Switch C are on the same subnet, Switch D can receive the multicast messages from Switch C without being enabled with the multicast functions and can be synchronized to Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD-Vlan-interface2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 3.0.1.31
```

```
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 8.31 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch C, and the clock stratum level of Switch D is 3, while that of Switch C is 2.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD-Vlan-interface2] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset  delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0   2   254     64    62   -16.0    31.0   16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

3) Configuration on Switch B:

Because Switch A and Switch C are on different subnets, you must enable the multicast functions on Switch B before Switch A can receive multicast messages from Switch C.

Enable IP multicast routing and IGMP.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] pim dm
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/1
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] igmp enable
[SwitchB-Vlan-interface3] quit
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] igmp-snooping static-group 224.0.1.1 vlan 3
```

4) Configuration on Switch A:

Enable IP multicast routing and IGMP.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 3
```

Configure Switch A to work in the multicast client mode and receive multicast messages on VLAN-interface 3.

```
[SwitchA-Vlan-interface3] ntp-service multicast-client
```

View the NTP status of Switch A after clock synchronization.

```
[SwitchA-Vlan-interface3] display ntp-service status
 Clock status: synchronized
```

```
Clock stratum: 3
Reference clock ID: 3.0.1.31
Nominal frequency: 64.0000 Hz
Actual frequency: 64.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 40.00 ms
Root dispersion: 10.83 ms
Peer dispersion: 34.30 ms
Reference time: 16:02:49.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

As shown above, Switch A has been synchronized to Switch C, and the clock stratum level of Switch A is 3, while that of Switch C is 2.

View the NTP session information of Switch A, which shows that an association has been set up between Switch A and Switch C.

```
[SwitchA-Vlan-interface3] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset   delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0    2   255     64    26   -16.0    40.0   16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

Note

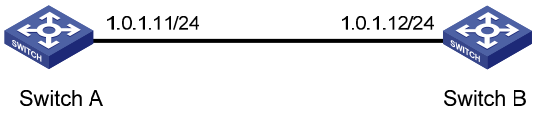
Refer to *IGMP Configuration* in the *IP Multicast* volume for how to configure IGMP and PIM.

Configuring NTP Client/Server Mode with Authentication

Network requirements

- The local clock of Switch A is to be used as the master clock, with a stratum level of 2.
- Switch B works in the client mode and Switch A is to be used as the NTP server of Switch B, with Switch B as the client.
- NTP authentication is to be enabled on both Switch A and Switch B.

Figure 1-11 Network diagram for configuration of NTP client/server mode with authentication



Configuration procedure

Configuration on Switch B:

Enable NTP authentication on Switch B.

```
<SwitchB> system-view
[SwitchB] ntp-service authentication enable
```

Set an authentication key.

```
[SwitchB] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

Specify the key as a trusted key.

```
[SwitchB] ntp-service reliable authentication-keyid 42
```

Specify Switch A as the NTP server.

```
[SwitchB] ntp-service unicast-server 1.0.1.11 authentication-keyid 42
```

Before Switch B can synchronize its clock to that of Switch A, you need to enable NTP authentication for Switch A.

Perform the following configuration on Switch A:

Enable NTP authentication.

```
[SwitchA] ntp-service authentication enable
```

Set an authentication key.

```
[SwitchA] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

Specify the key as a trusted key.

```
[SwitchA] ntp-service reliable authentication-keyid 42
```

View the NTP status of Switch B after clock synchronization.

```
[SwitchB] display ntp-service status
 Clock status: synchronized
 Clock stratum: 3
 Reference clock ID: 1.0.1.11
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 1.05 ms
 Peer dispersion: 7.81 ms
 Reference time: 14:53:27.371 UTC Sep 19 2005 (C6D94F67.5EF9DB22)
```

As shown above, Switch B has been synchronized to Switch A, and the clock stratum level of Switch B is 3, while that of Switch A is 2.

View the NTP session information of Switch B, which shows that an association has been set up Switch B and Switch A.

```
[SwitchB] display ntp-service sessions
      source      reference   stra  reach  poll  now  offset  delay  disper
**************************************************************************
[12345] 1.0.1.11  127.127.1.0    2    63    64    3    -75.5    31.0  16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

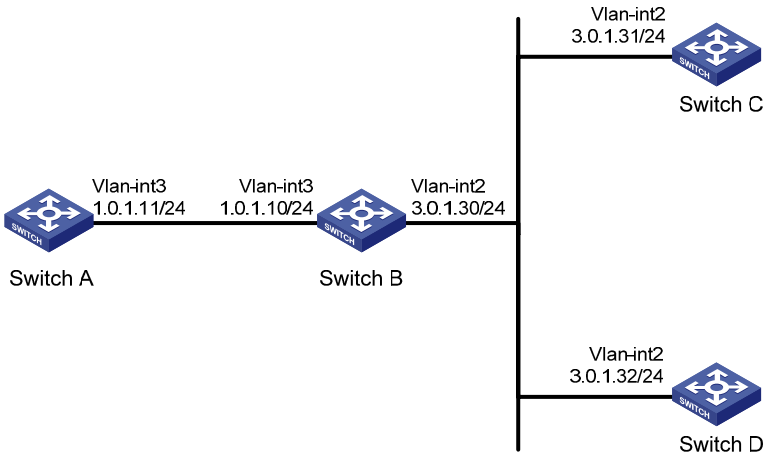
Configuring NTP Broadcast Mode with Authentication

Network requirements

- The local clock of Switch C is to be used as the master clock, with a stratum level of 3.

- Switch C works in the broadcast server mode and sends out broadcast messages from VLAN-interface 2.
- Switch D works in the broadcast client mode and receives broadcast messages through VLAN-interface 2.
- NTP authentication is enabled on both Switch C and Switch D.

Figure 1-12 Network diagram for configuration of NTP broadcast mode with authentication



Configuration procedure

1) Configuration on Switch C:

Configure NTP authentication.

```
<SwitchC> system-view
[SwitchC] ntp-service authentication enable
[SwitchC] ntp-service authentication-keyid 88 authentication-mode md5 123456
[SwitchC] ntp-service reliable authentication-keyid 88
```

Specify Switch C as an NTP broadcast server, and specify an authentication key.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server authentication-keyid 88
```

2) Configuration on Switch D:

Configure NTP authentication.

```
<SwitchD> system-view
[SwitchD] ntp-service authentication enable
[SwitchD] ntp-service authentication-keyid 88 authentication-mode md5 123456
[SwitchD] ntp-service reliable authentication-keyid 88
```

Configure Switch D to work in the NTP broadcast client mode.

```
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service broadcast-client
```

Now, Switch D can receive broadcast messages through VLAN-interface 2, and Switch C can send broadcast messages through VLAN-interface 2. Upon receiving a broadcast message from Switch C, Switch D synchronizes its clock to that of Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD-Vlan-interface2] display ntp-service status
```

```
 Clock status: synchronized
 Clock stratum: 4
 Reference clock ID: 3.0.1.31
 Nominal frequency: 64.0000 Hz
 Actual frequency: 64.0000 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 31.00 ms
 Root dispersion: 8.31 ms
 Peer dispersion: 34.30 ms
 Reference time: 16:01:51.713 UTC Sep 19 2005 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch C, and the clock stratum level of Switch D is 4, while that of Switch C is 3.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD-Vlan-interface2] display ntp-service sessions
      source    reference    stra  reach  poll  now  offset   delay  disper
**************************************************************************
[1234] 3.0.1.31  127.127.1.0   3   254     64    62   -16.0    32.0   16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

Table of Contents

1 VRRP Configuration

When configuring VRRP, go to these sections for information you are interested in:

Note

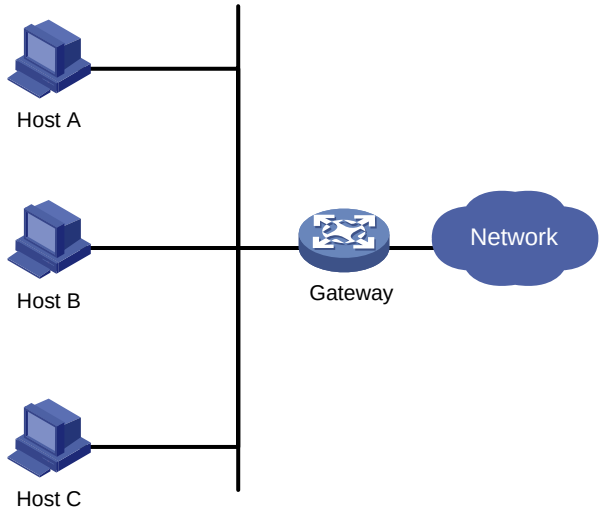
- The term router in this document refers to a router in a generic sense or a Layer 3 switch.
- At present, the interfaces that VRRP involves can only be VLAN interfaces unless otherwise specified.

Introduction to VRRP

VRRP Overview

Normally, as shown in Figure 1-1, you can configure a default route with the gateway as the next hop for every host on a network segment. All packets destined to other network segments are sent over the default route to the gateway and then be forwarded by the gateway. However, when the gateway fails, all the hosts using the gateway as the default next-hop router fail to communicate with the external network.

Figure 1-1 LAN networking



Configuring a default route for network hosts facilitates your configuration, but also requires high performance stability of the device acting as the gateway. Using more egress gateways is a common way to improve system reliability, introducing the problem of routing among the multiple egresses.

Virtual Router Redundancy Protocol (VRRP) is designed to address this problem. VRRP adds routers that can act as network gateways to a VRRP group, which forms a virtual router. Routers in the VRRP group elect a master through the VRRP election mechanism to take the responsibility of a gateway, and hosts on a LAN only need to configure the virtual router as their default network gateway.

VRRP is an error-tolerant protocol, which improves the network reliability and simplifies configurations on hosts. Deploying VRRP on multicast and broadcast LANs such as Ethernet, you can ensure that the system can still provide highly reliable default links without changing configurations (such as dynamic routing protocols, route discovery protocols) when a device fails, and prevent network interruption due to failure of a single link.

VRRP has two versions: VRRPv2 and VRRPv3. VRRPv2 is based on IPv4, and VRRPv3 is based on IPv6. The two versions implement the same functions but provide different commands.

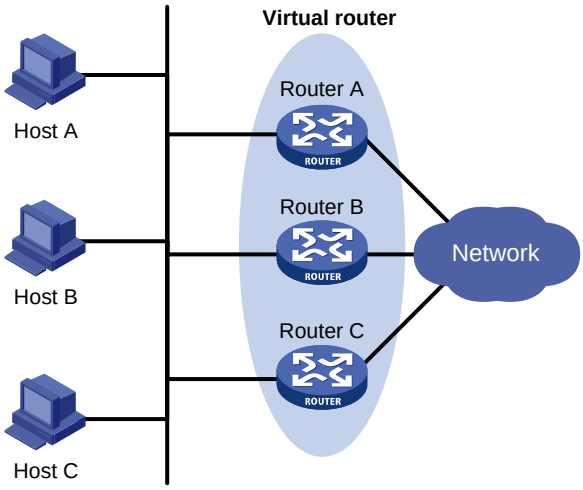
VRRP Group Overview

VRRP combines a group of routers (including a master and multiple backups) on a LAN into a virtual router called VRRP group.

A VRRP group has the following features:

- A virtual router has an IP address. A host on the LAN only needs to know the IP address of the virtual router and uses the IP address as the next hop of the default route.
- Every host on the LAN communicates with external networks through the virtual router.
- Routers in the VRRP group elect the gateway according to their priorities. When the master acting as the gateway fails, to ensure that the hosts in the network segment can communicate with the external networks uninterruptedly, the other routers in the VRRP group elect a new gateway to undertake the responsibility of the failed router.

Figure 1-2 Network diagram for VRRP



As shown in Figure 1-2, Router A, Router B, and Router C form a virtual router, which has its own IP address. Hosts on the Ethernet use the virtual router as the default gateway.

The router with the highest priority of the three routers is elected as the master to act as the gateway, and the other two are backups.

Caution

- The IP address of the virtual router can be either an unused IP address on the segment where the VRRP group resides or the IP address of an interface on a router in the VRRP group. In the latter case, the router is called the IP address owner.
- In a VRRP group, you can configure only one IP address owner.

VRRP priority

VRRP determines the role (master or backup) of each router in the VRRP group by priority. A router with a higher priority has more opportunity to become the master.

VRRP priority is in the range of 0 to 255. A bigger number means a higher priority. Priorities 1 to 254 are configurable. Priority 0 is reserved for special uses and priority 255 for the IP address owner. When a router acts as the IP address owner, its running priority is always 255. That is, the IP address owner in a VRRP group acts as the master as long as it works properly.

Working mode

A router in a VRRP group works in one of the following two modes:

- Non-preemptive mode

When a router in the VRRP group becomes the master, it stays as the master as long as it operates normally, even if a backup is assigned a higher priority later.

- Preemptive mode

When a backup finds its priority higher than that of the master, the backup sends VRRP advertisements to start a new master election in the VRRP group and becomes the master. Accordingly, the original master becomes a backup..

Authentication mode

VRRP provides two authentication modes:

- **simple**: Simple text authentication

You can adopt the simple text authentication mode in a network facing possible security problems. A router sending a packet fills an authentication key into the packet, and the router receiving the packet compares its local authentication key with that of the received packet. If the two authentication keys are the same, the received VRRP packet is considered real and valid; otherwise, the received packet is considered invalid.

- **md5**: MD5 authentication

You can adopt MD5 authentication in a network facing severe security problems. The router encrypts a packet to be sent using the authentication key and MD5 algorithm and saves the encrypted packet in the authentication header. The router receiving the packet uses the authentication key to decrypt the packet and checks the validity of the packet.

On a secure network, you do not need to set the authentication mode.

VRRP Timers

VRRP timers include VRRP advertisement interval timer and VRRP preemption delay timer.

VRRP advertisement interval timer

The master in a VRRP group sends VRRP advertisements periodically to inform the other routers in the VRRP group that it operates properly.

You can adjust the interval for sending VRRP advertisements by setting the VRRP advertisement interval timer. If a backup receives no advertisements in a period three times the interval, the backup regards itself as the master and sends VRRP advertisements to start a new master election.

VRRP preemption delay timer

In an unstable network, a backup can fail to receive the packets from the master due to network congestion and thus the members in the group change their states frequently. Set the VRRP preemption delay timer to address the problem.

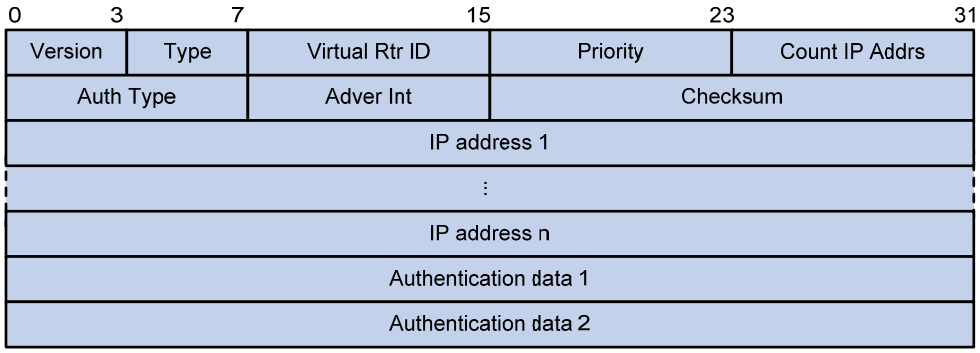
With the VRRP preemption delay timer set, if a backup receives no advertisement in a period three times the advertisement interval and then the preemption delay, it considers that the master fails. In this case, the backup regards itself as the master and sends VRRP advertisements to start a new master election in the VRRP group.

Format of VRRP Packets

VRRP uses multicast packets. The router acting as the master sends VRRP packets periodically to declare its existence. VRRP packets are also used for checking the parameters of the virtual router and electing the master.

IPv4-based VRRP packet format

Figure 1-3 Format of IPv4-based VRRP packet



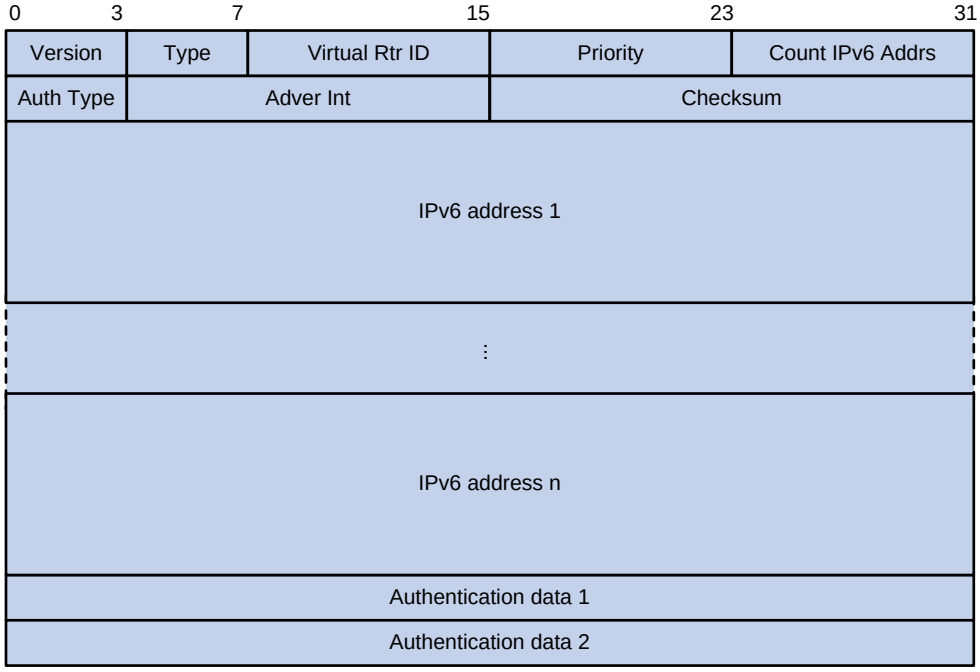
As shown in Figure 1-3, an IPv4-based VRRP packet consists of the following fields:

- Version: Version number of the protocol, 2 for VRRPv2.
- Type: Type of the VRRP packet. Only one VRRP packet type is present, that is, VRRP advertisement, which is represented by 1.
- Virtual Rtr ID (VRID): Serial number of the virtual router, that is, serial number of the VRRP group. It ranges from 1 to 255.
- Priority: Priority of the router in the VRRP group, in the range 0 to 255. A greater value represents a higher priority.
- Count IP Addrs: Number of virtual IP addresses for the VRRP group. A VRRP group can have multiple virtual IP addresses.
- Auth Type: Authentication type. 0 means no authentication, 1 means simple text authentication, and 2 means MD5 authentication.

- Adver Int: Interval for sending advertisement packets, in seconds. The default is 1.
- Checksum: 16-bit checksum for validating the data in VRRP packets.
- IP Address: Virtual IP address entry of the VRRP group. The Count IP Addrs field defines the number of the virtual IP addresses.
- Authentication Data: Authentication key. Currently, this field is used only for simple authentication and is 0 for any other authentication modes.

IPv6-based VRRP packet format

Figure 1-4 Format of IPv6-based VRRP packet



As shown in Figure 1-4, an IPv6-based VRRP packet consists of the following fields:

- Version: Version number of the protocol, 3 for VRRPv3.
- Type: Type of the VRRP packet. Only one VRRP packet type is present, that is, VRRP advertisement, which is represented by 1.
- Virtual Rtr ID (VRID): Serial number of the virtual router, that is, serial number of the VRRP group. It ranges from 1 to 255.
- Priority: Priority of the router in the VRRP group, in the range 0 to 255. A greater value represents a higher priority.
- Count IPv6 Addrs: Number of virtual IPv6 addresses for the VRRP group. A VRRP group can have multiple virtual IPv6 addresses.
- Auth Type: Authentication type. 0 means no authentication, and 1 means simple authentication. VRRPv3 does not support MD5 authentication.
- Adver Int: Interval for sending advertisement packets, in centiseconds. The default is 100.
- Checksum: 16-bit checksum for validating the data in VRRPv3 packets.
- IPv6 Address: Virtual IPv6 address entry of the VRRP group. The Count IPv6 Addrs field defines the number of the virtual IPv6 addresses.
- Authentication Data: Authentication key. Currently, this field is used only for simple authentication and is 0 for any other authentication modes.

Principles of VRRP

- With VRRP enabled, the routers decide their respective roles in the VRRP group by priority. The router with the highest priority becomes the master, and the others are the backups. The master sends VRRP advertisements periodically to notify the backups that it is working properly, and each of the backups starts a timer to wait for advertisements from the master.
- In preemptive mode, when a backup receives a VRRP advertisement, it compares the priority in the packet with that of its own. If the priority of the backup is higher, the backup becomes the master; otherwise, it remains a backup.
- In non-preemptive mode, the router in the VRRP group remains as a master or backup as long as the master does not fail. The backup does not become the master even if the backup is configured with a higher priority.
- If the timer of a backup expires but the backup still does not receive any VRRP advertisement, it considers that the master fails. In this case, the backup considers itself as the master and sends VRRP advertisements to start a new master election.

VRRP Tracking

Tracking a specified interface

The interface tracking function expands the backup functionality of VRRP. It provides backup not only when the interface to which a VRRP group is assigned fails but also when other interfaces (such as uplink interfaces) on the router become unavailable.

If the uplink interface of a router in a VRRP group fails, normally the VRRP group cannot be aware of the uplink failure. If the router is the master of the VRRP group, hosts on the LAN are not able to access the external network because of the uplink failure. You can solve the problem through the function of tracing a specified interface. In this case, it is the uplink interface. After you configure to monitor the uplink interface, when the uplink interface goes down, the priority of the master is automatically decreased by a specified value and a higher priority router in the VRRP group becomes the master.

Tracking a Track object

By monitoring a Track object, you can:

- Monitor the upper link. If there is a fault on the upper link, hosts in the LAN cannot access the external network through the router. In this case, the state of the monitored Track object changes to negative and the priority of the router decreases by a specified value. After that, a higher priority router in the VRRP group becomes the master to maintain the proper communication between the hosts in the LAN and the external network.
- Monitor the master on a backup. If there is a fault on the master, the backup working in the mode switches to the master immediately to maintain normal communication.

Note

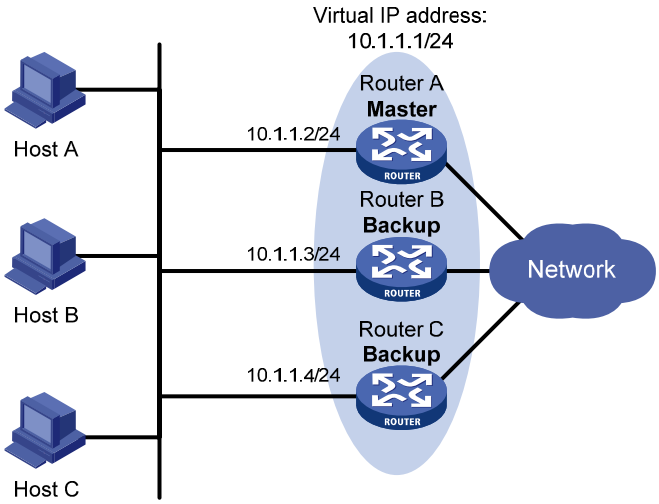
For details of Track object tracking, refer to *Track Configuration* in the *System Volume*.

VRRP Application (Taking IPv4-Based VRRP for Example)

Master/backup

In master/backup mode, only one router, the master, provides services. When the master fails, a new master is elected from the original backups. This mode requires only one VRRP group, in which each router holds a different priority and the one with the highest priority becomes the master, as shown in Figure 1-5.

Figure 1-5 VRRP in master/backup mode



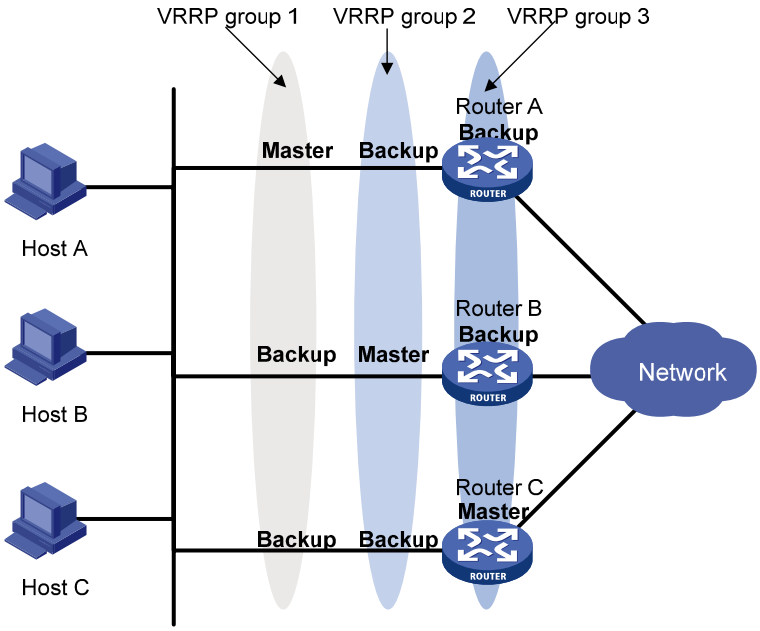
At the beginning, Router A is the master and therefore can forward packets to external networks, whereas Router B and Router C are backups and are thus in the state of listening. If Router A fails, Router B and Router C elect for a new master. The new master takes over the forwarding task to provide services to hosts on the LAN.

Load balancing

You can create more than one VRRP group on an interface of a router, and allow the router to be the master of one VRRP group but a backup of another at the same time.

In load balancing mode, multiple routers provide services at the same time. This mode requires two or more VRRP groups, each of which includes a master and one or more backups. The masters of the VRRP groups can be assumed by different routers, as shown in Figure 1-6.

Figure 1-6 VRRP in load balancing mode



A router can be in multiple VRRP groups and hold a different priority in different group.

In Figure 1-6, three VRRP groups are present:

- VRRP group 1: Router A is the master; Router B and Router C are the backups.
- VRRP group 2: Router B is the master; Router A and Router C are the backups.
- VRRP group 3: Router C is the master; Router A and Router B are the backups.

For load balancing among Router A, Router B, and Router C, hosts on the LAN need to be configured to use VRRP group 1, 2, and 3 as the default gateways respectively. When configuring VRRP priorities, make sure that each router holds such a priority in each VRRP group that it will take the expected role in the group.

Configuring VRRP for IPv4

VRRP for IPv4 Configuration Task List

Complete these tasks to configure VRRP for IPv4:

Task	Remarks
Configuring the Association Between Virtual IP Address and MAC Address	Optional
Creating VRRP Group and Configuring Virtual IP Address	Required
Configuring Router Priority, Preemptive Mode and Tracking Function	Optional
Configuring VRRP Packet Attributes	Optional
Enabling the Trap Function of VRRP	Optional

Configuring the Association Between Virtual IP Address and MAC Address

After the virtual IP address of a VRRP group is associated with a MAC address, the master takes the configured MAC address as the source MAC address of the packets to be sent, so that the hosts in the

internal network can learn the association between the IP address and the MAC address and thus forward the packets to be forwarded to the other network segments to the master.

There are two types of association between virtual IP address and MAC address:

- Virtual IP address is associated with virtual router MAC address

By default, a MAC address is created for a VRRP group after the VRRP group is created, and the virtual IP address is associated with the virtual MAC address. With such association adopted, the hosts in the internal network do not need to update the association between IP address and MAC address when the master changes.

- Virtual IP address is associated with real MAC address of the interface

If an IP address owner exists in a VRRP group and you associate the virtual IP address with the virtual MAC address, two MAC addresses are associated with an IP address. In this case, you can associate the virtual IP address of the VRRP group with the real MAC address, so that the packets from a host are forwarded to the IP address owner according the real MAC address.

Follow these steps to configure the association between MAC address and virtual IP address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the association between virtual IP address and MAC address	**vrrp method** { **real-mac** | **virtual-mac** }	Optional The virtual MAC address is associated with the virtual IP address by default.

Caution

You should configure this function before creating a VRRP group. Otherwise, you cannot modify the mapping between the virtual IP address and the MAC address.

Creating VRRP Group and Configuring Virtual IP Address

You need to configure a virtual IP address for a VRRP group when creating the VRRP group on an interface. If the interface connects to multiple sub-networks, you can configure multiple virtual IP addresses for the VRRP group to realize router backup on different sub-networks.

A VRRP group is created automatically when you specify the first virtual IP address for the VRRP group. If you specify another virtual IP address for the VRRP group later, the virtual IP address is added to the virtual IP address list of the VRRP group.

Caution

It is not recommended to create VRRP groups on the VLAN interface of a super VLAN. Otherwise, network performance may be affected.

Configuration prerequisites

Before creating a VRRP group and configuring a virtual IP address on an interface, you should first configure an IP address for the interface and ensure that the virtual IP address to be configured is in the same network segment as the IP address of the interface.

Configuration procedure

Follow these steps to create VRRP group and configure virtual IP address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the specified interface view	**interface** *interface-type interface-number*	—
Create VRRP group and configure virtual IP address of the VRRP group	**vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*	Required VRRP group is not created by default.

Caution

- For the 3com Switches 4800G, the maximum number of VRRP groups on a switch is 32; and the maximum number of virtual IP addresses for a VRRP group is 5.
- A VRRP group is removed after you remove all the virtual IP addresses in it. In addition, configurations on that VRRP group no longer take effect.
- The virtual IP address of the virtual router can be either an unused IP address on the segment where the VRRP group resides or the IP address of an interface on a router in the VRRP group. In the latter case, the router is called the IP address owner.
- Removal of the VRRP group on the IP address owner will cause IP address collision. In such a case, it is recommended to modify the IP address of the interface on the IP address owner to resolve the collision.
- The virtual IP address of the VRRP group cannot be 0.0.0.0, 255.255.255.255, loopback addresses, non class A/B/C addresses or other illegal IP addresses such as 0.0.0.1.
- Only when the configured virtual IP address and the interface IP address belong to the same segment and are legal host addresses can the VRRP group operate normally. If the configured virtual IP address and the interface IP address do not belong to the same network segment, or the configured IP address is the network address or network broadcast address of the network segment that the interface IP address belongs to, the state of the VRRP group is always **initialize** though you can perform the configuration successfully, that is, VRRP does not take effect in this case.

Configuring Router Priority, Preemptive Mode and Tracking Function

Configuration prerequisites

Before you configure these features, you should first create a VRRP group on the interface and configure a virtual IP address for it.

Configuration procedure

By configuring router priority, preemptive mode, interface tracking, or a Track object, you can decide which router in the VRRP group serves as the Master.

Follow these steps to configure router priority, preemptive mode and the Track object tracking function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter interface view	**interface** *interface-type interface-number*	—
Configure router priority in the VRRP group	**vrrp vrid** *virtual-router-id* **priority** *priority-value*	Optional 100 by default.
Configure the router in the VRRP group to work in preemptive mode and configure preemption delay	**vrrp vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	Optional The router in the VRRP group works in preemptive mode and the preemption delay is 0 seconds by default. If the router in the VRRP group works in non preemptive mode, the preemption delay changes to zero seconds automatically.
Configure the interface to be tracked	**vrrp vrid** *virtual-router-id* **track interface** *interface-type interface-number* [**reduced** *priority-reduced*]	Optional No interface is being tracked by default.
Configure VRRP to track a specified Track object	**vrrp vrid** *virtual-router-id* **track** *track-entry-number* [**reduced** *priority-reduced* | **switchover**]	Optional Not configured by default.

Caution

- The running priority of an IP address owner is always 255 and you do not need to configure it. An IP address owner always works in the preemptive mode.
- Do not configure VRRP tracking of an interface or an object on an IP address owner.
- If the state of the interface under tracking changes from down to up, the priority of the device corresponding to the interface is restored automatically.
- If the state of a Track object changes from negative to positive, the priority of the device corresponding to the Track object is restored automatically.

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create a VRRP group and configure a virtual IP address.

Configuration procedure

Follow these steps to configure VRRP packet attributes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the specified interface view	**interface** *interface-type interface-number*	—
Configure the authentication mode and authentication key when the VRRP groups send and receive VRRP packets	**vrrp vrid** *virtual-router-id* **authentication-mode** { **md5** | **simple** } *key*	Optional Authentication is not performed by default
Configure the time interval for the Master in the VRRP group to send VRRP advertisement	**vrrp vrid** *virtual-router-id* **timer advertise** *adver-interval*	Optional 1 second by default
Disable TTL check on VRRP packets	**vrrp un-check ttl**	Optional Enabled by default You do not need to create a VRRP group before executing this command.

Note

- You may configure different authentication modes and authentication keys for the VRRP groups on an interface. However, the members of the same VRRP group must use the same authentication mode and authentication key.
- Excessive traffic or different timer setting on routers can cause the Backup timer to time out abnormally and trigger a change of the state. To solve this problem, you can prolong the time interval to send VRRP packets and configure a preemption delay.

Enabling the Trap Function of VRRP

After the trap function is enabled for a VRRP module, the VRRP module will generate traps with severity level **errors** to report its key events. The generated traps will be sent to the information center of the device, where you can configure whether to output the trap information and the output destination. For information center configurations, refer to *Information Center Configuration* in the *System Volume*.

Follow these steps to enable the trap function of VRRP:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the trap function of VRRP	**snmp-agent trap enable vrrp** [**authfailure** | **newmaster**]	Optional Enabled by default.

Note

For detailed description on the **snmp-agent trap enable vrrp** command, refer to command **snmp-agent trap enable** in *SNMP Commands* in the *System Volume*.

Displaying and Maintaining VRRP for IPv4

To do…	Use the command…	Remarks
Display VRRP group status	**display vrrp** [**verbose**] [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Display VRRP group statistics	**display vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Clear VRRP group statistics	**reset vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in user view

Configuring VRRP for IPv6

VRRP for IPv6 Configuration Task List

Complete these tasks to configure VRRP for IPv6:

Task	Remarks
Configuring the Association Between Virtual IPv6 Address and MAC Address	Optional
Creating VRRP Group and Configuring Virtual IPv6 Address	Required
Configuring Router Priority, Preemptive Mode and Interface Tracking	Optional
Configuring VRRP Packet Attributes	Optional

Configuring the Association Between Virtual IPv6 Address and MAC Address

After the virtual IPv6 address of a VRRP group is associated with the MAC address, the master takes the configured MAC address as the source MAC address of the packets to be sent, so that the hosts in the internal network can learn the association between the IPv6 address and the MAC address and thus forward the packets to be forwarded to the other network segments to the master.

There are two types of association between virtual IPv6 address and MAC address:

- Virtual IPv6 address is associated with virtual router MAC address

By default, a MAC address is created for a VRRP group after the VRRP group is created, and the virtual IPv6 address is associated with the virtual MAC address. With such association adopted, the hosts in the internal network do not need to update the association between IPv6 address and MAC address when the master changes.

- Virtual IPv6 address is associated with real MAC address of the interface

If an IP address owner exists in a VRRP group and you associate the virtual IPv6 address with the virtual MAC address, two MAC addresses are associated with an IPv6 address. In this case, you can associate the virtual IPv6 address of the VRRP group with the real MAC address, so that the packets from a host is forwarded to the IP address owner according the real MAC address.

Follow these steps to configure the association between MAC address and virtual IPv6 address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the association between virtual IPv6 address and MAC address	**vrrp ipv6 method** { **real-mac** | **virtual-mac** }	Optional The virtual MAC address of the VRRP group is associated with the virtual IPv6 address by default.

Caution

You should configure this function before creating a VRRP group. Otherwise, you cannot modify the mapping between the virtual IPv6 address and the MAC address.

Creating VRRP Group and Configuring Virtual IPv6 Address

You need to configure a virtual IPv6 address for a VRRP group when creating the VRRP group. You can configure multiple virtual IPv6 addresses for a VRRP group.

A VRRP group is created automatically when you specify the first virtual IPv6 address for the VRRP group. If you specify another virtual IPv6 address for the VRRP group later, the virtual IPv6 address is added to the virtual IPv6 address list of the VRRP group.

Caution

It is not recommended to create VRRP groups on the VLAN interface of a super VLAN. Otherwise, network performance may be affected.

Configuration prerequisites

Before creating a VRRP group and configuring a virtual IPv6 address, you should first configure an IPv6 address of the interface and ensure that the virtual IPv6 address to be configured is in the same network segment as the IPv6 address of the interface.

Configuration procedure

Follow these steps to create VRRP group and configure its virtual IPv6 address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the specified interface view	**interface** *interface-type interface-number*	—
Create a VRRP group and configure its virtual IPv6 address	**vrrp ipv6 vrid** *virtual-router-id* **virtual-ip** *virtual-address* [**link-local**]	Required No VRRP group is created by default. The first virtual IPv6 address of the VRRP group must be a link local address. Only one link local address is allowed in a VRRP group, and must be removed the last.

Caution

- For the 3com Switches 4800G, the maximum number of IPv6 VRRP groups on a switch is 32, and the maximum number of virtual IPv6 addresses for a VRRP group is 5.
- A VRRP group is removed after you remove all the virtual IPv6 addresses in it. In addition, configurations on that VRRP group no longer take effect.
- Removal of the VRRP group on the IP address owner will cause IP address collision. In such a case, it is recommended to modify the IPv6 address of the interface on the IP address owner to resolve the collision.

Configuring Router Priority, Preemptive Mode and Interface Tracking

Configuration prerequisites

Before configuring these features, you should first create a VRRP group and configure a virtual IPv6 address.

Configuration procedure

By configuring router priority, preemptive mode and interface tracking, you can decide which router in the VRRP group serves as the Master.

Follow these steps to configure router priority, preemptive mode and interface tracking:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the specified interface view	**interface** *interface-type interface-number*	—
Configure the priority of the router in the VRRP group	**vrrp ipv6 vrid** *virtual-router-id* **priority** *priority-value*	Optional 100 by default

To do…	Use the command…	Remarks
Configure the router in the VRRP group to work in preemptive mode and configure preemption delay of the VRRP group	**vrrp ipv6 vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	Optional The router in the VRRP group works in preemptive mode and the preemption delay is zero seconds by default.
Configure the interface to be tracked	**vrrp ipv6 vrid** *virtual-router-id* **track interface** *interface-type interface-number* [**reduced** *priority-reduced*]	Optional No interface is being tracked by default.

Caution

- The running priority of an IP address owner is always 255 and you do not need to configure it. An IP address owner always works in the preemptive mode.
- Interface tracking is not configurable on an IP address owner.
- If the state of the interface under tracking changes from down to up, the priority of the device corresponding to the interface is restored automatically.

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create a VRRP group and configure a virtual IPv6 address.

Configuration procedure

Follow these steps to configure VRRP packet attributes:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter the specified interface view	**interface** *interface-type interface-number*	—
Configure the authentication mode and authentication key when the VRRP groups send or receive VRRP packets	**vrrp ipv6 vrid** *virtual-router-id* **authentication-mode simple** *key*	Optional Authentication is not performed by default
Configure the time interval for the Master in the VRRP group to send VRRP advertisement	**vrrp ipv6 vrid** *virtual-router-id* **timer advertise** *adver-interval*	Optional 100 centiseconds by default

You may configure different authentication modes and authentication keys for the VRRP groups on an interface. However, the members of the same VRRP group must use the same authentication mode and authentication key.

Excessive traffic or different timer setting on routers can cause the Backup timer to time out abnormally and change the state. To solve this problem, you can prolong the time interval to send VRRP packets and configure a preemption delay.

Displaying and Maintaining VRRP for IPv6

To do…	Use the command…	Remarks
Display VRRP group status	**display vrrp ipv6** [**verbose**] [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Display VRRP group statistics	**display vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Clear VRRP group statistics	**reset vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in user view

IPv4-Based VRRP Configuration Examples

This section provides these configuration examples:

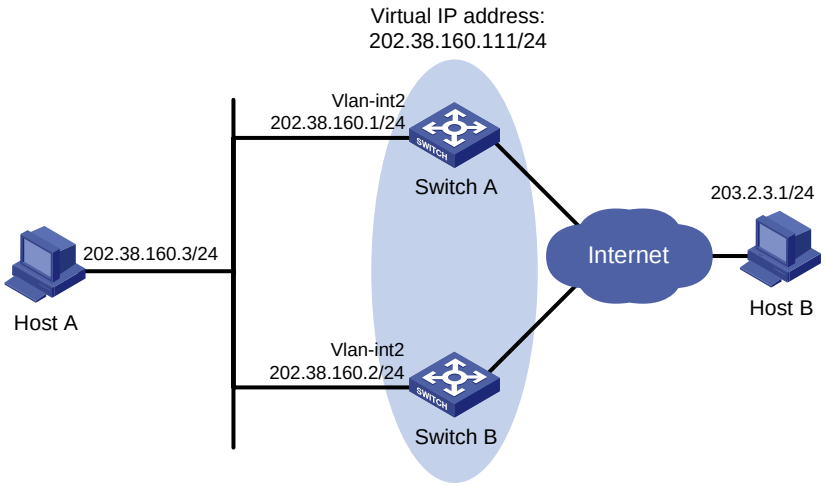
- Single VRRP Group Configuration Example
- VRRP Interface Tracking Configuration Example
- Multiple VRRP Group Configuration Example

Single VRRP Group Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 202.38.160.111/24 as its default gateway.
- Switch A and Switch B belong to VRRP group 1 with the virtual IP address of 202.38.160.111/24.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A fails, packets sent from Host A to Host B are forwarded by Switch B.

Figure 1-7 Network diagram for single VRRP group configuration



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
```

```
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

Create VRRP group 1 and set its virtual IP address to be 202.38.160.111.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Set Switch A to work in preemptive mode. The preemption delay is five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-Vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

Create VRRP group 1 and set its virtual IP address to be 202.38.160.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Set Switch B to work in preemptive mode. The preemption delay is five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

3) Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp verbose** command to verify the configuration.

Display detailed information of VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 1
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : NONE
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1
```

Display detailed information of VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
```

```
Total number of virtual routers: 1
Interface       : Vlan-interface2
VRID            : 1                   Adver. Timer    : 1
Admin Status    : UP                  State           : Backup
Config Pri      : 100                 Run Pri         : 100
Preempt Mode    : YES                 Delay Time      : 5
Auth Type       : NONE
Virtual IP      : 202.38.160.111
Master IP       : 202.38.160.1
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A fails, you can still ping through Host B on Host A. Use the **display vrrp verbose** command to view the detailed information of the VRRP group on Switch B.

If Switch A fails, the detailed information of VRRP group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 1
 Admin Status    : UP                  State           : Master
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : NONE
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.2
```

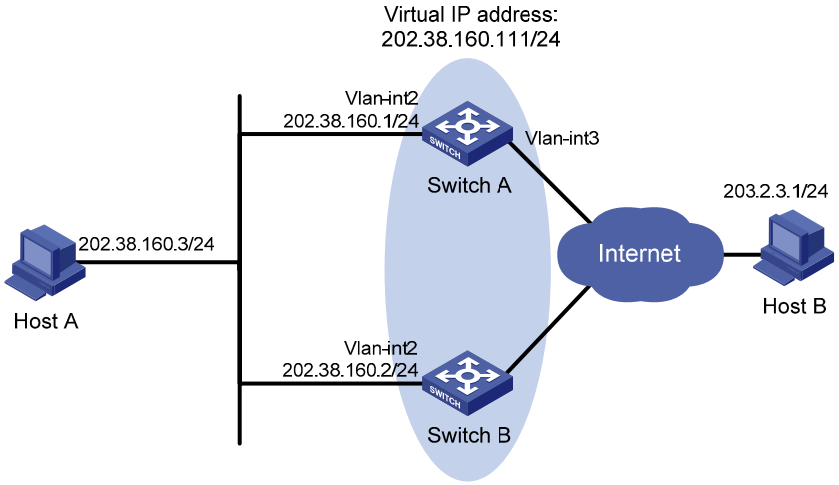
The above information indicates that if Switch A fails, Switch B becomes the master, and packets sent from Host A to Host B are forwarded by Switch B.

VRRP Interface Tracking Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 202.38.160.111/24 as its default gateway.
- Switch A and Switch B belong to VRRP group 1 with the virtual IP address of 202.38.160.111/24.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if VLAN-interface 3 through which Switch A connects to the Internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Figure 1-8 Network diagram for VRRP interface tracking



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

Create a VRRP group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Configure the priority of Switch A in the VRRP group to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Configure the authentication mode of the VRRP group as **simple** and authentication key as **hello**.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Set the interval for Master to send VRRP advertisement to five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

Set the interface to be tracked.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track interface vlan-interface 3 reduced 30
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

Create a VRRP group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Configure the authentication mode of the VRRP group as **simple** and authentication key as **hello**.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Set the interval for Master to send VRRP advertisement to five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

3) Verify the configuration

After the configuration, Host B can be pinged successfully on Host A. You can use the **display vrrp verbose** command to verify the configuration.

Display detailed information of VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                    Adver. Timer    : 5
 Admin Status    : UP                   State           : Master
 Config Pri      : 110                  Run Pri         : 110
 Preempt Mode    : YES                  Delay Time      : 0
 Auth Type       : SIMPLE TEXT          Key             : hello
 Track IF        : Vlan3                Pri Reduced     : 30
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1
```

Display detailed information of VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                    Adver. Timer    : 5
 Admin Status    : UP                   State           : Backup
 Config Pri      : 100                  Run Pri         : 100
 Preempt Mode    : YES                  Delay Time      : 0
 Auth Type       : SIMPLE TEXT          Key             : hello
 Virtual IP      : 202.38.160.111
 Master IP       : 202.38.160.1
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If interface VLAN-interface 3 through which Switch A connects to the Internet is not available, you can still ping Host B successfully on Host A. Use the **display vrrp verbose** command to view the detailed information of the VRRP group.

If VLAN-interface 3 on Switch A is not available, the detailed information of VRRP group 1 on Switch A is displayed.

```
[SwitchA-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
```

```
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Backup
 Config Pri      : 110                 Run Pri         : 80
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : SIMPLE TEXT         Key             : hello
 Track IF        : Vlan3               Pri Reduced     : 30
 Virtual IP      : 202.38.160.111
 Master IP       : 202.38.160.2
```

If VLAN-interface 3 on Switch A is not available, the detailed information of VRRP group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 5
 Admin Status    : UP                  State           : Master
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : SIMPLE TEXT         Key             : hello
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.2
```

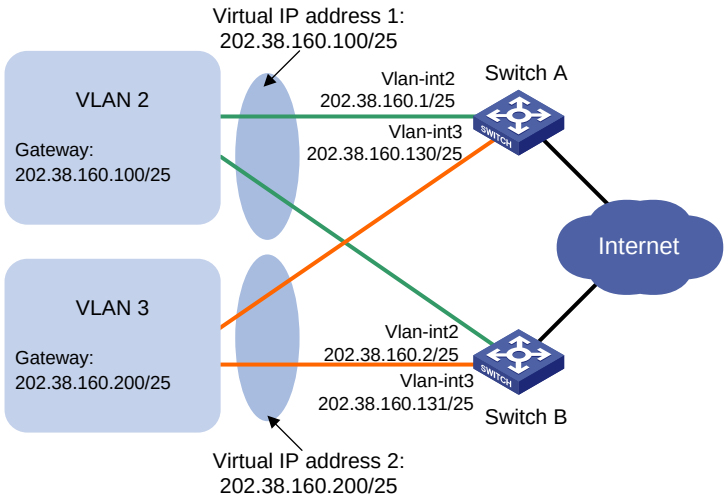
The above information indicates that if VLAN-interface 3 on Switch A is not available, the priority of Switch A is reduced to 80 and it becomes the backup. Switch B becomes the master and packets sent from Host A to Host B are forwarded by Switch B.

Multiple VRRP Group Configuration Example

Network requirements

- Hosts in VLAN 2 use 202.38.160.100/25 as their default gateway and hosts in VLAN 3 use 202.38.160.200/25 as their default gateway.
- Switch A and Switch B belong to both VRRP group 1 and VRRP group 2. The virtual IP address of VRRP group 1 is 202.38.160.100/25, and that of VRRP group 2 is 202.38.160.200/25.
- In VRRP group 1, Switch A has a higher priority than Switch B. In VRRP group 2, Switch B has a higher priority than Switch A. In this case, hosts in VLAN 2 and VLAN 3 can communicate with the outside through Switch A and Switch B respectively, and if Switch A or Switch B fails, the hosts can use the other switch to communicate with the outside, so as to avoid communication interruption.

Figure 1-9 Network diagram for multiple VRRP group configuration



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.128
```

Create a VRRP group 1 and set its virtual IP address to 202.38.160.100.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.100
```

Configure the priority of Switch A in VRRP group 1 as 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
[SwitchA-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchA] vlan 3
[SwitchA-vlan3] port gigabitethernet 1/0/6
[SwitchA-vlan3] quit
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ip address 202.38.160.130 255.255.255.128
```

Create a VRRP group 2 and set its virtual IP address to 202.38.160.200.

```
[SwitchA-Vlan-interface3] vrrp vrid 2 virtual-ip 202.38.160.200
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
```

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.128
```

Create a VRRP group 1 and set its virtual IP address to 202.38.160.100.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.100
[SwitchB-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/6
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ip address 202.38.160.131 255.255.255.128
```

Create a VRRP group 2 and set its virtual IP address to 202.38.160.200.

```
[SwitchB-Vlan-interface3] vrrp vrid 2 virtual-ip 202.38.160.200
```

Configure the priority of Switch B in VRRP group 2 to 110.

```
[SwitchB-Vlan-interface3] vrrp vrid 2 priority 110
```

3) Verify the configuration

You can use the **display vrrp verbose** command to verify the configuration.

Display detailed information of the VRRP group on Switch A.

```
[SwitchA-Vlan-interface3] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 2
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 1
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : 202.38.160.100
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1

 Interface       : Vlan-interface3
 VRID            : 2                   Adver. Timer    : 1
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : 202.38.160.200
 Master IP       : 202.38.160.131
```

Display detailed information of the VRRP group on Switch B.

```
[SwitchB-Vlan-interface3] display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 2
```

```
Interface       : Vlan-interface2
VRID            : 1                     Adver. Timer    : 1
Admin Status    : UP                    State           : Backup
Config Pri      : 100                   Run Pri         : 100
Preempt Mode    : YES                   Delay Time      : 0
Auth Type       : NONE
Virtual IP      : 202.38.160.100
Master IP       : 202.38.160.1

Interface       : Vlan-interface3
VRID            : 2                     Adver. Timer    : 1
Admin Status    : UP                    State           : Master
Config Pri      : 110                   Run Pri         : 110
Preempt Mode    : YES                   Delay Time      : 0
Auth Type       : NONE
Virtual IP      : 202.38.160.200
Virtual MAC     : 0000-5e00-0102
Master IP       : 202.38.160.131
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and hosts with the default gateway of 202.38.160.100/25 accesses the Internet through Switch A; in VRRP group 2 Switch A is the backup, Switch B is the master and hosts with the default gateway of 202.38.160.200/25 accesses the Internet through Switch B.

IPv6-Based VRRP Configuration Examples

This section provides these configuration examples:

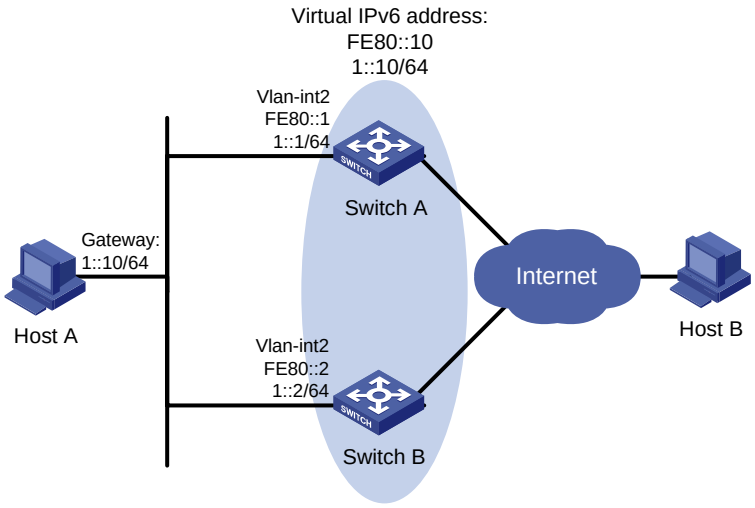
- Single VRRP Group Configuration Example
- VRRP Interface Tracking Configuration Example
- Multiple VRRP Group Configuration Example

Single VRRP Group Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 1::10/64 as its default gateway.
- Switch A and Switch B belong to VRRP group 1 with the virtual IP addresses of 1::10/64 and FE80::10.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A fails, packets sent from Host A to Host B are forwarded by Switch B.

Figure 1-10 Network diagram for single VRRP group configuration



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create a VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Set Switch A to work in preemptive mode, with the preemption delay set to 5 seconds.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

Enable Switch A to send RA messages.

```
[SwitchA-Vlan-interface2] undo ipv6 nd ra halt
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
```

```
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create a VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Configure Switch B to work in the preemptive mode, with the preemption delay set to 5 seconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

Enable Switch B to send RA messages.

```
[SwitchB-Vlan-interface2] undo ipv6 nd ra halt
```

3) Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp ipv6 verbose** command to verify the configuration.

Display detailed information of VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : NONE
 Virtual IP      : FE80::10
                   1::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1
```

Display detailed information of VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : NONE
 Virtual IP      : FE80::10
                   1::10
 Master IP       : FE80::1
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A fails, you can still ping through Host B on Host A. You can use the **display vrrp ipv6 verbose** command to view the detailed information of the VRRP group on Switch B.

If Switch A fails, the detailed information of VRRP group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                    Adver. Timer    : 100
 Admin Status    : UP                   State           : Master
 Config Pri      : 100                  Run Pri         : 100
 Preempt Mode    : YES                  Delay Time      : 5
 Auth Type       : NONE
 Virtual IP      : FE80::10
                   1::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::2
```

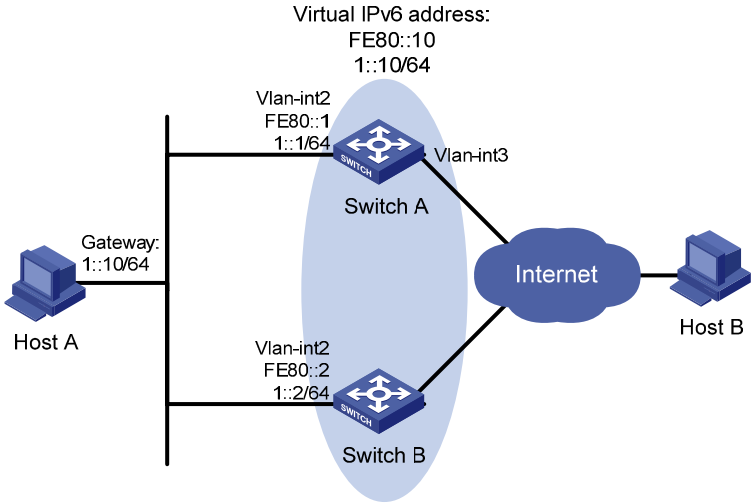
The above information indicates that if Switch A fails, Switch B becomes the master, and packets sent from Host A to Host B are forwarded by Switch B.

VRRP Interface Tracking Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 1::10/64 as its default gateway.
- Switch A and Switch B belong to VRRP group 1 with the virtual IP addresses of 1::10/64 and FE80::10.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if VLAN-interface 3 through which Switch A connects to the Internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Figure 1-11 Network diagram for VRRP interface tracking



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create a VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Set the authentication mode for VRRP group 1 to **simple** and authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simple hello
```

Set the VRRP advertisement interval to 500 centiseconds.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

Set Switch A work in preemptive mode. The preemption delay is five seconds.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

Set the interface to be tracked.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 track interface vlan-interface 3 reduced 30
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create a VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Set the authentication mode for VRRP group 1 to **simple** and authentication key to **hello**.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simple hello
```

Set the VRRP advertisement interval to 500 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

Set Switch B to work in preemptive mode. The preemption delay is five seconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

3) Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp ipv6 verbose** command to verify the configuration.

Display detailed information of VRRP group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 500
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Track IF        : Vlan3               Pri Reduced     : 30
 Virtual IP      : FE80::10
                   1::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1
```

Display detailed information of VRRP group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 500
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Virtual IP      : FE80::10
                   1::10
 Master IP       : FE80::1
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If interface VLAN-interface 3 on Switch A is not available, you can still ping Host B successfully on Host A. You can use the **display vrrp ipv6 verbose** command to view the detailed information of the VRRP group.

If interface VLAN-interface 3 on Switch A is not available, the detailed information of VRRP group 1 on Switch A is displayed.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 500
 Admin Status    : UP                  State           : Backup
```

```
 Config Pri      : 110                 Run Pri         : 80
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Track IF        : Vlan3               Pri Reduced     : 30
 Virtual IP      : FE80::10
                   1::10
 Master IP       : FE80::2
```

If interface VLAN-interface 3 on Switch A is not available, the detailed information of VRRP group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 1
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 500
 Admin Status    : UP                  State           : Master
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 5
 Auth Type       : SIMPLE TEXT         Key             : hello
 Virtual IP      : FE80::10
                   1::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::2
```

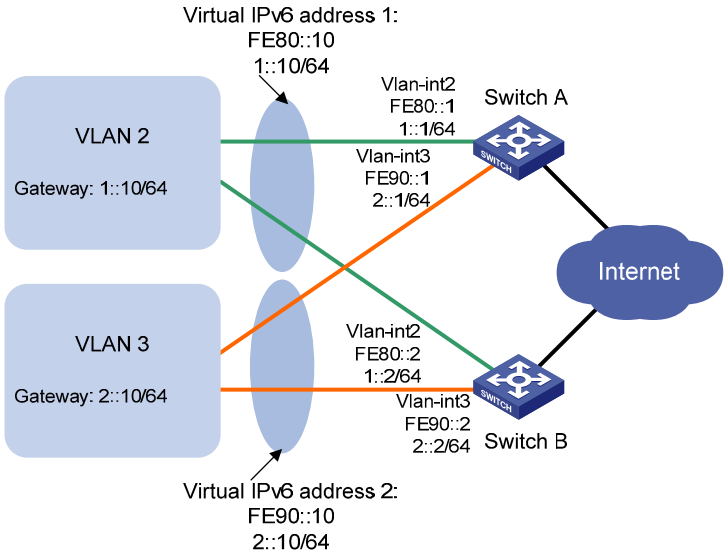
The above information indicates that if VLAN-interface 3 on Switch A is not available, the priority of Switch A is reduced to 80 and Switch A becomes the backup. Switch B becomes the master and packets sent from Host A to Host B are forwarded by Switch B.

Multiple VRRP Group Configuration Example

Network requirements

- Hosts in VLAN 2 use 1::10/64 as their default gateway and hosts in VLAN 3 use 2::10/64 as their default gateway.
- Switch A and Switch B belong to both VRRP group 1 and VRRP group 2. The virtual IPv6 addresses of VRRP group 1 are 1::10/64 and FE80::10, and those of VRRP group 2 are 2::10/64 and FE90::10.
- In VRRP group 1, Switch A has a higher priority than Switch B. In VRRP group 2, Switch B has a higher priority than Switch A. In this case, hosts in VLAN 1 and VLAN can communicate with the outside through Switch A and Switch B respectively, and if Switch A or Switch B fails, the hosts can use the other switch to communicate with the outside, so as to avoid communication interruption.

Figure 1-12 Network diagram for multiple VRRP group configuration



Configuration procedure

1) Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create VRRP group 1 and set its virtual IPv6 addresses to FE80::10 to 1::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
[SwitchA-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchA] vlan 3
[SwitchA-vlan3] port gigabitethernet 1/0/6
[SwitchA-vlan3] quit
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ipv6 address fe90::1 link-local
[SwitchA-Vlan-interface3] ipv6 address 2::1 64
```

Create VRRP group 2 and set its virtual IPv6 addresses to FE90::10 and 2::10.

```
[SwitchA-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip fe90::10 link-local
[SwitchA-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip 2::10
```

2) Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB-vlan2] port gigabitethernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create VRRP group 1 and set its virtual IPv6 addresses to FE80::10 and 1::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip 1::10
[SwitchB-Vlan-interface2] quit
```

Configure VLAN 3.

```
[SwitchB] vlan 3
[SwitchB-vlan3] port gigabitethernet 1/0/6
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] ipv6 address fe90::2 link-local
[SwitchB-Vlan-interface3] ipv6 address 2::2 64
```

Create VRRP group 2 and set its virtual IPv6 addresses to FE90::10 and 2::10.

```
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip fe90::10 link-local
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 virtual-ip 2::10
```

Set the priority of Switch B in VRRP group 2 to 110.

```
[SwitchB-Vlan-interface3] vrrp ipv6 vrid 2 priority 110
```

3) Verify the configuration

You can use the **display vrrp ipv6 verbose** command to verify the configuration.

Display detailed information of the VRRP group on Switch A.

```
[SwitchA-Vlan-interface3] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 2
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : FE80::10
                   1::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1

 Interface       : Vlan-interface3
 VRID            : 2                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Backup
```

```
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : FE90::10
                   2::10
 Master IP       : FE90::2
```

Display detailed information of the VRRP group on Switch B.

```
[SwitchB-Vlan-interface3] display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Total number of virtual routers: 2
 Interface       : Vlan-interface2
 VRID            : 1                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Backup
 Config Pri      : 100                 Run Pri         : 100
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : FE80::10
                   1::10
 Master IP       : FE80::1

 Interface       : Vlan-interface3
 VRID            : 2                   Adver. Timer    : 100
 Admin Status    : UP                  State           : Master
 Config Pri      : 110                 Run Pri         : 110
 Preempt Mode    : YES                 Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : FE90::10
                   2::10
 Virtual MAC     : 0000-5e00-0202
 Master IP       : FE90::2
```

The above information indicates that in VRRP group 1 Switch A is the master, Switch B is the backup and hosts with the default gateway of 1::10/64 accesses the Internet through Switch A; in VRRP group 2 Switch A is the backup, Switch B is the master and hosts with the default gateway of 2::10/64 accesses the Internet through Switch B.

Note

Multiple VRRP groups are commonly used in actual networking. In IPv6 network, to implement load sharing among multiple VRRP groups, you need to manually configure the default gateway for hosts.

Troubleshooting VRRP

Symptom 1:

The console screen displays error prompts frequently.

Analysis:

This error is probably caused by the following:

- Inconsistent configuration of the devices in the VRRP group.
- A device is attempting to send illegitimate VRRP packets.

Solution:

- In the first case, modify the configuration.
- In the latter case, you have to resort to non-technical measures.

Symptom 2:

Multiple masters are present in the same VRRP group.

Analysis:

- Multiple masters coexist for a short period: This is normal and requires no manual intervention.
- Multiple masters coexist for a long period: This is because devices in the VRRP group cannot receive VRRP packets, or the received VRRP packets are illegal.

Solution:

Ping between these masters, and do the following:

- If the ping fails, check network connectivity.
- If the ping succeeds, check that their configurations are consistent in terms of number of virtual IP addresses, virtual IP addresses, advertisement interval, and authentication.

Symptom 3:

Frequent VRRP state transition.

Analysis:

The VRRP advertisement interval is set too short.

Solution:

Increase the interval to sent VRRP advertisement or introduce a preemption delay.

Table of Contents

1 Hotfix Configuration

When configuring hotfix, go to these sections for information you are interested in:

- Hotfix Overview
- Hotfix Configuration Task List
- Displaying and Maintaining Hotfix
- Hotfix Configuration Examples

Hotfix Overview

Hotfix is a fast and cost-effective method to repair software defects of a device. Compared with another method, software version upgrade, hotfix can upgrade the software without interrupting the running services of the device, that is, it can repair the software defects of the current version without rebooting the device.

Basic Concepts in Hotfix

Patch and patch file

A patch, also called patch unit, is a package to fix software defects. Generally, patches are released as patch files A patch file may contain one or more patches for different defects. After loaded from the Flash to the memory area, each patch is assigned a unique number, which starts from 1, for identification, management, and operation. For example, if a patch file has three patch units, they will be numbered as 1, 2, and 3 respectively.

Incremental patch

Patches in a patch file are all incremental patches. An incremental patch means that the patch is dependent on the previous patch units. For example, if a patch file has three patch units, patch 3 can be running only after patch 1 and 2 take effect. You cannot run patch 3 separately.

Common patch and temporary patch

Patches fall into two types, common patches and temporary patches.

- Common patches are those formally released through the version release flow.
- Temporary patches are those not formally released through the version release flow, but temporarily provided to solve the emergent problems.

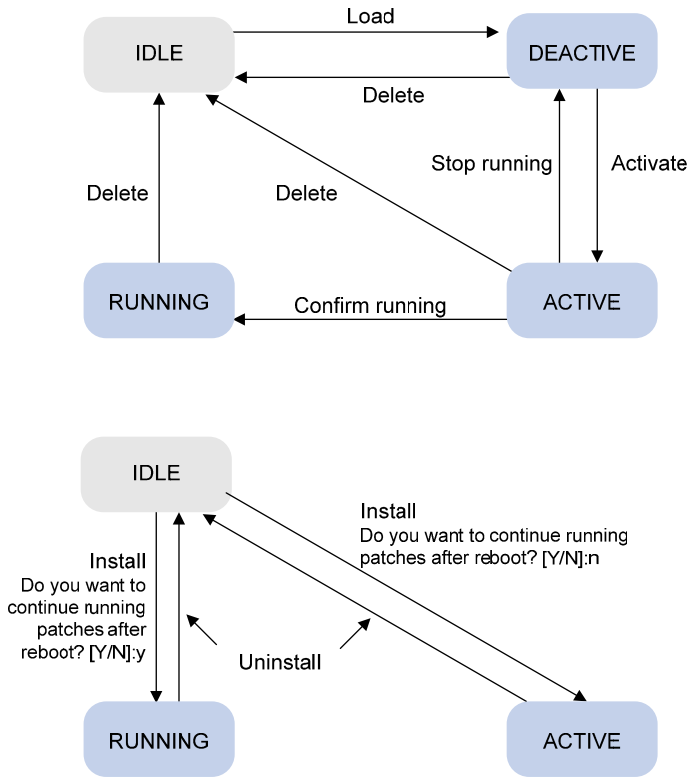
The common patches always include the functions of the previous temporary patches, so as to replace them. The patch type affects the patch loading process only: the system will delete all the temporary patches before it loads the common patch.

Patch Status

Each patch has its status, which can be switched by command lines. The relationship between patch state changes and command actions is shown in Figure 1-1. The patch can be in the state of IDLE, DEACTIVE, ACTIVE, and RUNNING. Load, run temporarily, confirm running, stop running, delete,

install, and uninstall represent operations, corresponding to commands of **patch load**, **patch active**, **patch run**, **patch deactive**, **patch delete**, **patch install**, and **undo patch install**. For example, if you execute the **patch active** command for the patches in the DEACTIVE state, the patches turn to the ACTIVE state.

Figure 1-1 Relationship between patch state changes and command actions



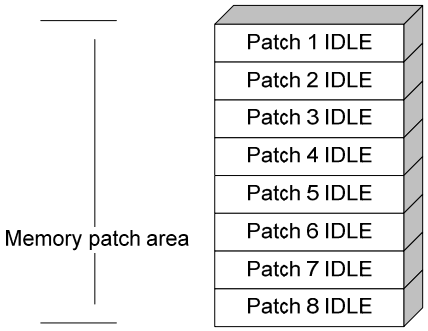
Note

Information about patch states is saved in file **patchstate** on the flash. It is recommended not to operate this file.

IDLE state

Patches in the IDLE state are not loaded. You cannot install or run the patches, as shown in Figure 1-2 (suppose the memory patch area can load up to eight patches). The patches that are in the IDLE state will be still in the IDLE state after system reboot.

Figure 1-2 Patches are not loaded to the memory patch area



Note

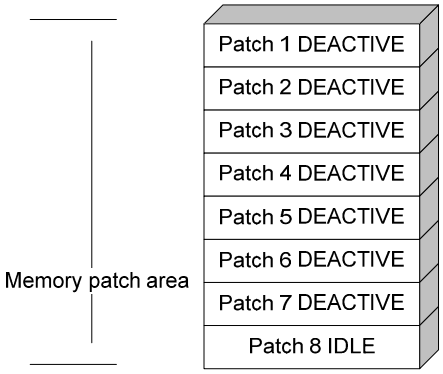
Currently, the system patch area supports up to 200 patches.

DEACTIVE state

Patches in the DEACTIVE state have been loaded to the memory patch area but have not run in the system yet. Suppose that there are seven patches in the patch file to be loaded. After the seven patches successfully pass the version check and CRC check, they will be loaded to the memory patch area and are in the DEACTIVE state. At this time, the patch states in the system are as shown in Figure 1-3.

The patches that are in the DEACTIVE state will be still in the DEACTIVE state after system reboot.

Figure 1-3 A patch file is loaded to the memory patch area

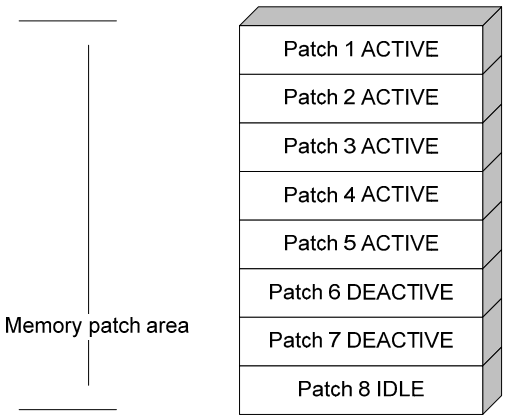


ACTIVE state

Patches in the ACTIVE state are those that have run temporarily in the system and will become DEACTIVE after system reboot. For the seven patches in Figure 1-3, if you activate the first five patches, the state of them will change from DEACTIVE to ACTIVE. At this time, the patch states in the system are as shown in Figure 1-4.

The patches that are in the ACTIVE state will be in the DEACTIVE state after system reboot.

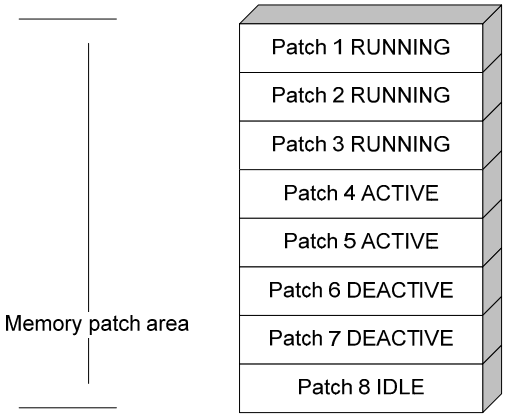
Figure 1-4 Patches are activated



RUNNING state

After you confirm the running of the ACTIVE patches, the state of the patches will become RUNNING and will be in the RUNNING state after system reboot. For the five patches in Figure 1-4, if you confirm the running the first three patches, their states will change from ACTIVE to RUNNING. At this time, the patch states of the system are as shown in Figure 1-5.

Figure 1-5 Patches are running



The patches that are in the RUNNING state will be still in the RUNNING state after system reboot.

Hotfix Configuration Task List

Task		Remarks
Install patches	One-Step Patch Installation	Use either approach. The step-by-step patch installation allows you to control the patch status.
	Step-by-Step Patch Installation	
Uninstall patches	One-Step Patch Uninstallation	Use either approach. The step-by-step patch uninstallation allows you to control the patch status.
	Step-by-Step Patch Uninstallation	

Configuration Prerequisites

Patches are released per device model type. Before patching the system, you need to save the appropriate patch files to the storage media of the device using FTP or TFTP. When saving the patch files, note that:

- The patch files match the device model and software version. If they are not matched, the hotfixing operation will fail.
- Name the patch file properly. Otherwise, the system cannot locate the patch file and the hotfixing operation will fail. The name is in the format of "patch_*PATCH-FLAG suffix*.bin". The PATCH-FLAG is pre-defined and support for the PATCH-FLAG depends on device model. The first three characters of the version item (using the **display patch information** command) represent the PATCH-FLAG suffix. The system searches the root directory of the storage medium (flash by default) for patch files based on the PATCH-FLAG. If there is a match, the system loads patches to or install them on the memory patch area.

Table 1-1 describes the default patch name for device.

Table 1-1 Default patch names for device

Product	PATCH-FLAG	Default patch name
4800G	PATCH-XXX	patch_xxx.bin

Note

The loading and installation are performed on all member devices. Before these operations, save the same patch files to the root directories in the storage media of all member devices.

One-Step Patch Installation

You can use the **patch install** command to install patches in one step. After you execute the command, the system displays the message "Do you want to continue running patches after reboot? [Y/N]:".

- Entering **y** or **Y**: All the specified patches are installed, and turn to the RUNNING state from IDLE. This equals execution of the commands **patch location**, **patch load**, **patch active**, and **patch run**. The patches remain RUNNING after system reboot.
- Entering **n** or **N**: All the specified patches are installed and turn to the ACTIVE state from IDLE. This equals execution of the commands **patch location**, **patch load** and **patch active**. The patches turn to the DEACTIVE state after system reboot.

Follow these steps to install the patches in one step:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Install the patches in one step	**patch install** *patch-location*	Required

Note

- The patch matches the device type and software version.
- The **patch install** command changes the patch file location specified with the **patch location** command to the directory specified by the *patch-location* argument of the **patch install** command.

Step-by-Step Patch Installation

Step-by-Step Patch Installation Task List

Task	Remarks
Configuring the Patch File Location	Optional
Loading a Patch File	Required
Activating Patches	Required
Confirm Running Patches	Optional

Configuring the Patch File Location

Follow these steps to configure the patch file location:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the patch file location	**patch location** *patch-location*	Optional **flash:** by default

Note

- The directory specified by the *patch-location* argument must exist on each member device. If one member device does not have such directory, the system cannot locate the patch file on the member device.
- The **patch install** command changes patch file location specified with the **patch location** command to the directory specified by the *patch-location* argument of the **patch install** command. For example, if you execute the **patch location** *xxx* command and then the **patch install** *yyy* command, the patch file location automatically changes from xxx to yyy.

Loading a Patch File

Loading the right patch files is the basis of other hotfixing operations. The system loads a patch file from the flash by default. It will failed if the system cannot find the patch file on the flash.

Caution

Set the file transfer mode to binary mode before using FTP or TFTP to upload/download patch files to/from the flash of the device. Otherwise, patch file cannot be parsed properly.

Follow the steps below to load a patch file:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Load the patch file on from the storage medium to the specified memory patch area	**patch load slot** *slot-number*	Required

Activating Patches

After you activate a patch, the patch will take effect and is in the test-run stage. After the device is reset or rebooted, the patch becomes invalid.

If you find that an ACTIVE patch is of some problem, you can reboot the device to deactivate the patch, so as to avoid a series of running faults resulting from patch error.

Follow the steps below to activate patches:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Activate the specified patches	**patch active** *patch-number* **slot** *slot-number*	Required

Confirm Running Patches

After you confirm the running of a patch, the patch state becomes RUNNING, and the patch is in the normal running stage. After the device is reset or rebooted, the patch is still valid.

Follow the steps below to confirm the running of patches:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Confirm the running of the specified patches	**patch run** *patch-number* [**slot** *slot-number*]	Required

Note

This operation is applicable to patches in the ACTIVE state only.

One-Step Patch Uninstallation

You can use the **undo patch install** command to uninstall patches from all the member devices. The patches then turn to the IDLE state. This equals the execution of the commands **patch deactive** and **patch delete** on each member device.

Follow these steps to uninstall the patches in one step:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Uninstall the patches	**undo patch install**	Required

Step-by-Step Patch Uninstallation

Step-by-Step Patch Uninstallation Task List

Task	Remarks
Stop Running Patches	Required
Deleting Patches	Required

Stop Running Patches

After you stop running a patch, the patch state becomes DEACTIVE, and the system runs in the way before it is installed with the patch.

Follow the steps below to stop running patches:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Stop running the specified patches	**patch deactive** *patch-number* **slot** *slot-number*	Required

Deleting Patches

Deleting patches only removes the patches from the memory patch area, and does not delete them from the storage medium. The patches turn to IDLE state after this operation. After a patch is deleted, the system runs in the way before it is installed with the patch.

Follow the steps below to delete patches:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Delete the specified patches from the memory patch area	**patch delete** *patch-number* **slot** *slot-number*	Required

Displaying and Maintaining Hotfix

To do…	Use the command…	Remarks
Display the patch information	**display patch information**	Available in any view

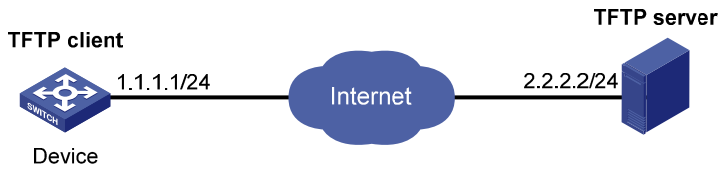
Hotfix Configuration Examples

Hotfix Configuration Example (Single Device)

Network requirements

- The software running on Device is of some problem, and thus hotfixing is needed.
- The patch file **patch_xxx.bin** is saved on the TFTP server.
- The IP address of Device is 1.1.1.1/24, and IP address of TFTP Server is 2.2.2.2/24. An available route exists between Device and TFTP server.

Figure 1-6 Network diagram of hotfix configuration



Configuration procedure

1) Configure TFTP Server. Note that the configuration varies depending on server type and the configuration procedure is omitted.
- Enable the TFTP server function.
- Save the patch file **patch_xxx.bin** to the directory of the TFTP server.

2) Configure Device

Caution

Make sure the free flash space of the device is big enough to store the patch file.

Before upgrading the software, use the **save** command to save the current system configuration. The configuration procedure is omitted.

Load the patch file **patch_xxx.bin** from the TFTP server to the root directory of the device storage medium.

```
<Device> tftp 2.2.2.2 get patch_xxx.bin
```

Install the patch.

```
<Device> system-view
[Device] patch install flash:
Patches will be installed. Continue? [Y/N]:y
```

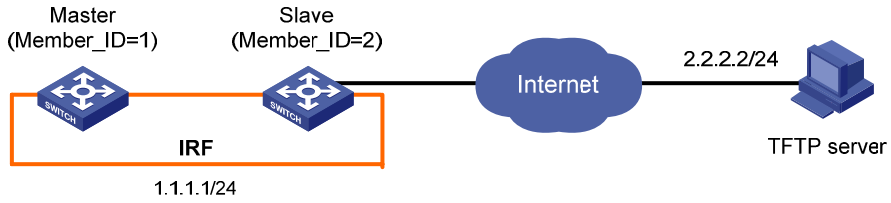
```
Do you want to continue running patches after reboot? [Y/N]:y
Installing patches........
Installation completed, and patches will continue to run after reboot.
```

Hotfix Configuration Example (IRF Stack Device)

Network requirements

- IRF refers to an IRF stack in this example and it consists of two stack devices, Master and Slave. The software running on the stack devices are of some problem, and thus hotfixing is needed.
- The patch file **patch_xxx.bin** is saved on the TFTP server.
- The IP address of IRF is 1.1.1.1/24, and IP address of TFTP Server is 2.2.2.2/24. An available route exists between IRF and TFTP server.

Figure 1-7 Network diagram for hotfix configuration



Note: the orange line indicates the dedicated line for stacking.

Configuration procedure

1) Configure the TFTP server. Note that the configuration varies depending on server type, and the configuration procedure is omitted.
- Enable the TFTP server function.
- Save the patch file **patch_xxx.bin** to the directory of TFTP server.

2) Configure Device.

Caution

Make sure the free flash space of the device is big enough to store the patch files.

Before upgrading the software, use the **save** command to save the current system configuration. The configuration procedure is omitted.

Load the patch file **patch_xxx.bin** from the TFTP server to the root directory of the Master's storage medium.

```
<Device> tftp 2.2.2.2 get patch_xxx.bin
```

Load the patch file **patch_xxx.bin** from the TFTP server to the root directory of the Slave's storage medium.

```
<Device> tftp 2.2.2.2 get patch_xxx.bin slot2#flash:/patch_xxx.bin
```

Install the patch.

```
<Device> system-view
```

```
[Device] patch install flash:
Patches will be installed. Continue? [Y/N]:y
Do you want to continue running patches after reboot? [Y/N]:y
Installing patches........
Installation completed, and patches will continue to run after reboot.
```

Table of Contents

1 Cluster Management Configuration

When configuring cluster management, go to these sections for information you are interested in:

Cluster Management Overview

Cluster Management Definition

A cluster is a group of network devices. Cluster management is to implement management of large numbers of distributed network devices. Cluster management offers the following advantages:

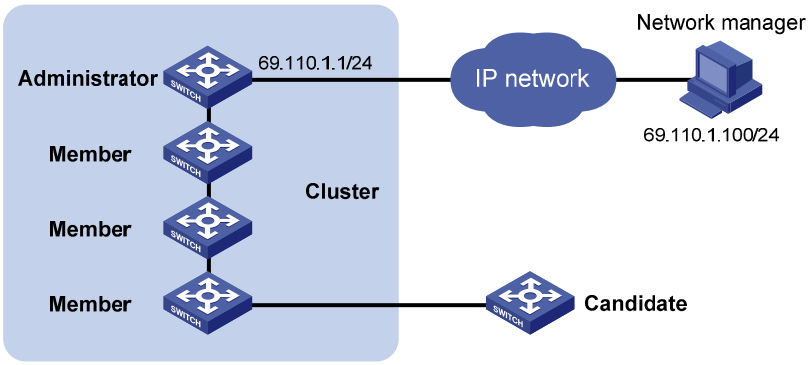
- Saving public IP address resource
- Simplifying configuration and management tasks. By configuring a public IP address on one device, you can configure and manage a group of devices without the trouble of logging in to each device separately.
- Providing topology discovery and display function, which is useful for network monitoring and debugging
- Allowing simultaneous software upgrading and parameter configuration on multiple devices, free of topology and distance limitations

Roles in a Cluster

The devices in a cluster play different roles according to their different functions and status. You can specify the following three roles for the devices:

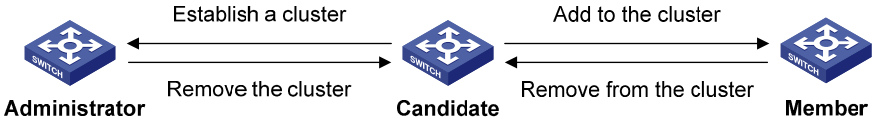
- Management device (Administrator): The device providing management interfaces for all devices in a cluster and the only device configured with a public IP address. You can specify one and only one management device for a cluster. Any configuration, management, and monitoring of the other devices in a cluster can only be implemented through the management device. When a device is specified as the management device, it collects related information to discover and define candidate devices.
- Member device (Member): A device managed by the management device in a cluster.
- Candidate device (Candidate): A device that does not belong to any cluster but can be added to a cluster. Different from a member device, its topology information has been collected by the management device but it has not been added to the cluster.

Figure 1-1 Network diagram for a cluster



As shown in Figure 1-1, the device configured with a public IP address and performs the management function is the management device, the other managed devices are member devices, and the device that does not belong to any cluster but can be added to a cluster is a candidate device. The management device and the member devices form the cluster.

Figure 1-2 Role change in a cluster



As shown in Figure 1-2, a device in a cluster changes its role according to the following rules:

- A candidate device becomes a management device when you create a cluster on it. A management device becomes a candidate device only after the cluster is removed.
- A candidate device becomes a member device after being added to a cluster. A member device becomes a candidate device after it is removed from the cluster.

How a Cluster Works

Cluster management is implemented through HW Group Management Protocol version 2 (HGMPv2), which consists of the following three protocols:

- Neighbor Discovery Protocol (NDP)
- Neighbor Topology Discovery Protocol (NTDP)
- Cluster

A cluster configures and manages the devices in it through the above three protocols. Cluster management involves topology information collection and the establishment and maintenance of a cluster. Topology information collection and cluster maintenance are independent from each other, with the former starting before the cluster is created:

- All devices use NDP to collect the information of the directly connected neighbors, including their software version, host name, MAC address and port number.
- The management device uses NTDP to collect the information of the devices within user-specified hops and the topology information of all devices and specify the candidate devices of the cluster.
- The management device adds or deletes a member device and modifies cluster management configuration according to the candidate device information collected through NTDP.

Introduction to NDP

NDP is used to discover the information about directly connected neighbors, including the device name, software version, and connecting port of the adjacent devices. NDP works in the following ways:

- A device running NDP periodically sends NDP packets to its neighbors. An NDP packet carries NDP information (including the device name, software version, and connecting port, etc.) and the holdtime, which indicates how long the receiving devices will keep the NDP information. At the same time, the device also receives (but does not forward) the NDP packets from its neighbors.
- A device running NDP stores and maintains an NDP table. The device creates an entry in the NDP table for each neighbor. If a new neighbor is found, meaning the device receives an NDP packet sent by the neighbor for the first time, the device adds an entry in the NDP table. If the NDP information carried in the NDP packet is different from the stored information, the corresponding entry and holdtime in the NDP table are updated; otherwise, only the holdtime of the entry is updated. If no NDP information from the neighbor is received when the holdtime times out, the corresponding entry is removed from the NDP table.

NDP runs on the data link layer, and therefore supports different network layer protocols.

Introduction to NTDP

NTDP provides information required for cluster management; it collects topology information about the devices within the specified hop count. Based on the neighbor information stored in the neighbor table maintained by NDP, NTDP on the management device advertises NTDP topology collection requests to collect the NDP information of all the devices in a specific network range as well as the connection information of all its neighbors. The information collected will be used by the management device or the network management software to implement required functions.

When a member device detects a change on its neighbors through its NDP table, it informs the management device through handshake packets. Then the management device triggers its NTDP to collect specific topology information, so that its NTDP can discover topology changes timely.

The management device collects topology information periodically. You can also administratively launch a topology information collection. The process of topology information collection is as follows:

- The management device periodically sends NTDP topology collection request from the NTDP-enabled ports.
- Upon receiving the request, the device sends NTDP topology collection response to the management device, copies this response packet on the NTDP-enabled port and sends it to the adjacent device. Topology collection response includes the basic information of the NDP-enabled device and NDP information of all adjacent devices.
- The adjacent device performs the same operation until the NTDP topology collection request is sent to all the devices within specified hops.

When the NTDP topology collection request is advertised in the network, large numbers of network devices receive the NTDP topology collection request and send NTDP topology collection response at the same time, which may cause congestion and the management device busyness. To avoid such case, the following methods can be used to control the speed of the NTDP topology collection request advertisement:

- Upon receiving an NTDP topology collection request, each device does not forward it, instead, it waits for a period of time and then forwards the NTDP topology collection request on the first NTDP-enabled port.
- On the same device, except the first port, each NTDP-enabled port waits for a period of time and

then forwards the NTDP topology collection request after its prior port forwards the NTDP topology collection request.

Cluster management maintenance

1) Adding a candidate device to a cluster

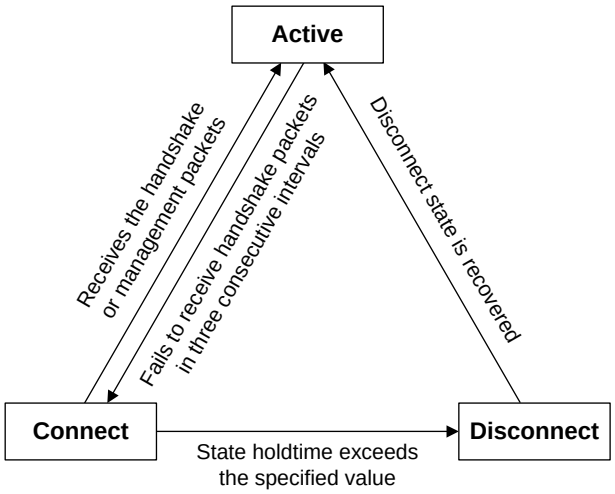
You should specify the management device before creating a cluster. The management device discovers and defines a candidate device through NDP and NTDP protocols. The candidate device can be automatically or manually added to the cluster.

After the candidate device is added to the cluster, it can obtain the member number assigned by the management device and the private IP address used for cluster management.

2) Communication within a cluster

In a cluster the management device communicates with its member devices by sending handshake packets to maintain connection between them. The management/member device state change is shown in Figure 1-3.

Figure 1-3 Management/member device state change



- After a cluster is created, a candidate device is added to the cluster and becomes a member device, the management device saves the state information of its member device and identifies it as Active. And the member device also saves its state information and identifies itself as Active.
- After a cluster is created, its management device and member devices begin to send handshake packets. Upon receiving the handshake packets from the other side, the management device or a member device simply remains its state as Active, without sending a response.
- If the management device does not receive the handshake packets from a member device in an interval three times of the interval to send handshake packets, it changes the status of the member device from Active to Connect. Likewise, if a member device fails to receive the handshake packets from the management device in an interval three times of the interval to send handshake packets, the status of itself will also be changed from Active to Connect.
- If this management device, in information holdtime, receives the handshake or management packets from its member device which is in Connect state, it changes the state of its member device to Active; otherwise, it changes the state of its member device to Disconnect, in which case the management device considers its member device disconnected. If this member device, which is in Connect state, receives handshake or management packets from the management device in information holdtime, it changes its state to Active; otherwise, it changes its state to Disconnect.
- If the communication between the management device and a member device is recovered, the

member device which is in Disconnect state will be added to the cluster. After that, the state of the member device locally and on the management device will be changed to Active.

Besides, a member device informs the management device using handshake packets when there is a neighbor topology change.

Management VLAN

The management VLAN is a VLAN used for communication in a cluster; it limits the cluster management range. Through configuration of the management VLAN, the following functions can be implemented:

- Management packets (including NDP, NTDP and handshake packets) are restricted within the management VLAN, therefore isolated from other packets, which enhances security.
- The management device and the member devices communicate with each other through the management VLAN.

For a cluster to work normally, you must set the packets from the management VLAN to pass the ports connecting the management device and the member/candidate devices (including the cascade ports). Therefore:

- If the packets from the management VLAN cannot pass a port, the device connected with the port cannot be added to the cluster. Therefore, if the ports (including the cascade ports) connecting the management device and the member/candidate devices prohibit the packets from the management VLAN, you can set the packets from the management VLAN to pass the ports on candidate devices with the management VLAN auto-negotiation function.
- Only when the default VLAN ID of the cascade ports and the ports connecting the management device and the member/candidate devices is that of the management VLAN can you set the packets without tags from the management VLAN to pass the ports; otherwise, only the packets with tags from the management VLAN can pass the ports.

Note

- If a candidate device is connected to a management device through another candidate device, the ports between the two candidate devices are cascade ports.
- For information about VLAN, refer to *VLAN Configuration* in the *Access Volume*.

Cluster Configuration Task List

Before configuring a cluster, you need to determine the roles and functions the devices play. You also need to configure the related functions, preparing for the communication between devices within the cluster.

Complete these tasks to configure a cluster:

Task		Remarks
Configuring the Management Device	Enabling NDP Globally and for Specific Ports	Optional
	Configuring NDP Parameters	Optional
	Enabling NTDP Globally and for Specific Ports	Optional
	Configuring NTDP Parameters	Optional
	Manually Collecting Topology Information	Optional
	Enabling the Cluster Function	Optional
	Establishing a Cluster	Required
	Enabling Management VLAN Auto-negotiation	Required
	Configuring Communication Between the Management Device and the Member Devices Within a Cluster	Optional
	Configuring Cluster Management Protocol Packets	Optional
	Cluster Member Management	Optional
Configuring the Member Devices	Enabling NDP	Optional
	Enabling NTDP	Optional
	Manually Collecting Topology Information	Optional
	Enabling the Cluster Function	Optional
	Deleting a Member Device from a Cluster	Optional
Configuring Access Between the Management Device and Its Member Devices		Optional
Adding a Candidate Device to a Cluster		Optional
Configuring Advanced Cluster Functions	Configuring Topology Management	Optional
	Configuring Interaction for a Cluster	Optional
	SNMP Configuration Synchronization Function	Optional
	Configuring Web User Accounts in Batches	Optional

Caution

- Disabling the NDP and NTDP functions on the management device and member devices after a cluster is created will not cause the cluster to be dismissed, but will influence the normal operation of the cluster.
- When both the cluster function and the 802.1x function (or the MAC address authentication) are enabled on devices, you need to enable HABP on the devices. Otherwise, the management device of the cluster cannot manage the devices connected with it. For description of HABP, refer to *HABP Configuration* in the *Security Volume*.
- If the routing table of the management device is full when a cluster is established, that is, entries with the destination address as a candidate device cannot be added to the routing table, all candidate devices will be added to and removed from the cluster repeatedly.
- If the routing table of a candidate device is full when the candidate device is added to a cluster, that is, the entry with the destination address as the management device cannot be added to the routing table, the candidate device will be added to and removed from the cluster repeatedly.

Configuring the Management Device

Enabling NDP Globally and for Specific Ports

For NDP to work normally, you must enable NTDP both globally and on specific ports.

Follow these steps to enable NDP globally and for specific ports:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Enable NDP globally		**ndp enable**	Optional Enabled by default.
Enable the NDP feature for the port(s)	In system view	**ndp enable interface***interface-list*	Use either command By default, NDP is enabled globally and also on all ports.
	In Ethernet port view or Layer 2 aggregate interface view	**interface** *interface-type interface-number*	
		ndp enable	

Note

You are recommended to disable NDP on the port which connects with the devices that do not need to join the cluster, preventing the management device from adding the device which needs not to join the cluster and collecting the topology information of this device.

Configuring NDP Parameters

A port enabled with NDP periodically sends NDP packets to its neighbors. If no NDP information from the neighbor is received when the holdtime times out, the corresponding entry is removed from the NDP table.

Follow these steps to configure NDP parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the interval for sending NDP packets	**ndp timer hello** *hello-time*	Optional 60 seconds by default.
Configure the period for the receiving device to keep the NDP packets	**ndp timer aging** *aging-time*	Optional 180 seconds by default.

Caution

The time for the receiving device to hold NDP packets cannot be shorter than the interval for sending NDP packets; otherwise, the NDP table may become instable.

Enabling NTDP Globally and for Specific Ports

For NTDP to work normally, you must enable NTDP both globally and on specific ports.

Follow these steps to enable NTDP globally and for specific ports:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable NTDP globally	**ntdp enable**	Optional Enabled by default
Enable NTDP for the port	**interface** *interface-type interface-number*	Optional NTDP is enabled on all ports by default.
	ntdp enable	

Note

You are recommended to disable NTDP on the port which connects with the devices that do not need to join the cluster, preventing the management device from adding the device which needs not to join the cluster and collecting the topology information of this device.

Configuring NTDP Parameters

By configuring the maximum hops for collecting topology information, you can get topology information

of the devices in a specified range, thus avoiding unlimited topology collection.

After the interval for collecting topology information is configured, the device collects the topology information at this interval.

To avoid network congestion caused by large amounts of topology responses received in short periods:

- Upon receiving an NTDP topology collection request, a device does not forward it, instead, it waits for a period of time and then forwards the NTDP topology collection request on its first NTDP-enabled port.
- On the same device, except the first port, each NTDP-enabled port waits for a period of time and then forwards the NTDP topology collection request after the previous port forwards the NTDP topology collection request.

Follow these steps to configure NTDP parameters:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the maximum hops for topology collection	**ntdp hop** *hop-value*	Optional 3 by default.
Configure the interval to collect topology information	**ntdp timer** *interval-time*	Optional 1 minute by default.
Configure the delay to forward topology-collection request packets on the first port	**ntdp timer hop-delay** *time*	Optional 200 ms by default.
Configure the port delay to forward topology collection request on other ports	**ntdp timer port-delay** *time*	Optional 20 ms by default.

Note

The two delay values should be configured on the topology collecting device. A topology collection request sent by the topology collecting device carries the two delay values, and a device that receives the request forwards the request according to the delays.

Manually Collecting Topology Information

The management device collects topology information periodically after a cluster is created. In addition, you can configure to manually initiate topology information collection, thus managing and monitoring the device on real time, regardless of whether a cluster is created.

Follow these steps to configure to manually collect topology information:

To do…	Use the command…	Remarks
Manually collect topology information	**ntdp explore**	Required

Enabling the Cluster Function

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable the cluster function globally	**cluster enable**	Optional Enabled by default.

Establishing a Cluster

Before establishing a cluster, you need to specify the management VLAN, and you cannot modify the management VLAN after a device is added to the cluster.

In addition, you need to configure a private IP address pool for the devices to be added to the cluster on the device to be configured as the management device before establishing a cluster. Meanwhile, the IP addresses of the VLAN interfaces of the management device and member devices cannot be in the same network segment as that of the cluster address pool; otherwise, the cluster cannot work normally. When a candidate device is added to a cluster, the management device assigns it a private IP address for it to communicate with other devices in the cluster.

You can establish a cluster in two ways: manually and automatically. With the latter, you can establish a cluster according to the prompt information. The system:

1) Prompts you to enter a name for the cluster you want to establish;
2) Lists all the candidate devices within your predefined hop count;
3) Starts to automatically add them to the cluster.

You can press **Ctrl+C** anytime during the adding process to exit the cluster auto-establishment process. However, this will only stop adding new devices into the cluster, and devices already added into the cluster are not removed.

Follow these steps to manually establish a cluster:

To do…		Use the command…	Remarks
Enter system view		**system-view**	—
Specify the management VLAN		**management-vlan** *vlan-id*	Optional By default, VLAN 1 is the management VLAN.
Enter cluster view		**cluster**	—
Configure the private IP address range for member devices		**ip-pool** *administrator-ip-address* { *mask* | *mask-length* }	Required Not configured by default.
Establish a cluster	Manually establish a cluster	**build** *name*	Required Use either approach By default, the device is not the management device.
	Automatically establish a cluster	**auto-build** [**recover**]	

Enabling Management VLAN Auto-negotiation

The management VLAN limits the cluster management range. If the device discovered by the management device does not belong to the management VLAN, meaning the cascade ports and the ports connecting with the management device do not allow the packets from the management VLAN to pass, and the new device cannot be added to the cluster. Through the configuration of the management VLAN auto-negotiation function, the cascade ports and the ports directly connected to the management device can be automatically added to the management VLAN.

Follow these steps to configure management VLAN auto-negotiation:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Enable management VLAN auto-negotiation	**management-vlan synchronization enable**	Required Disabled by default.

Configuring Communication Between the Management Device and the Member Devices Within a Cluster

In a cluster, the management device and member devices communicate by sending handshake packets to maintain connection between them. You can configure interval of sending handshake packets and the holdtime of a device on the management device. This configuration applies to all member devices within the cluster. For a member device in Connect state:

- If the management device does not receive handshake packets from a member device within the holdtime, it changes the state of the member device to Disconnect. When the communication is recovered, the member device needs to be re-added to the cluster (this process is automatically performed).
- If the management device receives handshake packets from the member device within the holdtime, the state of the member device remains Active.

Follow these steps to configure communication between the management device and the member devices within a cluster:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Configure the interval to send handshake packets	**timer** *interval-time*	Optional 10 seconds by default
Configure the holdtime of a device	**holdtime** *seconds*	Optional 60 seconds by default

Configuring Cluster Management Protocol Packets

By default, the destination MAC address of cluster management protocol packets (including NDP, NTDP and HABP packets) is a multicast MAC address 0180-C200-000A, which IEEE reserved for later use. Since some devices cannot forward the multicast packets with the destination MAC address of

0180-C200-000A, cluster management packets cannot traverse these devices. For a cluster to work normally in this case, you can modify the destination MAC address of a cluster management protocol packet without changing the current networking.

The management device periodically sends MAC address negotiation broadcast packets to advertise the destination MAC address of the cluster management protocol packets.

Follow these steps to configure the destination MAC address of the cluster management protocol packets:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Configure the destination MAC address for cluster management protocol packets	**cluster-mac** *mac-address*	Required The destination MAC address is 0180-C200-000A by default.
Configure the interval to send MAC address negotiation broadcast packets	**cluster-mac syn-interval** *interval-time*	Optional One minute by default.

Caution

When you configure the destination MAC address for cluster management protocol packets:

- If the interval for sending MAC address negotiation broadcast packets is 0, the system automatically sets it to 1 minute.
- If the interval for sending MAC address negotiation broadcast packets is not 0, the interval remains unchanged.

Cluster Member Management

You can manually add a candidate device to a cluster, or remove a member device from a cluster.

If a member device needs to be rebooted for software upgrade or configuration update, you can remotely reboot it through the management device.

Adding a member device

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Add a candidate device to the cluster	**add-member** [*member-number*] **mac-address** *mac-address* [**password** *password*]	Required

Removing a member device

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Remove a member device from the cluster	**delete-member** *member-number* [**to-black-list**]	Required

Rebooting a member device

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Reboot a specified member device	**reboot member** { *member-number* | **mac-address** *mac-address* } [**eraseflash**]	Required

Configuring the Member Devices

Enabling NDP

Refer to Enabling NDP Globally and for Specific Ports.

Enabling NTDP

Refer to Enabling NTDP Globally and for Specific Ports.

Manually Collecting Topology Information

Refer to Manually Collecting Topology Information.

Enabling the Cluster Function

Refer to Enabling the Cluster Function.

Deleting a Member Device from a Cluster

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Delete a member device from the cluster	**undo administrator-address**	Required

Configuring Access Between the Management Device and Its Member Devices

After having successfully configured NDP, NTDP and cluster, you can configure, manage and monitor

the member devices through the management device. You can manage member devices in a cluster through switching from the operation interface of the management device to that of a member device or configure the management device by switching from the operation interface of a member device to that of the management device.

Follow these steps to configure access between member devices of a cluster:

To do…	Use the command…	Remarks
Switch from the operation interface of the management device to that of a member device	**cluster switch-to** { *member-number* | **mac-address** *mac-address* | **sysname** *member-sysname* }	Required
Switch from the operation interface of a member device to that of the management device	**cluster switch-to administrator**	Required

Caution

Telnet connection is used in the switching between the management device and a member device. Note the following when switching between them:

- Authentication is required when you switch from a member device to the management device. The switching fails if authentication is not passed. Your user level is allocated according to the predefined level by the management device if authentication is passed.
- When a candidate device is added to a cluster and becomes a member device, its super password will be automatically synchronized to the management device. Therefore, after a cluster is established, it is not recommended to modify the super password of any member (including the management device and member devices) of the cluster; otherwise, the switching may fail because of an authentication failure.
- If the member specified in this command does not exist, the system prompts error when you execute the command; if the switching succeeds, your user level on the management device is retained.
- If the Telnet users on the device to be logged in reach the maximum number, the switching fails.
- To prevent resource waste, avoid ring switching when configuring access between cluster members. For example, if you switch from the operation interface of the management device to that of a member device and then need to switch back to that of the management device, use the **quit** command to end the switching, but not the **cluster switch-to administrator** command to switch to the operation interface of the management device.

Adding a Candidate Device to a Cluster

Follow these steps to add a candidate device to a cluster:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—

To do…	Use the command…	Remarks
Add a candidate device to the cluster	**administrator-address** *mac-address* **name** *name*	Required

Configuring Advanced Cluster Functions

This section covers these topics:

- Configuring Topology Management
- Configuring Interaction for a Cluster
- SNMP Configuration Synchronization Function
- Configuring Web User Accounts in Batches

Configuring Topology Management

The concepts of blacklist and whitelist are used for topology management. An administrator can diagnose the network by comparing the current topology (namely, the information of a node and its neighbors in the cluster) and the standard topology.

- Topology management whitelist (standard topology): A whitelist is a list of topology information that has been confirmed by the administrator as correct. You can get the information of a node and its neighbors from the current topology. Based on the information, you can manage and maintain the whitelist by adding, deleting or modifying a node.
- Topology management blacklist: Devices in a blacklist are not allowed to join a cluster. A blacklist contains the MAC addresses of devices. If a blacklisted device is connected to a network through another device not included in the blacklist, the MAC address and access port of the latter are also included in the blacklist. The candidate devices in a blacklist can be added to a cluster only if the administrator manually removes them from the list.

The whitelist and blacklist are mutually exclusive. A whitelist member cannot be a blacklist member, and vice versa. However, a topology node can belong to neither the whitelist nor the blacklist. Nodes of this type are usually newly added nodes, whose identities are to be confirmed by the administrator.

You can back up and restore the whitelist in the following two ways:

- Backing them up on the FTP server shared by the cluster. You can manually restore the whitelist and blacklist from the FTP server.
- Backing them up in the Flash of the management device. When the management device restarts, the whitelist and blacklist will be automatically restored from the Flash. When a cluster is re-established, you can choose whether to restore the whitelist and blacklist from the Flash automatically, or you can manually restore them from the Flash of the management device.

Follow these steps to configure cluster topology management:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Add a device to the blacklist	**black-list add-mac** *mac-address*	Optional
Remove a device from the blacklist	**black-list delete-mac** { **all** | *mac-address* }	Optional

To do…	Use the command…	Remarks
Confirm the current topology and save it as the standard topology	**topology accept** { **all** [**save-to** { **ftp-server** | **local-flash** }] | **mac-address** *mac-address* | **member-id** *member-number* }	Optional
Save the standard topology to the FTP server or the local Flash	**topology save-to** { **ftp-server** | **local-flash** }	Optional
Restore the standard topology information from the FTP server or the local Flash	**topology restore-from** { **ftp-server** | **local-flash** }	Optional

Configuring Interaction for a Cluster

After establishing a cluster, you can configure FTP/TFTP server, NM host and log host for the cluster on the management device.

- After you configure an FTP/TFTP server for a cluster, the members in the cluster access the FTP/TFTP server configured through the management device.
- After you configure a log host for a cluster, all the log information of the members in the cluster will be output to the configured log host in the following way: first, the member devices send their log information to the management device, which then converts the addresses of log information and sends them to the log host.
- After you configure an NM host for a cluster, the member devices in the cluster send their Trap messages to the shared SNMP NM host through the management device.

If the port of an access NM device (including FTP/TFTP server, NM host and log host) does not allow the packets from the management VLAN to pass, the NM device cannot manage the devices in a cluster through the management device. In this case, on the management device, you need to configure the VLAN interface of the access NM device (including FTP/TFTP server, NM host and log host) as the NM interface.

Follow these steps to configure the interaction for a cluster:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Configure the FTP server shared by the cluster	**ftp-server** *ip-address* [**user-name** *username* **password** { **simple** | **cipher** } *password*]	Required By default, no FTP server is configured for a cluster.
Configure the TFTP server shared by the cluster	**tftp-server** *ip-address*	Required By default, no TFTP server is configured for a cluster.
Configure the log host shared by the member devices in the cluster	**logging-host** *ip-address*	Required By default, no log host is configured for a cluster.
Configure the SNMP NM host shared by the cluster	**snmp-host** *ip-address* [**community-string read** *string1* **write** *string2*]	Required By default, no SNMP host is configured.

To do…	Use the command…	Remarks
Configure the NM interface of the management device	**nm-interface vlan-interface** *vlan-interface-id*	Optional

Caution

To isolate management protocol packets of a cluster from packets outside the cluster, you are recommended to configure to prohibit packets from the management VLAN from passing the ports that connect the management device with the devices outside the cluster and configure the NM interface for the management device.

SNMP Configuration Synchronization Function

SNMP configuration synchronization function facilitates management of a cluster, with which you can perform SNMP-related configurations on the management device and synchronize them to the member devices on the whitelist. This operation is equal to configuring multiple member devices at one time, simplifying the configuration process. Follow these steps to configure the SNMP configuration synchronization function:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Configure the SNMP community name shared by a cluster	**cluster-snmp-agent community** { **read** | **write** } *community-name* [**mib-view** *view-name*]	Required
Configure the SNMPv3 group shared by a cluster	**cluster-snmp-agent group v3** *group-name* [**authentication** | **privacy**] [**read-view** *read-view*] [**write-view** *write-view*] [**notify-view** *notify-view*]	Required
Create or update information of the MIB view shared by a cluster	**cluster-snmp-agent mib-view included** *view-name oid-tree*	Required By default, the name of the MIB view shared by a cluster is ViewDefault and a cluster can access the ISO subtree.
Add a user for the SNMPv3 group shared by a cluster	**cluster-snmp-agent usm-user v3** *user-name group-name* [**authentication-mode** { **md5** | **sha** } *auth-password*] [**privacy-mode des56** *priv-password*]	Required

Note

- The SNMP-related configurations are retained when a cluster is dismissed or the member devices are removed from the whitelist.
- For information about SNMP, refer to *SNMP Configuration* in the *System Volume*.

Configuring Web User Accounts in Batches

Configuring Web user accounts in batches enables you to configure on the management device the username and password used to log in to the devices (including the management device and member devices) within a cluster through Web and synchronize the configurations to the member devices in the whitelist. This operation is equal to performing the configurations on the member devices. You need to enter your username and password when you log in to the devices (including the management device and member devices) in a cluster through Web.

Follow these steps to configure Web user accounts in batches:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enter cluster view	**cluster**	—
Configure Web user accounts in batches	**cluster-local-user** *username* **password** { **cipher** | **simple** } *password*	Required

Note

If a cluster is dismissed or the member devices are removed from the whitelist, the configurations of Web user accounts are still retained.

Displaying and Maintaining Cluster Management

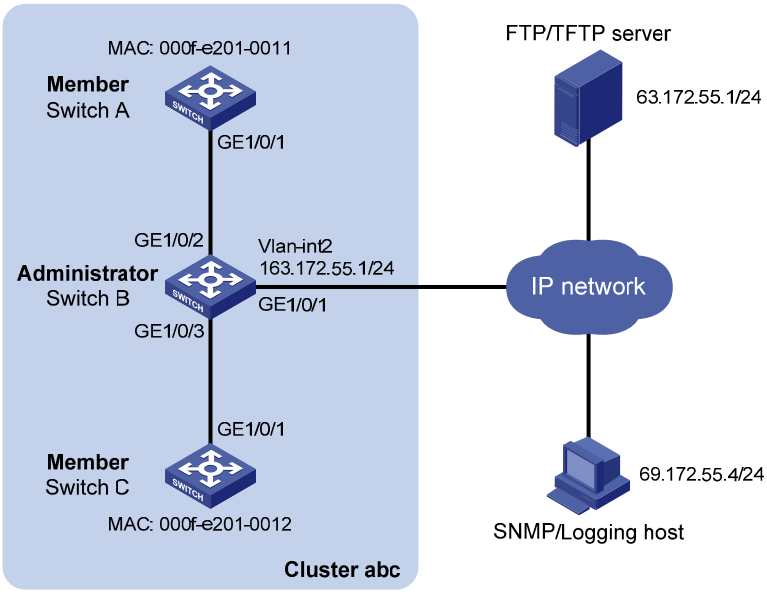
To do…	Use the command…	Remarks
Display NDP configuration information	**display ndp** [**interface** *interface-list*]	Available in any view
Display the global NTDP information	**display ntdp**	
Display the device information collected through NTDP	**display ntdp device-list** [**verbose**]	
Display the detailed NTDP information of a specified device	**display ntdp single-device mac-address** *mac-address*	
View cluster state and statistics	**display cluster**	
View the standard topology information	**display cluster base-topology** [**mac-address** *mac-address* | **member-id** *member-number*]	
View the current blacklist of the cluster	**display cluster black-list**	
View the information of candidate devices	**display cluster candidates** [**mac-address** *mac-address* | **verbose**]	
Display the current topology information or the topology path between two devices	**display cluster current-topology** [**mac-address** *mac-address* [**to-mac-address** *mac-address*] | **member-id** *member-number* [**to-member-id** *member-number*]]	
Display members in a cluster	**display cluster members** [*member-number* | **verbose**]	
Clear NDP statistics	**reset ndp statistics** [**interface** *interface-list*]	Available in user view

Cluster Management Configuration Example

Network requirements

- Three switches form cluster **abc**, whose management VLAN is VLAN 10. In the cluster, Switch B serves as the management device (Administrator), whose network management interface is VLAN-interface 2; Switch A and Switch C are the member devices (Member).
- All the devices in the cluster use the same FTP server and TFTP server on host 63.172.55.1/24, and use the same SNMP NMS and log services on host IP address: 69.172.55.4/24.
- Add the device whose MAC address is 000f-e201-0013 to the blacklist.

Figure 1-4 Network diagram for cluster management configuration



Configuration procedure

1) Configure the member device Switch A

Enable NDP globally and for port GigabitEthernet 1/0/1.

```
<SwitchA> system-view
[SwitchA] ndp enable
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] ndp enable
[SwitchA-GigabitEthernet1/0/1] quit
```

Enable NTDP globally and for port GigabitEthernet 1/0/1.

```
[SwitchA] ntdp enable
[SwitchA] interface gigabitethernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] ntdp enable
[SwitchA-GigabitEthernet1/0/1] quit
```

Enable the cluster function.

```
[SwitchA] cluster enable
```

2) Configure the member device Switch C

As the configurations of the member devices are the same, the configuration procedure of Switch C is omitted here.

3) Configure the management device Switch B

Enable NDP globally and for ports GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.

```
<SwitchB> system-view
[SwitchB] ndp enable
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] ndp enable
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] ndp enable
```

```
[SwitchB-GigabitEthernet1/0/3] quit
```

Configure the period for the receiving device to keep NDP packets as 200 seconds.

```
[SwitchB] ndp timer aging 200
```

Configure the interval to send NDP packets as 70 seconds.

```
[SwitchB] ndp timer hello 70
```

Enable NTDP globally and for ports GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.

```
[SwitchB] ntdp enable
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] ntdp enable
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] ntdp enable
[SwitchB-GigabitEthernet1/0/3] quit
```

Configure the hop count to collect topology as 2.

```
[SwitchB] ntdp hop 2
```

Configure the delay to forward topology-collection request packets on the first port as 150 ms.

```
[SwitchB] ntdp timer hop-delay 150
```

Configure the delay to forward topology-collection request packets on the first port as 15 ms.

```
[SwitchB] ntdp timer port-delay 15
```

Configure the interval to collect topology information as 3 minutes.

```
[SwitchB] ntdp timer 3
```

Configure the management VLAN of the cluster as VLAN 10.

```
[SwitchB] vlan 10
[SwitchB-vlan10] quit
[SwitchB] management-vlan 10
```

Configure ports GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 as Trunk ports and allow packets from the management VLAN to pass.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type trunk
[SwitchB-GigabitEthernet1/0/2] port trunk permit vlan 10
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port link-type trunk
[SwitchB-GigabitEthernet1/0/3] port trunk permit vlan 10
[SwitchB-GigabitEthernet1/0/3] quit
```

Enable the cluster function.

```
[SwitchB] cluster enable
```

Configure a private IP address range for the member devices, which is from 172.16.0.1 to 172.16.0.7.

```
[SwitchB] cluster
[SwitchB-cluster] ip-pool 172.16.0.1 255.255.255.248
```

Configure the current device as the management device, and establish a cluster named **abc**.

```
[SwitchB-cluster] build abc
```

```
Restore topology from local flash file,for there is no base topology.
(Please confirm in 30 seconds, default No). (Y/N)
N
```

Enable management VLAN auto-negotiation.

```
[abc_0.SwitchB-cluster] management-vlan synchronization enable
```

Configure the holdtime of the member device information as 100 seconds.

```
[abc_0.SwitchB-cluster] holdtime 100
```

Configure the interval to send handshake packets as 10 seconds.

```
[abc_0.SwitchB-cluster] timer 10
```

Configure the FTP Server, TFTP Server, Log host and SNMP host for the cluster.

```
[abc_0.SwitchB-cluster] ftp-server 63.172.55.1
[abc_0.SwitchB-cluster] tftp-server 63.172.55.1
[abc_0.SwitchB-cluster] logging-host 69.172.55.4
[abc_0.SwitchB-cluster] snmp-host 69.172.55.4
```

Add the device whose MAC address is 00E0-FC01-0013 to the blacklist.

```
[abc_0.SwitchB-cluster] black-list add-mac 00e0-fc01-0013
[abc_0.SwitchB-cluster] quit
```

Add port GigabitEthernet 1/0/1 to VLAN 2, and configure the IP address of VLAN-interface 2.

```
[abc_0.SwitchB] vlan 2
[abc_0.SwitchB-vlan2] port gigabitethernet 1/0/1
[abc_0.SwitchB] quit
[abc_0.SwitchB] interface vlan-interface 2
[abc_0.SwitchB-Vlan-interface2] ip address 163.172.55.1 24
[abc_0.SwitchB-Vlan-interface2] quit
```

Configure VLAN-interface 2 as the network management interface.

```
[abc_0.SwitchB] cluster
[abc_0.SwitchB-cluster] nm-interface vlan-interface 2
```

Table of Contents

1 IRF Stack Configuration

When configuring IRF stack, go to these sections for information you are interested in:

IRF Stack Overview

Introduction

Intelligent Resilient Framework (IRF) allows you to build an IRF stack, namely a united device, by interconnecting multiple devices through stack ports. You can manage all the devices in the IRF stack by managing the united device.

In an IRF stack, every single device is a stack member, and plays one of the following two roles according to its function:

- Master: A stack member. It is elected to manage the entire stack. An IRF stack has only one master at one time.
- Slave: A stack member. It is managed by the master and operates as a backup of the master. In an IRF stack, except for the master, all the other devices are the slaves.

Role election defines the roles of stack members and is discussed in a later section.

Stack Connections

Pysical stack port

To make an IRF stack operate normally, you need to connect the stack members physically. Physical ports that are dedicated to stack connection on devices are called physical stack ports. For the Switch 4800G series, the 10 GE interface modules can be inserted into the expansion module slots on the rear panel of the switch to provide physical stack ports. The following 10 GE interface modules can be used to provide physical stack ports:

- One-port 10 GE XFP interface module
- Dual-port 10 GE XFP interface module
- Short-haul dual-port 10 GE CX4 interface module

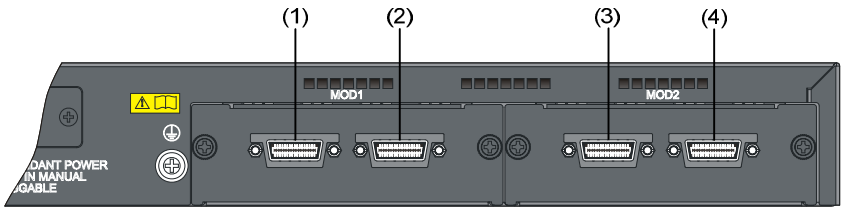
Note

For the details of the interface modules, refer to *3Com Switch 4500G-4800G Family 10G Module Installation Guide*.

You can connect physical stack ports of the Switch 4800G series with either the CX4 dedicated cables or fibers according to the interface type on the interface module. Dedicated cables provide higher reliability and performance; whereas fibers connect physical devices located very far from each other and provide flexible application.

The physical stack ports are numbered according to their physical locations on the rear panel of the Switch 4800G series. With the rear panel facing you, the physical stack ports are numbered successively from left to right: ports on the interface module in slot 1 are numbered 1 and 2, and ports on the interface module in slot 2 are numbered 3 and 4, as shown in Figure 1-1, which illustrates an example of inserting a CX4 dual-port interface module.

Figure 1-1 Numbering physical stack ports



Caution

If you insert a one-port interface module into the slot, then the number of the physical stack port corresponding to the module in slot 1 is 1, and the number of the physical stack port corresponding to the module in slot 2 is 3. For the number of physical stack ports, see Configuring Stack Ports.

Physical stack ports can be used for both stack connection and service data transmission. When establishing a stack, you need to specify that the physical stack ports are used for the stack, that is, bind them with logical stack port(s) to implement stack connection and establishment.

Logical stack port

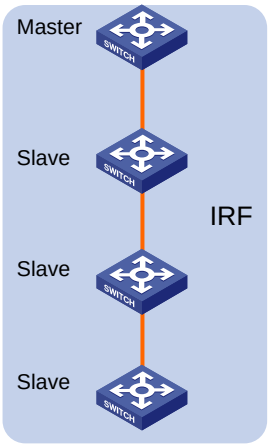
You can set a physical stack port as a logical stack port, which is used to transfer data and protocol packets among stack members. An Switch 4800G series can be configured with two logical stack ports, which are numbered 1 and 2, to connect other devices in a stack.

An IRF stack typically has a bus connection or a ring connection:

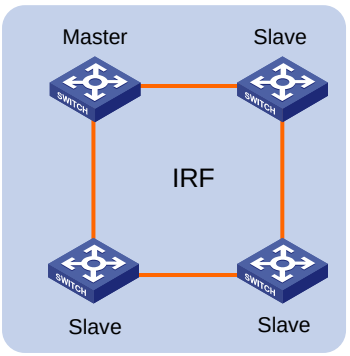
- Bus connection: Given a device, its logical stack port 1 is connected to logical stack port 2 of another device, and its logical stack port 2 is connected to logical stack port 1 of a third one; devices are connected to form a single straight connection, as shown in Figure 1-2.

- Ring connection: Given a device, its logical stack port 1 is connected to logical stack port 2 of another device, and its logical stack port 2 is connected to logical stack port 1 of a third one, as shown in Figure 1-2.

Figure 1-2 Physical connections of IRF stack



Bus connection



Ring connection

The orange lines in the figure represent stack cables, which are different from common Ethernet network cables. The stack cable can be composed of either one physical cable or multiple physical cables.

A ring connection is more reliable than a bus connection. The failure of one link in a ring connection does not affect the function and performance of the stack, whereas the failure of one link in a bus connection causes the split of the stack.

Note

You can connect at most nine Switch 4800G series Ethernet switches to form a stack.

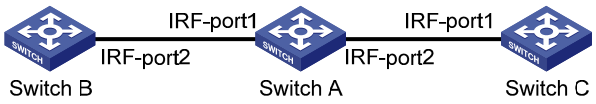
Correspondence between a logical stack port and a physical stack port

The connection of logical stack ports is based on that of physical stack ports; therefore, you need to bind a logical stack port with physical stack port(s). A logical stack port can be bound to one physical stack port or, to realize link backup and bandwidth expansion, bound to two physical stack ports (aggregated as an aggregate stack port).

You need to specify the correspondence between a logical stack port and physical stack port(s) through command line. When you specify that a logical stack port is bound to one physical stack port, the serial number of the physical stack port bound to logical stack port 1 must be smaller than that of the physical stack port bound to logical stack port 2; when you specify that a logical stack port is bound to two physical stack ports (aggregate stack port), these two physical stack ports must be on the same module.

As shown in Figure 1-3. Switch A connects to Switch B and Switch C through logical stack ports IRF-port 1 and IRF-port 2 respectively.

Figure 1-3 Stack port correspondence



Based on the type and number of the interface module inserted on Switch A, you can adopt one of the following typical correspondences to establish a stack connection.

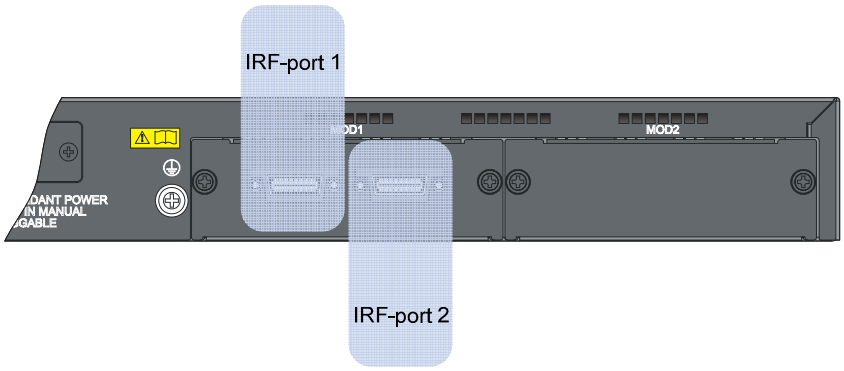
Note

- The dual-port 10 GE CX4 interface module is used in the following examples to introduce correspondence between the logical stack port and the physical stack port(s).
- When the dual-port 10 GE SFP interface module is used, the correspondence between the logical stack port and the physical stack port(s) is similar.

---

1) Stack port correspondence for one interface module

Figure 1-4 Stack port correspondence for one interface module



When a dual-port interface module is installed, you need to bind IRF-port 1 to physical stack port 1, and IRF-port 2 to physical stack port 2 (as shown in Figure 1-4), because the serial number of the physical stack port bound to IRF-port 1 must be smaller than that of the physical stack port bound to IRF-port 2. Therefore, you cannot bind IRF-port 1 to physical stack port 2, and IRF-port 2 to physical port 1.

Note

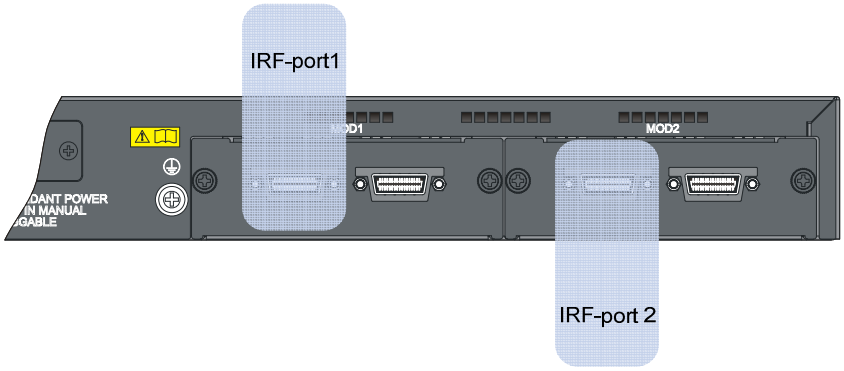
- If only one single-port interface module is installed, the device can be used only as Switch B or Switch C in Figure 1-3, that is, the device should be at either end of a bus connection.
- In this situation, because only one logical stack port is needed on Switch B or Switch C, IRF-port 2 or IRF-port 1 can be bound to any physical port on the device.

---

2) Stack port correspondence for two interface modules

- Correspondence in non-aggregate mode

Figure 1-5 Correspondence in non-aggregate mode for two interface modules



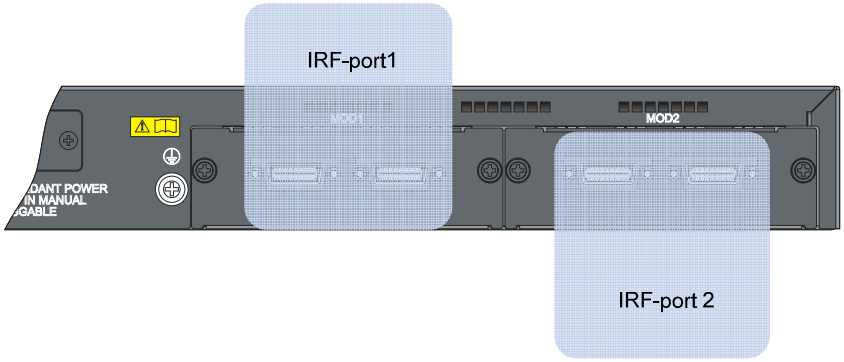
When two dual-port interface modules are installed, if the correspondence is not in the aggregate mode, you can bind a logical stack port to any physical stack port (Figure 1-5 only shows one possibility). However, you must ensure that the serial number of the physical stack port bound to IRF-port 1 is smaller than that of the physical stack port bound to IRF-port 2, namely, the physical stack port bound to IRF-port 2 should be located on the right side of the physical stack port bound to IRF-port 1. The two physical stack ports bound to the logical stack ports can be located either on one interface module or on different interface modules.

Note

- If two single-port interface modules are installed, you need to bind IRF-port 1 to physical stack port 1, and IRF-port 2 to physical stack port 3.
- If one dual-port interface module and one single-port interface module are installed, the correspondence is the same with that when you install two dual-port interface modules. In this situation, IRF-port 2 or IRF-port 1 can be bound to any physical port on the device, because only one logical stack port is needed on Switch B or Switch C.

- Correspondence in aggregate mode

Figure 1-6 Correspondence in aggregate mode for two interface modules



Because the two physical stack ports bound to an aggregate stack port must be located on the same interface module, two logical stack ports (that is, two aggregate stack ports) can only be bound to the two physical stack ports on each of the two interface modules respectively (as shown in Figure 1-6). In

addition, you can only bind IRF-port 1 to physical stack ports 1 and 2, and IRF-port 2 to physical ports 3 and 4.

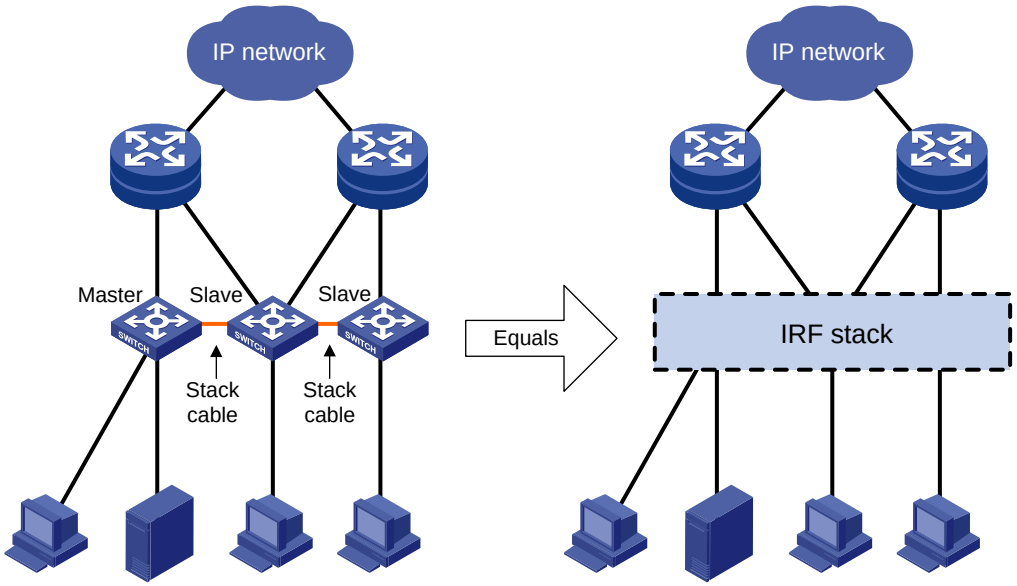
Note

If one dual-port interface module and one single-port interface module are installed, you can bind two physical stack ports on the dual-port interface module to the logical stack port in aggregate mode, and bind the physical stack port on the single-port interface module to the other logical stack port in non-aggregate mode. In this situation, IRF-port 2 or IRF-port 1 can be bound to any physical port on the device, because only one logical stack port is needed on Switch B or Switch C.

Application and Advantages

Typically, you can use an IRF stack in the distribution layer; and you can also apply it in the access layer. An IRF stack is a single logical device to the users and devices of the upper layer and lower layer, as shown in Figure 1-7.

Figure 1-7 IRF stack networking



IRF stack has the following advantages:

- Simple management

After an IRF stack is formed, you can log in to the IRF stack system by connecting to a port of any stack member. When you log in to the stack, actually you log in to the master device of the stack. You can manage all stack members by configuring the master, for example, allocating IP addresses to the members, interconnecting the members, and running routing protocols.

- Powerful network expansion capability

By adding member devices, you can increase the number of stack ports, expand network bandwidth, and improve processing capability of the stack system.

- High reliability

Not only the physical stack ports of members can be aggregated, but also the physical links between the stack system and the upper or lower layer devices can be aggregated, and thus the reliability of the stack system is increased through the link backup.

The stack system comprises multiple member devices: the master runs, manages and maintains the stack, whereas the slaves process services as well as function as the backups. When the master fails, the stack system elects a new master immediately to prevent service interruption and implement 1:N backup.

IRF Stack Working Process

IRF stack management can be divided into three stages: topology collection, role election, and stack maintenance.

Topology Collection

Each device in a stack exchanges hello packets with the directly connected neighbors to collect topology of the entire stack. The hello packets carry topology information, including stack port connection states, member IDs, priorities, and bridge MAC addresses.

Each member records its known topology information locally. At the initiation of the collection, the members record their own topology information. When a stack port of a member becomes up, the member sends its known topology information from this port periodically. Upon receiving the topology information, the directly connected neighbor updates the local topology information.

The collection process lasts for a period of time. When all members have obtained the complete topology information (known as topology convergence), the stack will enter the next stage: role election.

Role Election

A stack is composed of multiple member devices; each member has a role, which is either master or slave. The process of defining the role of stack members is role election.

Role election is held when the topology is instable, such as, forming a stack, adding a new member, stack split, or stack merge. The master is elected according to the following principles one by one, until the only winner is found out:

- The current master wins, even if a new member has a higher priority.
- A member with a higher priority wins.
- A member with the longest system up-time wins.
- A member with the lowest bridge MAC address wins.

In this stage, member ID collision, software version loading and stack merging are also handled, which are discussed in the later sections.

When a device is booted, it first collects topology information and then participates in the role election. After that, the stack system can run normally. When the role election is finished, the stack enters the next stage: stack maintenance.

Note

- Stack merge: The process of connecting two existing IRF stacks with stack cables. After the mergence, stack election is held, and members of the loser side reboot and join the winner side as slaves.
- Stack split: In an IRF stack, the failure of stack cables or power-off of a member causes physical disconnection between two devices, and the process is stack split.
- Member number restriction: The number of stack members has an upper limit, which may vary with device models.

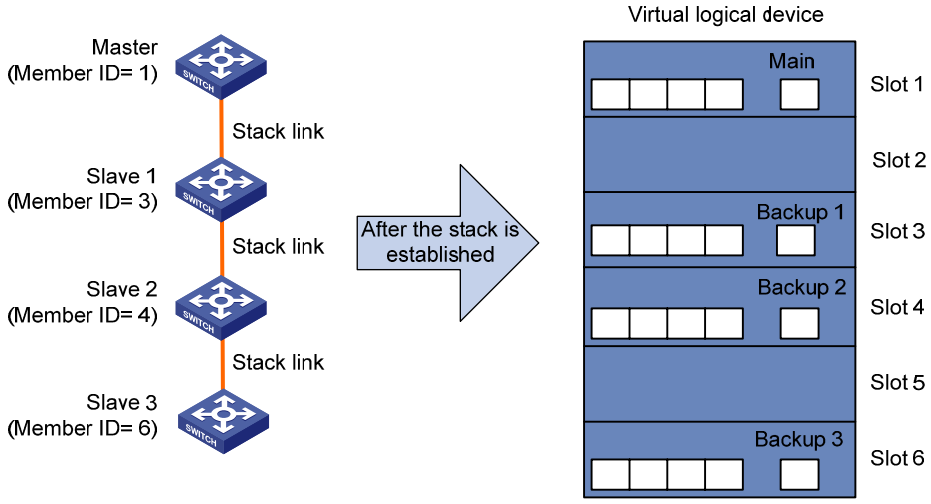
Stack Management

Member ID

A stack system uses member IDs to uniquely identify and manage member devices. Because a stack of centralized devices functions like a logical distributed device, each member device is a board of the logical distributed device: the master is the active switching and routing processing unit (SRPU), and a slave is a standby SRPU and functions like a line processing unit (LPU). Therefore, the member ID is also called the slot number.

As shown in Figure 1-8, a stack system comprises four members, which are numbered 1, 3, 4 and 6. After the stack is established, the stack system functions like a distributed device: slots 1, 3, 4 and 6 are inserted with boards, and each board has its own power supply unit (PSU), fan, CPU, console port and Ethernet interfaces.

Figure 1-8 IRF stack



Interface name

For a device operating independently (that is, the device does not belong to any stack), its interface name is in the following format: member ID/slot number/interface serial number, where

- By default, member ID is 1.
- After a device leaves a stack, it continues using the member ID when it was in the stack as its device ID.

- Subslot number is the number of the slot in which the LPU resides. For a box-type device, LPUs are fixed on the device, so the slot number is a fixed value. On the Switch 4800G series, the subslot on the front panel is numbered 0, and subslots of the two expansion slots on the rear panel are numbered 1 and 2 from left to right.
- Interface serial number is dependent on the number of interfaces supported by the device. View the silkscreen on the LPU for the number of supported interfaces.

For example, GigabitEthernet 1/0/1 is an interface on the independently operating device **Sysname**. To set the link type of GigabitEthernet 1/0/1 to trunk, perform the following steps:

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] port link-type trunk
```

For a stack member, the interface name also adopts the previously introduced format: member ID/slot number/interface serial number, where

- The member ID identifies the stack member on which the interface resides
- Meaning and value of the subslot number and the interface serial number are the same as those on an independently operating device.

For example, GigabitEthernet 1/0/1 is an interface on stack member slave 6 (member ID is 6). To set the link type of GigabitEthernet 1/0/1 to trunk, perform the following steps:

```
<Master> system-view
[Master] interface gigabitethernet 6/0/1
[Master-GigabitEthernet6/0/1] port link-type trunk
```

Configuration file management

1) Configuration file synchronization

IRF stack uses a strict configuration file synchronization mechanism to ensure that devices in a stack can work as a single device on the network, and to ensure that after the master fails, the other devices can operate normally.

- When a slave starts up, it automatically finds out the master, synchronizes the master's configuration file, and executes the configuration file; if all devices in a stack start up simultaneously, the slaves synchronize the master's initial configuration file and execute it.
- When the IRF stack operates normally, all your configurations will be recorded into the current configuration file of the master, and are synchronized to each device in the stack; when you save the current configuration file of the master as the initial configuration file by using the **save** command, all slaves execute the same saving operation to make the initial configuration files of all devices consistent.

Through the real-time synchronization, all devices in the stack keep the same configuration file. If the master fails, all the other devices can execute various functions according to the same configuration file.

2) Configuration file application

The configuration file can be divided into two parts: global configuration and port configuration. When a slave applies these two kinds of configurations of the master, it deals with them in different ways:

- Global configuration: All slaves execute the current global configuration on the master exactly, that is, all members in the stack apply the same global configuration.

- Port configuration: When a slave applies the port configuration on the master, it cares about the configuration related to its own port, for example, the slave with the member ID of 3 only cares about the configuration related to the GigabitEthernet 3/0/x port on the master. If there is a configuration related to its own port, it will apply the configuration; if not, no matter what configuration has been made to the port before the slave joins the stack, the slave will function using null-configuration.

File system name

You can use the name of the storage device to access the file system of an independently operating device. For the naming rules of a storage device, refer to the *File System Management Configuration* in the *System Volume*.

For example, flash is the storage device on the independently operating device Sysname. To back up the file **aa.cfg** under the root directory of the flash to the **test** folder, perform the following steps:

```
<Sysname> mkdir test
...
%Created dir flash:/test.

<Sysname>copy aa.cfg test/aa(20080714).cfg
Copy flash:/aa.cfg to flash:/test/aa(20080714).cfg?[Y/N]:y
..
%Copy file flash:/aa.cfg to flash:/test/aa(20080714).cfg...Done.
<Sysname> cd test
<Sysname> dir
Directory of flash:/test/

   0     -rw-      1568  Jul 14 2008 11:54:04   aa(20080714).cfg

30861 KB total (20956 KB free)
```

To access the file system of the master, use the name of the storage device; to access the file system of a slave, use the name in the following format: Member-ID#Storage-device-name.

For example:

1) To access the **test** folder under the root directory of the flash on the master, perform the following steps:

```
<Master> mkdir test
...
%Created dir flash:/test.
<Master> dir
Directory of flash:/
   0     -rw-  10105088  Apr 26 2000 13:44:57   test.app
   1     -rw-      2445  Apr 26 2000 15:18:19   config.cfg
   2     drw-         -  Jul 14 2008 15:20:35   test
30861 KB total (20961 KB free)
```

2) To create and access the **test** folder under the root directory of the flash on stack member slave 6, perform the following steps:

```
<Master> mkdir slot6#flash:/test
%Created dir slot6#flash:/test.
```

```
<Master> cd slot6#flash:/test
<Master> pwd
slot6#flash:/test
```

Or:

```
<Master> cd slot6#flash:/
<Master> mkdir test
%Created dir slot6#flash:/test.
```

3) To copy the **test.app** file on the master to the root directory of the flash on stack member slave 6, perform the following steps:

```
<Master> pwd
slot6#flash:
```

//The above information indicates that the current working path is the root directory of the flash on slave 6.

```
<Master> cd flash:/
<Master> pwd
flash:
```

// The above operations indicate that the current working path is the root directory of the flash on the master.

```
<Master> copy test.app slot6#flash:/
Copy flash:/test.app to slot6#flash:/test.app?[Y/N]:y
%Copy file flash:/test.app to slot6#flash:/test.app...Done.
```

Stack Maintenance

In an IRF stack, direct neighbors exchange hello packets periodically (the period is 200 ms). Without receiving any hello packet from a direct neighbor for ten periods, a member considers that the hello packets timed out, and the stack isolates the expired device in the topology and updates its topology database.

When a stack port of a member becomes down, the member broadcasts the information to all the other members immediately. If the stack port of the master is down, an election is triggered.

IRF Stack Configuration Task List

Before configuring an IRF stack, you need to define the roles and functions of all the members for better planning. Because the configuration of some parameters takes effect after device reboot, you are recommended to first configure parameters, power off the devices, connect devices physically, power on the devices, and finally the devices will join in the stack automatically. After an IRF stack is formed, you can configure and manage the stack by logging in to any device in the stack. The operations you make take effect on the master, and will be applied to the member devices in the stack. For easy fault location and device maintenance, the Switch 4800G switch provides slave view, where you can execute the **display**, **terminal**, and **debug** commands.

Complete the following tasks to configure IRF stack:

Task		Remarks
Configuring IRF Stack	Configuring Stack Ports	Required
	Setting a Member ID for a Device	Optional
	Specifying a Priority for a Stack Member	Required
	Specifying the Preservation Time of Stack Bridge MAC Address	Optional
	Enabling Auto Upgrade of Boot Files	Optional
	Setting the Delay Time for the Link Layer to Report a Link-Down Event	Optional
Connect the physical stack ports of devices by using stack cables (a ring connection is recommended), and then power on the devices.		
Logging In to an IRF Stack	Logging In to the Master	Required
	Logging In to a Slave	Optional

Configuring IRF Stack

Configuring Stack Ports

IRF stack can be enabled on a device only after the logical stack ports are bound with physical stack ports).

For how to bind the logical stack port and physical stack port(s) on an Switch 4800G series, see Correspondence between a logical stack port and a physical stack port.

Follow these steps to configure stack ports

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Bind physical stack ports to a logical stack port, and enable IRF stack on the current device	**irf member** *member-id* **irf-port** *irf-port-id* **port** *port-list*	Required By default, no logical stack port is configured.

Caution

- The above configuration takes effect after the reboot of the device.
- A logical stack port that is bound with multiple physical stack ports is an aggregation stack port, which increases the bandwidth and reliability on the stack port. If you specify multiple physical stack ports with the *port-list* argument, you can configure an aggregation stack port. You can configure at most two physical stack ports as an aggregation stack port for an Switch 4800G series switch, and you can only aggregate stack ports 1 and 2, and stack ports 3 and 4.
- The *irf-port-id* argument represents the logical stack port number. The *port-list* argument represents the physical stack port number. For the correspondence of stack ports, refer to Correspondence between a logical stack port and a physical stack port.
- When you insert a one-port interface module into the slot on the rear panel, if the interface module is in slot 1, the port on it will be numbered 1; and if the interface module is in slot 2, the port on it will be numbered 3.

Setting a Member ID for a Device

The member ID of a device defaults to 1. During the establishment of a stack, when the devices that form the stack have duplicated member IDs, the member ID of the master is decided first, and then the member IDs of slaves are decided one by one according to their distances to the master, that is, the nearest slave gets the smallest available ID, and the nearer slave gets the smaller available ID, and so forth; after the stack is established, if the newly added device and another member have duplicated IDs, the stack system assigns the smallest available ID for the new member. You can also set the member IDs according to network planning.

For a device that is already in a stack, you can use commands in Table 1-1 to modify the member ID of the device, and this modification will be effective after the reboot of the device.

For a device that is not in a stack, you are recommended to set its member ID in the following way:

1) Plan the member IDs in advance. You can view the member IDs of a stack, and find out an unused ID for the new device.
2) Log in to the device to be added into the stack, and change its member ID to the unused ID found out in step 1.
3) Save the current configuration. Power off the device, connect the device with stack cables and power it on. Use the configuration introduced in Configuring Stack Ports to enable IRF stack on the device and add it into the IRF stack.

Table 1-1 Set a member ID for a device

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set a member ID for a device	**irf member** *member-id* **renumber** *new-id*	Optional The member ID of a device defaults to 1

Caution

- The above setting takes effect after the reboot of the device.
- You can use the **display irf configuration** command to view the current member ID of the device and the member ID will be used after the device reboot.
- In an IRF stack, member IDs are not only used to identify devices, but also used to identify the port configurations on different member devices in the configuration file. Therefore, modifying a member ID may cause device configuration changes or even losses, so modify member ID with caution. For example, three members (of same device model) with the member IDs of 1, 2 and 3 are connected to a stack port. Suppose that each member has several ports: change the member ID of device 2 to 3, change that of device 3 to 2, reboot both devices, and add them into the stack again. Then device 2 will use the original port configurations of device 3, and device 3 will use those of device 2.

Specifying a Priority for a Stack Member

Each stack member has a priority. During the master election, a member with the greatest priority will be elected as the master.

The priority of a device defaults to 1. You can modify the priority through command lines. The greater the priority value, the higher the priority. A member with a higher priority is more likely to be a master, and more likely to preserve its ID in a member ID collision.

Follow these steps to specify a priority for a stack member:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Specify a priority for a stack member	**irf member** *member-id* **priority** *priority*	Optional The priority of a stack member defaults to 1

Note

The setting of priority takes effect right after your configuration without the need of rebooting the device.

Specifying the Preservation Time of Stack Bridge MAC Address

A device uses the bridge MAC address when it communicates with the outside as a network bridge. A bridge device on the network has its unique bridge MAC address. Some Layer 2 protocols (like LACP) use bridge MAC addresses to identify different devices. During the forwarding of Layer 2 packets, if the destination MAC address of a packet is the bridge MAC address of a device, it means that the packet is sent to this device.

In an IRF stack, the bridge MAC address of a member device is called member bridge MAC address. The stack communicates with the outside as a single device; therefore, it also has a bridge MAC

address, which is called the stack bridge MAC address. Typically, a stack uses the bridge MAC address of the master device as the stack bridge MAC address.

You are recommended to configure the preservation time of stack bridge MAC address properly, otherwise, network problems will occur:

- If a master leaves a stack to join another stack or to operate independently and the stack is configured to preserve the bridge MAC address permanently, bridge MAC address collision occurs and thus causes network communication problem.
- If the master leaves the stack because of reboot or link failure and the stack is configured to change the stack bridge MAC address as soon as the master leaves, the unnecessary switch of bridge MAC address occurs and thus causes flow interruption.

Therefore, configure the preservation time stack bridge MAC address according to your network status:

- Preserve for six minutes: After the master leaves, the bridge MAC address will not change within six minutes. If the master does not come back after six minutes, the stack system will use the bridge MAC address of the newly elected master as that of the stack.
- Preserve permanently: No matter the master leaves the stack or not, the stack bridge MAC address remains unchanged.
- Not preserved: As soon as the master leaves, the system will use the bridge MAC address of the newly elected master as that of the stack.

Follow these steps to specify preservation time of stack bridge MAC address:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Configure the stack bridge MAC address to be preserved permanently after the master leaves	**irf mac-address persistent always**	Optional By default, stack bridge MAC address is preserved for 6 minutes after the master leaves.
Specify the preservation time of the stack bridge MAC address as 6 minutes after the master leaves	**irf mac-address persistent timer**	
Configure that the stack bridge MAC address changes as soon as the master leaves	**undo irf mac-address persistent**	

Caution

The change of the bridge MAC address may cause a temporary flow interruption.

Enabling Auto Upgrade of Boot Files

If this function is disabled, when the boot files of slaves and that of the master are in different versions, the new member or the member with a low priority will not boot normally. You need to update the device version manually and add the device into the stack again.

If this function is enabled, as soon as a device is added into a stack, the system compares its software version with that of the master. If the versions are not consistent, the device downloads the boot file

from the master automatically, reboots with the new boot file, and joins the stack again. If the downloaded boot file and the local file have duplicate file names, the local file is overwritten.

Follow these steps to enable auto upgrade of boot files in an IRF stack:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Enable auto upgrade of boot files in an IRF stack	**irf auto-update enable**	Optional Enabled by default

Caution

- Although IRF stack supports the auto upgrade of boot files, to shorten the time for stack establishment and reduce the influences caused by the stack establishment to the network, you are recommended to ensure that the device and the stack master have the same software version before adding a device into an IRF stack.
- After loading the master's boot file automatically, a slave configures the file as the boot file for the next boot and reboots automatically.
- Because system boot file occupies large memory space, to make the auto upgrade succeed, ensure that there is enough space on the storage media of the slave.

Setting the Delay Time for the Link Layer to Report a Link-Down Event

During the suppression time, the system cannot be aware of the switch between stack link states; after the suppression time, the link layer reports the link state changes to the system. Use this function to avoid the reboots of devices caused by the frequent link state changes of stack ports in a short time (for example, a stack splits and then merges quickly), preventing service interruption.

Follow these steps to set the delay time for the link layer to report a link-down event of a stack:

To do…	Use the command…	Remarks
Enter system view	**system-view**	—
Set the delay time for the link layer to report a link-down event of a stack	**irf link-delay** *interval*	Optional The function is disabled by default.

Caution

Do not set the time interval to a very long time; otherwise, the stack system will not be aware of the stack topology changes in time and thus the service will be recovered slowly.

Logging In to an IRF Stack

Logging In to the Master

After an IRF stack is formed, you can access the console of the stack system through the AUX or console port of any member device. Configure an IP address for the VLAN interface of a member device and make sure that the route is reachable, and then you can access the stack system remotely through Telnet, Web, or SNMP.

When you log in to the stack, actually you log in to the master device of the stack. The master is the configuration and control center of a stack. After you configure the stack on the master, the stack system synchronizes the configurations to the slaves.

Logging In to a Slave

When you log in to a stack, actually you log in to the master device of the stack. The operation interface of the access terminal displays the master console. However, the device can redirect you to a specified slave device. After you are redirected to a slave device, the user access terminal displays the console of the slave device instead of that of the master device. The system enters user view of the salve device and the command prompt is changed to <Sysname-member ID>, for example, <Sysname-2>. What you have input on the access terminal will be redirected to the specified slave device for processing. At present, only the following commands are allowed to be executed on a slave device:

- **display**
- **quit**
- **return**
- **system-view**
- **debugging**
- **terminal debugging**
- **terminal trapping**
- **terminal logging**

You can press **Ctrl+K** or use the **quit** or **return** command to return to the master console. At this time, the master console is reactivated, and therefore it can output system information and logs.

Follow the step below to log in to the specified slave device:

To do…	Use the command…	Remarks
Log in to the specified slave device of a stack	**irf switch-to** *member-id*	Required By default, you actually log in to the master device of a stack when you log in to the stack. Available in user view

Note

Because users' login to the stack system occupies large memory space, a stack system allows at most six users to log in at the same time. The permitted login user types are console and virtual type terminal (VTY).

Displaying and Maintaining IRF Stack

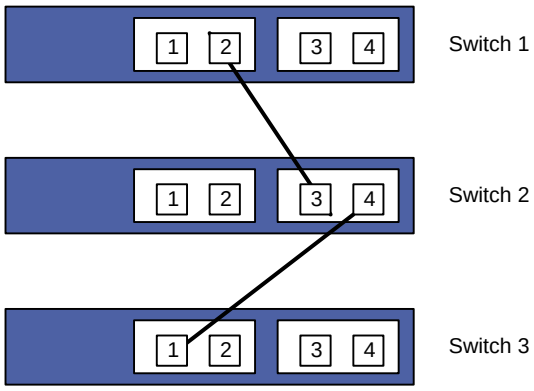
To do…	Use the command…	Remarks
Display related information of the stack	**display irf**	Available in any view
Display topology information of the stack	**display irf topology**	Available in any view
Display the pre-configurations of all members of the stack (The pre-configuration takes effect after the reboot of the device.)	**display irf configuration**	Available in any view
Display the master/slave switchover states of stack members	**display switchover state** [*member-id*]	Available in any view

IRF Stack Configuration Example

Network requirements

Three Switch 4800G series switches in an IRF stack form a bus connection. Their member IDs are 1, 2, and 3, as shown in Figure 1-9.

Figure 1-9 Network diagram for IRF stack



Configuration procedure

1) The three devices are not connected. Power them on and configure them separately.

Configure Switch 1.

```
<Switch1> system-view
[Switch1] irf member 1 renumber 1
Warning: Renumbering the switch number may result in configuration change or loss.
Continue?[Y/N]:y
[Switch1] irf member 1 irf-port 1 port 2
```

Configure Switch 2.

```
<Switch2>system-view
[Switch2] irf member 1 renumber 2
Warning: Renumbering the switch number may result in configuration change or loss.
Continue?[Y/N]:y
[Switch2] irf member 1 irf-port 1 port 2
```

```
[Switch2] irf member 1 irf-port 2 port 3
```

Configure Switch 3.

```
<Switch3> system-view
[Switch3] irf member 1 renumber 3
Warning: Renumbering the switch number may result in configuration change or loss.
Continue?[Y/N]:y
[Switch3] irf member 1 irf-port 2 port 3
```

2) Power off the three devices. Connect them as shown in Figure 1-9 with stack cables. Power them on, and the stack is formed.

Table of Contents

1 GR Overview

Go to these sections for information you are interested in:

Note

Throughout this chapter, the term router refers to a router in a Layer 3 switch running routing protocols.

Introduction to Graceful Restart

Graceful Restart ensures the continuity of packet forwarding when a protocol restarts.

The mechanism of Graceful Restart works as follows: When the protocol on a device restarts, the device will notify its neighbors to temporarily preserve the routing information and adjacency relationship with the device. After the protocol restarts, the neighbors will help the restarting device to update information (including various topology, routing and session information maintained by routing/MPLS related protocols that support Graceful Restart) and to restore it to the state prior to the restart in minimal time. No route flapping occurs during the restart, the packet forwarding path remains the same, and the whole system can forward data continuously. Hence, it is called “Graceful Restart”.

Basic Concepts in Graceful Restart

A router with the Graceful Restart function enabled is called a Graceful Restart-capable router. It can perform a Graceful Restart when its routing protocol restarts, ensuring consistent forwarding services. Routers that are not Graceful Restart capable will follow the normal restart procedures after a routing protocol restart.

- GR Restarter: Graceful restarting router, the router whose routing protocol has restarted due to administrator instructions or network failure. It must be Graceful Restart capable.
- GR Helper: The neighbor of the GR Restarter. It helps the GR Restarter to retain the routing information. It must be Graceful Restart capable.
- GR Session: A Graceful Restart session, which is the negotiation between the GR Restarter and the GR Helper. A GR session includes restart notification and communications during restart. Through this session, GR Restarter and GR Helper can know the GR capability of each other.
- GR Time: The time taken for the GR Restarter and the GR Helper to establish a session between them. Upon detection of the down state of a neighbor, the GR Helper will preserve the topology or routing information sent from the GR Restarter for a period as specified by the GR Time.

Graceful Restart Communication Procedure

Configure a device as GR Restarter in a network. This device and its GR Helper must support GR or be GR capable. Thus, when GR Restarter restarts, its GR Helper can know its restart process.

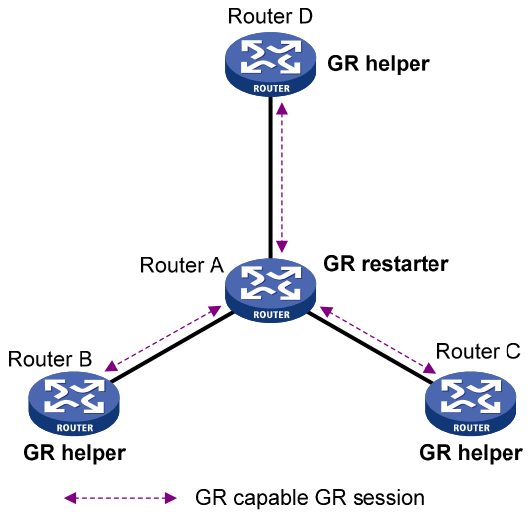
Note

In some cases, GR Restarter and GR Helper can replace each other.

The communication procedure between the GR Restarter and the GR Helper works as follows:

1) Establishing a GR session.

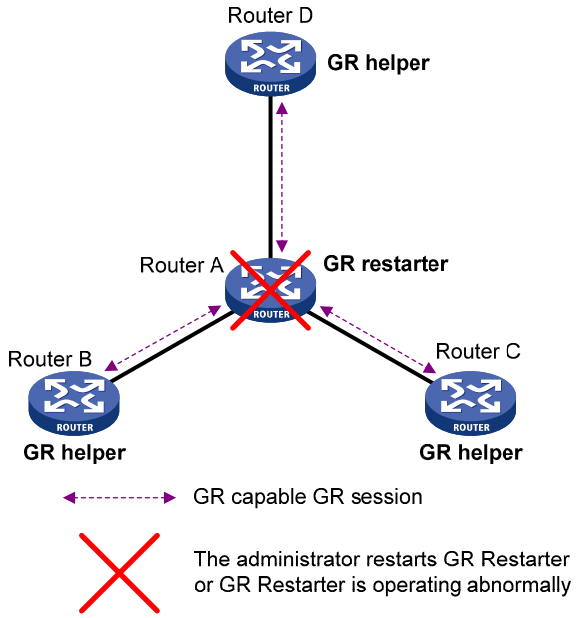
Figure 1-1 A GR session is established between the GR Restarter and the GR Helper



As illustrated in Figure 1-1, Router A works as GR Restarter, Router B, Router C and Router D are the GR Helpers of Router A. A GR session is established between the GR Restarter and the GR Helper.

2) Restarting GR Restarter

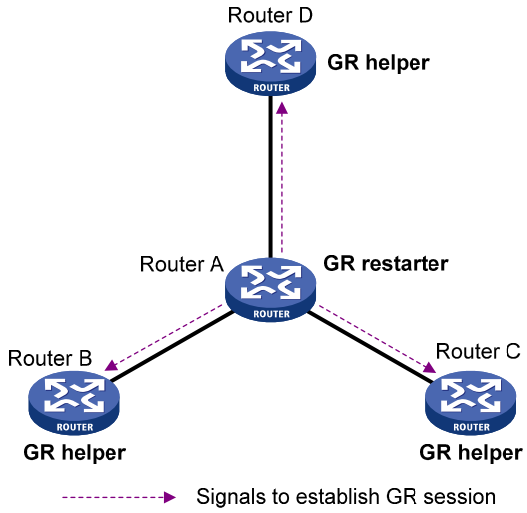
Figure 1-2 Restarting process for the GR Restarter



As illustrated in Figure 1-2. The GR Helper detects that the GR Restarter has restarted its routing protocol and assumes that it will recover within the GR Time. Before the GR Time expires, the GR Helper will neither terminate the session with the GR Restarter nor delete the topology or routing information of the latter.

3) Signaling to GR Helper

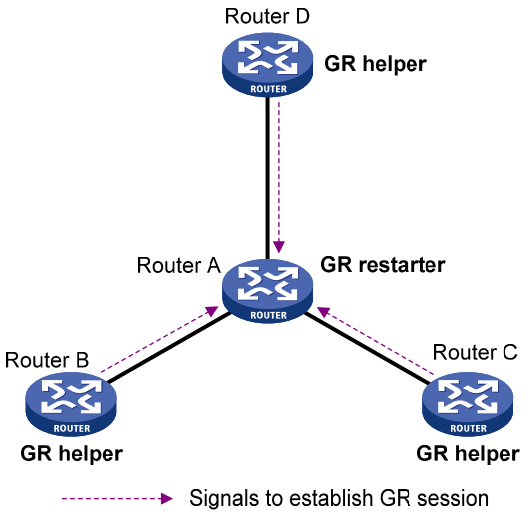
Figure 1-3 The GR Restarter signals to the GR Helper(s) after restart



As illustrated in Figure 1-3, after the GR Restarter has recovered, it will signal to all its neighbors and reestablish GR Session.

4) Obtaining topology and routing information

Figure 1-4 The GR Restarter obtains topology and routing information from the GR Helper



As illustrated in Figure 1-4, the GR Restarter obtains the necessary topology and routing information from all its neighbors through the GR sessions between them and calculates its own routing table based on this information.

Graceful Restart Mechanism for Several Commonly Used Protocols

Comware supports Graceful Restart based on Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS).

For the implementation and configuration procedure of the Graceful Restart mechanism of the above protocols, refer to *BGP Configuration*, *OSPF Configuration*, and *IS-IS Configuration* in the *IP Routing Volume*.

Table of Contents

1 Automatic Configuration

When configuring automatic configuration, go to these sections for information you are interested in:

- Introduction to Automatic Configuration
- Typical Networking of Automatic Configuration
- How Automatic Configuration Works

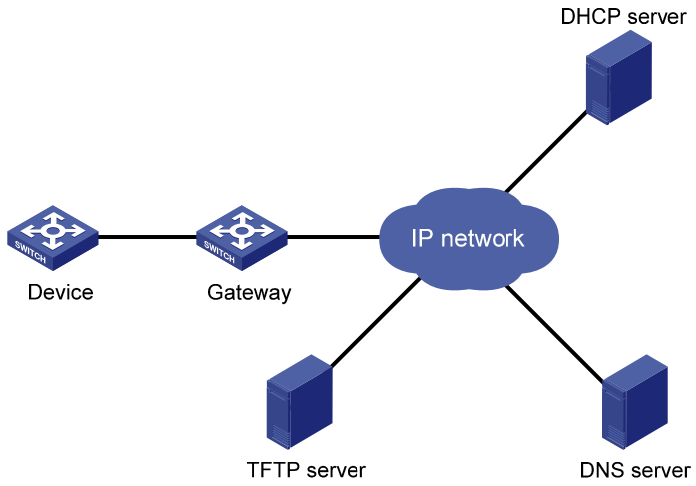
Introduction to Automatic Configuration

Automatic configuration enables a device to automatically obtain and execute the configuration file when it starts up without loading the configuration file.

Automatic configuration simplifies network configuration, facilitating centralized management of devices. Currently, enterprise networks are facing the problems of large distribution of devices and less administrators, resulting in the huge cost for administrators to manually configure each device. With the automatic configuration function, network administrators can save the configuration files on a specified server and the device can automatically obtain and execute the configuration files, therefore greatly reducing the workload of administrators.

Typical Networking of Automatic Configuration

Figure 1-1 Network diagram for automatic configuration



As shown in Figure 1-1, the device implements automatic configuration with the cooperation of a DHCP server, TFTP server and DNS server:

- DHCP server: Assigns an IP address, configure file name, TFTP server IP address, and DNS server IP address for the device that performs automatic configuration.
- TFTP server: Saves files needed in automatic configuration. A device obtains files needed from a TFTP server, for example, network intermediate file and the configuration file of the device.
- DNS server: Used for IP address-to-host name resolution. A device that performs automatic configuration can resolve an IP address to a host name through a DNS server to get the configuration file with the name **hostname.cfg** from a TFTP server; if the device gets the domain

name of the TFTP server from a DHCP response, the device can also resolve the domain name of the TFTP server to the IP address of the TFTP server through the DNS server.

If the DHCP server, TFTP server, DNS server, and the device that performs automatic configuration are not in the same segment, you need to configure DHCP relay on a device working as a gateway.

How Automatic Configuration Works

Basically, automatic configuration works in the following ways:

1) When a device starts up without loading any configuration file, the system sets the first active interface (if an active Layer 2 Ethernet interface exists, this first interface is a virtual interface corresponding with the default VLAN) as the DHCP client to request from the DHCP server for parameters, such as an IP address and name of a TFTP server, IP address of a DNS server, and the configuration file name.
2) After getting related parameters, the device will send a TFTP request to obtain the configuration file from the specified TFTP server for system initialization. If the client cannot get such parameters, it performs system initialization without loading any configuration file.

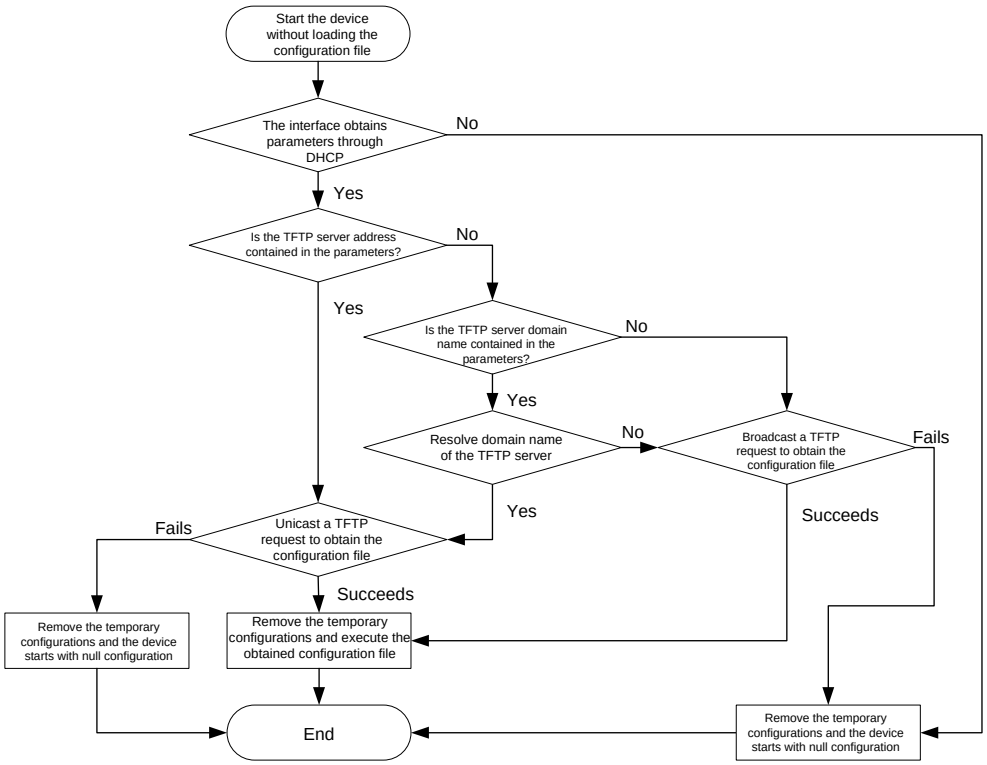
Note

- To implement auto-configuration, you need to configure some parameters on the DHCP server, DNS server and TFTP server, but you do not need to perform any configuration on the device that starts up without loading any configuration file. The configuration mode depends on the device model; it is omitted here.
- If you need to use the automatic configuration function, you are recommended to connect only the interfaces needed in automatic configuration to the network.

Work Flow of Automatic Configuration

The work flow of automatic configuration is as shown in Figure 1-2.

Figure 1-2 Work flow of automatic configuration



Obtaining the IP Address of an Interface and Related Information Through DHCP

Obtaining an IP address

When a device starts up without loading the configuration file, the system automatically configures the first active interface (if an active Layer 2 Ethernet interface exists, this first interface is a virtual interface corresponding with the default VLAN) of the device as obtaining its IP address through DHCP. The device broadcasts a DHCP request through this interface. The Option 55 field specifies the information (for example, the configuration file name, domain name and IP address of the TFTP server and DNS server needed for obtaining the automatic configuration files) that the device can obtain from the DHCP server.

Upon successfully obtaining its IP address through DHCP, the device resolves the Option 67 (or the file field, configuration file name) field, Option 66 (domain name of the TFTP server) field, Option 150 (IP address of the TFTP server) field and Option 6 (IP address of the DNS server) field. If failing to obtain its IP address, the device removes the temporary configuration and starts up without loading the configuration file.

Note

- The configuration file name is saved in the Option 67 or file field of the DHCP response. The device first resolves the Option 67 field; if this field contains the configuration file name, the device does not resolve the file field; otherwise, it resolves the file field.
- Temporary configuration contains two parts: the configuration on the interface where automatic configuration is performed when the device starts up with default configuration; and the executed **ip host** command when the device is resolving the network intermediate file (For the detailed description of the **ip host** command, refer to *Domain Name Resolution Commands* in the *IP Services Volume.*). Removal of the temporary configuration is to execute the **undo** commands.
- For the detailed introduction to DHCP, refer to *DHCP Configuration* in the *IP Services Volume.*

Principles for selecting an address pool on the DHCP server

The DHCP server selects IP addresses and other network configuration parameters from an address pool when assigning an IP address to a client. DHCP supports two mechanisms for IP address allocation.

- Dynamic address allocation: The DHCP server assigns an IP address and other configuration parameters in an address pool to a client.
- Manual address allocation: The DHCP server will select an address pool where an IP address is statically bound to the MAC address or ID of the client and assign the statically bound IP address and other configuration parameters to the client.

You can configure an address allocation mode as needed:

- Different devices with the same configuration file: You can configure dynamic address allocation on the DHCP server to assign IP addresses and the same configuration parameters (for example, configuration file name) to the devices. If this address allocation mode is adopted, the configuration file can only contain common configurations of the devices, and the specific configurations of each device need to be performed in other ways. For example, you need to specify to enable Telnet on a device through the configuration file obtained in automatic configuration and create a local user to facilitate the administrator to Telnet to each device to perform specific configurations (for example, configure the IP address of each interface).
- Different devices with different configuration files: You need to configure an address pool where an IP address is statically bound to the MAC address or ID of the client, to ensure that a specific client can be assigned with a fixed IP address and other configuration parameters. Through this address allocation mode, you can specify different configuration commands for each device, without the need to configure the device through other modes.

Note

You need to configure a client ID (when a device works as the DHCP client, it uses the client ID as its ID) of the static binding when you configure manual address allocation. Therefore, you need to obtain the client ID in this way: start the device that performs automatic configuration, enable the interface that performs automatic configuration to obtain its IP address through DHCP, after the IP address is successfully obtained, use the **display dhcp server ip-in-use** command to display address binding information on the DHCP server, thus to obtain the client ID of the device.

Obtaining the Configuration File from the TFTP Server

Configuration file type

The device can obtain the following types of configuration file from the TFTP server with the automatic configuration function enabled:

- The configuration file specified by the Option 67 or file field in the DHCP response
- The intermediate file, with the file name as **network.cfg**, used to save the mapping between the IP address and the host name. The mapping is defined in the following format:

ip host *hostname ip-address*

For example, the intermediate file can include the following:

```
ip host host1 101.101.101.101
ip host host2 101.101.101.102
ip host client1 101.101.101.103
ip host client2 101.101.101.104
```

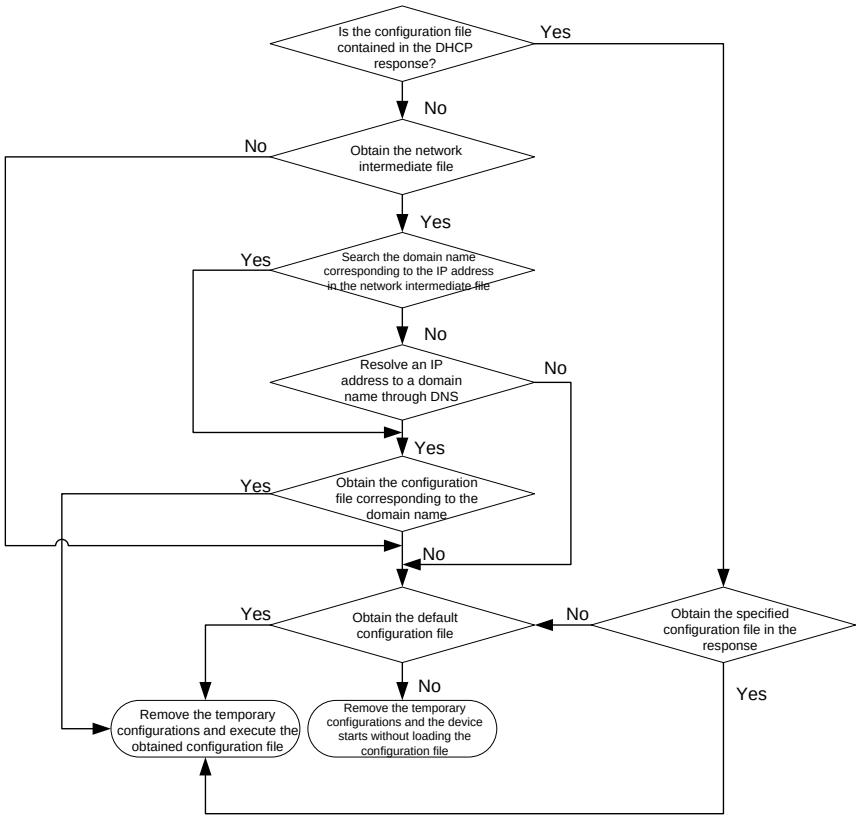
Caution

- There must be a space before the keyword **ip host**.
- The host name saved in the intermediate file must be the same with the configuration file name of the host. This host name is not the one saved in the DNS server, and their names can be the same or different.

- The configuration file corresponding with the host name of the device, with its file name as **hostname.cfg**. For example, if the host name of the device is **aaa**, then the configuration file name is **aaa.cfg**.
- Default configuration file, with the name as **device.cfg**.

Obtaining the configuration file

Figure 1-3 Obtain the configuration file



The device obtains the configuration file from the TFTP server based on its resolution of the configuration file name in the DHCP response:

- If the DHCP response contains information such as configuration file name, the device requests the specified configuration file from the TFTP server.
- If no information such as configuration file name is contained in the DHCP response, the device should obtain its host name first and then requests the configuration file corresponding with the host name. The device can obtain its host name in two steps: obtaining the intermediate file from the TFTP server and then searching in the intermediated file for its host name corresponding with the IP address of the device; if fails, the device obtains the host name from the DNS server.
- If the device fails to obtain the specified configuration file and resolve its host name or fails to obtain the configuration file corresponding with the host name, it requests the default configuration file from the TFTP server.

Sending mode of a TFTP request

The device selects the sending mode of the TFTP request based on its resolution of the TFTP server's domain name and IP address in the DHCP response:

- If a legitimate TFTP server IP address is contained in the DHCP response, the device unicasts a TFTP request to the TFTP server and does not resolve the domain name of the TFTP server. Otherwise, the device resolves the TFTP domain name.
- If a legitimate TFTP server domain name is contained in the DHCP response, the device resolves the IP address of the TFTP server through DNS server. If succeeds, the device unicasts a TFTP request to the TFTP server; if fails, the device broadcasts a TFTP request to the TFTP server.

- If the IP address and the domain name of the TFTP server are not contained in the DHCP response or they are illegitimate, the device broadcasts a TFTP request to the TFTP server.

 **Note**

- When broadcasting a TFTP request, the device obtains the configuration file from the TFTP server who responds the first. If the required configuration file does not exist on the TFTP server, then obtaining the configuration file fails, and the device removes the temporary configuration and starts up without loading the configuration file.
- When the device broadcasts a TFTP request to the TFTP server, you need to configure the UDP Helper function on a gateway to transfer broadcasts to unicasts and forwards the unicasts to the specified TFTP server if the device performs the automatic configuration and the TFTP server are not in the same segment because broadcasts can only be transmitted in a segment. For the detailed description of the UDP Helper function, refer to *UDP Helper Configuration* in the *IP Services Volume*.

Executing the Configuration File

Upon successfully obtaining the configuration file, the device removes the temporary configuration and executes the obtained configuration file; otherwise, it removes the temporary configuration and starts up without loading the configuration file.

 **Note**

After the device executes the configuration file obtained, the configuration file will be deleted. Therefore, you are recommended to save the configuration using the **save** command; otherwise, the device needs to perform the automatic configuration function after system reboot. For the detailed description of the **save** command, refer to *File System Management Configuration* in the *System Volume*.

Table of Contents

1 IPC Configuration

When configuring IPC, go to these sections for information you are interested in:

- IPC Overview
- Enabling IPC Performance Statistics
- Displaying and Maintaining IPC

IPC Overview

Introduction to IPC

Inter-Process Communication (IPC) is a reliable communication mechanism among different nodes. The following are the basic concepts in IPC.

Node

An IPC node is an entity supporting IPC; it is an independent processing unit. In actual application, an IPC node corresponds to one CPU.

- One centralized device has only one CPU, therefore corresponding to one node.
- An Intelligent Resilient Framework (IRF) is an interconnection of several centralized devices, with each member device corresponding to one node. Therefore, an IRF corresponds to multiple nodes.
- Typically a distributed device is available with multiple boards, each having one CPU, some boards are available with multiple CPUs. Some distributed devices may be available with multiple CPUs, for example service CPU and OAM CPU. Therefore, a distributed device corresponds to multiple nodes.

Therefore, in actual application, IPC is mainly applied on an IRF or distributed device; it provides a reliable transmission mechanism between different devices and boards.

Link

An IPC link is a connection between any two IPC nodes. There is one and only one link between any two nodes for packet sending and receiving. All IPC nodes are fully connected.

IPC links are created when the system is initialized: When a node starts up, it sends handshake packets to other nodes; a connection is established between them if the handshake succeeds.

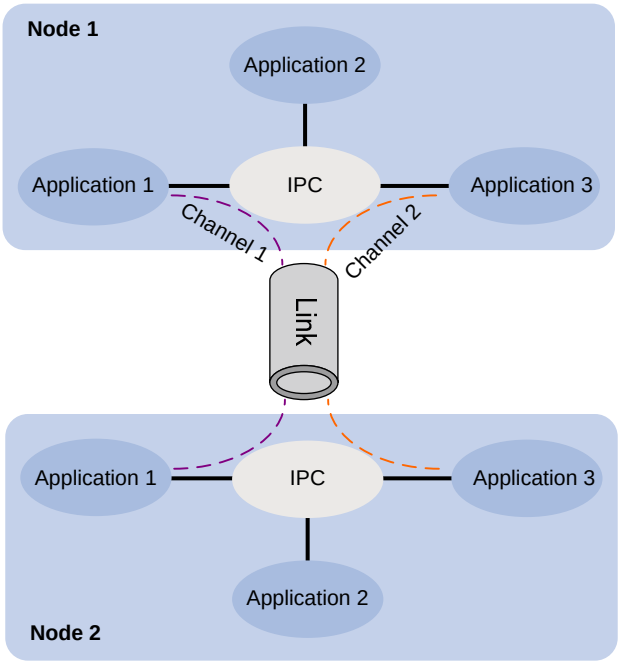
The system identifies the link connectivity between two nodes using link status. An IPC node can have multiple links, each having its own status.

Channel

A channel is a communication interface for an upper layer application module of a node to communicate with an application module of a peer node. Each node assigns a locally unique channel number to each upper layer application module.

Data of an upper layer application module is sent to the IPC module through a channel, and the IPC module sends the data to a peer node through the link. The relationship between a node, link and channel is as shown in Figure 1-1.

Figure 1-1 Relationship between a node, link and channel



Packet sending modes

IPC supports three packet sending modes: unicast, multicast (broadcast is considered as a special multicast), and mixcast, each having a corresponding queue. The upper layer application modules can select one as needed.

- Unicast: packet sending between two single nodes.
- Multicast: packet sending between a single node and multiple nodes. To use the multicast mode, a multicast group needs to be created first. Multicasts will be sent to all the nodes in the multicast group. An application can create multiple multicast groups. The creation and deletion of a multicast group and multicast group members depend on the application module.
- Mixcast, namely, both unicast and multicast are supported.

Enabling IPC Performance Statistics

When IPC performance statistics is enabled, the system collects statistics for packet sending and receiving of a node in a specified time range (for example, in the past 10 seconds, or in the past 1 minute). When IPC performance statistics is disabled, statistics collection is stopped. At this time, if you execute the **display** command, the system displays the statistics information at the time when IPC performance statistics was disabled.

Follow these steps to enable IPC performance statistics:

To do…	Use the command…	Remarks
Enable IPC performance statistics	**ipc performance enable** { **node** *node-id* | **self-node** } [**channel** *channel-id*]	Required Disabled by default Available in user view

Displaying and Maintaining IPC

To do…	Use the command…	Remarks
Display IPC node information	**display ipc node**	Available in any view
Display channel information of a node	**display ipc channel** { **node** *node-id* | **self-node** }	
Display queue information of a node	**display ipc queue** { **node** *node-id* | **self-node** }	
Display multicast group information of a node	**display ipc multicast-group** { **node** *node-id* | **self-node** }	
Display packet information of a node	**display ipc packet** { **node** *node-id* | **self-node** }	
Display link status information of a node	**display ipc link** { **node** *node-id* | **self-node** }	
Display IPC performance statistics information of a node	**display ipc performance** { **node** *node-id* | **self-node** } [**channel** *channel-id*]	
Clear IPC performance statistics information of a node	**reset ipc performance** [**node** *node-id* | **self-node**] [**channel** *channel-id*]	Available in user view