3Com® Switch 8800 Family
Configuration Guide

Advanced Software Version 5

Switch 8807
Switch 8810
Switch 8814

www.3Com.com
Part Number: 10016063 Rev. AA
Published: June, 2007

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

3Com Corporation reserves the right to revise this documentation and to make changes in content from time to time without obligation on the part of 3Com Corporation to provide notification of such revision or change.

3Com Corporation provides this documentation without warranty, term, or condition of any kind, either implied or expressed, including, but not limited to, the implied warranties, terms or conditions of merchantability, satisfactory quality, and fitness for a particular purpose. 3Com may make improvements or changes in the product(s) and/or the program(s) described in this documentation at any time.

If there is any software on removable media described in this documentation, it is furnished under a license agreement included with the product as a separate document, in the hard copy documentation, or on the removable media in a directory file named LICENSE.TXT or !LICENSE.TXT. If you are unable to locate a copy, please contact 3Com and a copy will be provided to you.

UNITED STATES GOVERNMENT LEGEND

If you are a United States government agency, then this documentation and the software described herein are provided to you subject to the following:

All technical data and computer software are commercial in nature and developed solely at private expense. Software is delivered as “Commercial Computer Software” as defined in DFARS 252.227-7014 (June 1995) or as a “commercial item” as defined in FAR 2.101(a) and as such is provided with only such rights as are provided in 3Com’s standard commercial license for the Software. Technical data is provided with limited rights only as provided in DFAR 252.227-7015 (Nov 1995) or FAR 52.227-14 (June 1987), whichever is applicable. You agree not to remove or deface any portion of any legend provided on any licensed program or documentation contained in, or delivered to you in conjunction with, this User Guide.

Unless otherwise indicated, 3Com registered trademarks are registered in the United States and may or may not be registered in other countries.

3Com and the 3Com logo are registered trademarks of 3Com Corporation.

Cisco is a registered trademark of Cisco Systems, Inc.

Funk RADIUS is a registered trademark of Funk Software, Inc.

Aegis is a registered trademark of Aegis Group PLC.

Intel and Pentium are registered trademarks of Intel Corporation. Microsoft, MS-DOS, Windows, and Windows NT are registered trademarks of Microsoft Corporation. Novell and NetWare are registered trademarks of Novell, Inc. UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Ltd.

IEEE and 802 are registered trademarks of the Institute of Electrical and Electronics Engineers, Inc.

All other company and product names may be trademarks of the respective companies with which they are associated.

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.

CONTENTS

6 Ethernet Interface Configuration

7 MAC Address Table Management Configuration

8 Link Aggregation Overview

9 Link Aggregation Configuration

10 Port Mirroring Configuration

11 MSTP Configuration

19 IP Routing Overview

20 ARP Configuration

21 Proxy ARP Configuration

22 IP Addressing Configuration

23 IPv6 Basics Configuration

24 IP Performance Configuration

41 IGMP Snooping Configuration

42 PIM Configuration

43 MSDP Configuration

44 MLD Configuration

45 MLD Snooping Configuration

64 PRIORITY MAPPING

65 CONGESTION AVOIDANCE

66 AGGREGATION CAR CONFIGURATION

67 VLAN POLICY CONFIGURATION

68 TRAFFIC MIRRORING CONFIGURATION

69 OUTBOUND TRAFFIC STATISTICS CONFIGURATION

70 AAA, RADIUS AND HWTACACS CONFIGURATION

77 DEVICE MANAGEMENT

78 POE CONFIGURATION

79 SYSTEM MAINTENANCE AND DEBUGGING

80 FILE SYSTEM MANAGEMENT CONFIGURATION

81 FTP CONFIGURATION

82 TFTP CONFIGURATION

ABOUT THIS GUIDE

This guide describes the 3Com® Switch 8800 and how to install hardware, configure and boot software, and maintain software and hardware. This guide also provides troubleshooting and support information for your switch.

This guide is intended for Qualified Service personnel who are responsible for configuring, using, and managing the switches. It assumes a working knowledge of local area network (LAN) operations and familiarity with communication protocols that are used to interconnect LANs.

Always download the Release Notes for your product from the 3Com World Wide Web site and check for the latest updates to software and product documentation:

`http://www.3com.com`

Conventions

Table 1 lists icon conventions that are used throughout this guide.

Table 1 Notice Icons

Icon	Notice Type	Description
i	Information note	Information that describes important features or instructions.
⚠	Caution	Information that alerts you to potential loss of data or potential damage to an application, system, or device.
⚡	Warning	Information that alerts you to potential personal injury.

Table 2 lists text conventions that are used throughout this guide.

Table 2 Text Conventions

Convention	Description
`Screen displays`	This typeface represents information as it appears on the screen.
Keyboard key names	If you must press two or more keys simultaneously, the key names are linked with a plus sign (+), for example: Press Ctrl+Alt+Del
The words “enter” and “type”	When you see the word “enter” in this guide, you must type something, and then press Return or Enter. Do not press Return or Enter when an instruction simply says “type.”

Table 2 Text Conventions

Convention	Description
Words in *italics*	Italics are used to: Emphasize a point. Denote a new term at the place where it is defined in the text. Identify menu names, menu commands, and software button names. Examples: From the *Help* menu, select *Contents*. Click *OK*.
Words in **bold**	Boldface type is used to highlight command names. For example, "Use the **display user-interface** command to..."

Related Documentation

The following manuals offer additional information necessary for managing your Switch 8800:

- *Switch 8800 Command Reference Guide* — Provides detailed descriptions of command line interface (CLI) commands, that you require to manage your Switch 8800.
- *Switch 8800 Configuration Guide*— Describes how to configure your Switch 8800 using the supported protocols and CLI commands.
- *Switch 8800 Release Notes* — Contains the latest information about your product. If information in this guide differs from information in the release notes, use the information in the *Release Notes*.

These documents are available in Adobe Acrobat Reader Portable Document Format (PDF) on the CD-ROM that accompanies your router or on the 3Com World Wide Web site:

http://www.3com.com/

About this Document

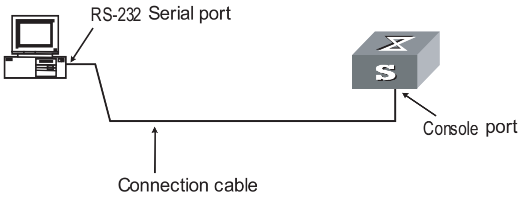
⚠ 3Com supports only the commands that are described in this guide. You may encounter commands in the device's command line interface (CLI) that are not described in this guide. Any command that you see in the CLI but is not described in this guide is not supported in this version of the software. Unsupported commands may result in a loss of data and you enter them at your own risk.

1 LOGGING IN TO A SWITCH

Setting up the Configuration Environment Through the Console Port

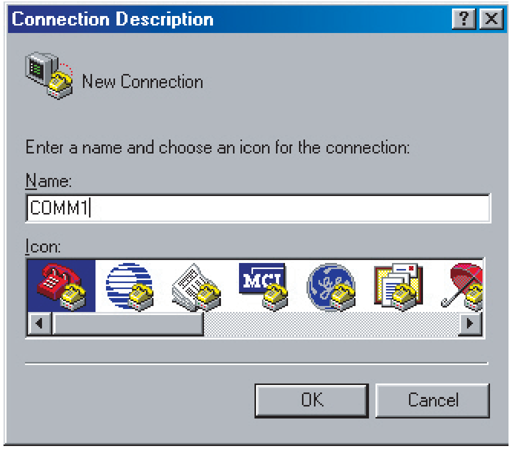
1 Set up the local configuration environment by connecting the serial port of the computer (or a terminal) with the Console port of the switch through a cable, as shown in Figure 1.

Figure 1 Set up the local configuration environment through the Console port



2 Run the terminal emulation program (Terminal in Windows 3.X or HyperTerminal in Windows 9X, etc.), and configure the terminal communication parameters as follows: baud rate as 9,600 bit/s, data bits as 8, stop bit as 1, parity and flow control as none, and terminal type as VT100, as shown in Figure 2, Figure 3, and Figure 4.

Figure 2 Create a connection

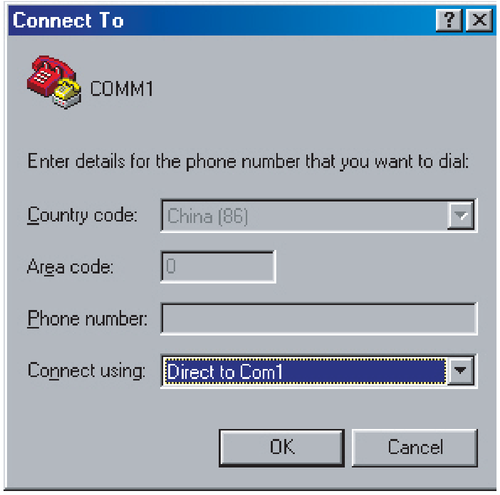


Enter a name and choose an icon for the connection:

Name:

COMM1

Figure 3 Configure the connection port



Enter details for the phone number that you want to dial:

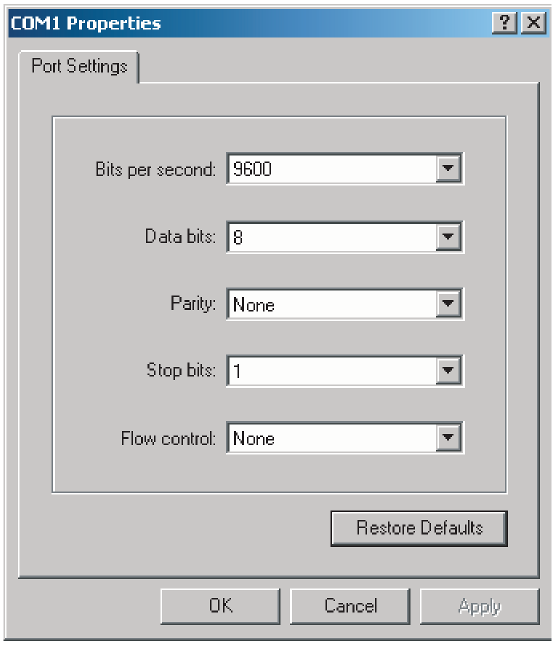
Country code: China (86)

Area code: 0

Phone number:

Connect using: Direct to Com1

Figure 4 Configure communication parameters of the port



Bits per second: 9600

Data bits: 8

Parity: None

Stop bits: 1

Flow control: None

3 Power on the switch to display the POST (power-on self test) information on the terminal. After the POST, the system will prompt you to press the <Enter> key and display the command line prompt (such as <SW8800>).

4 Enter the commands, configure the switch or view the running status of the switch. Enter "?" for help at any time, or refer to other chapters in this manual for specific commands.

Setting up the Configuration Environment Through Telnet

Telnetting a Switch from a PC (Terminal)

If you have properly configured the IP address of a VLAN interface through the Console port (using the **ip address** command in VLAN interface view) and specified the Ethernet port connecting the terminal to the VLAN (using the **port** command in VLAN view), you can log in to the switch through Telnet and configure the switch.

1 Before logging in to the switch through Telnet, set the username and password through the Console port.

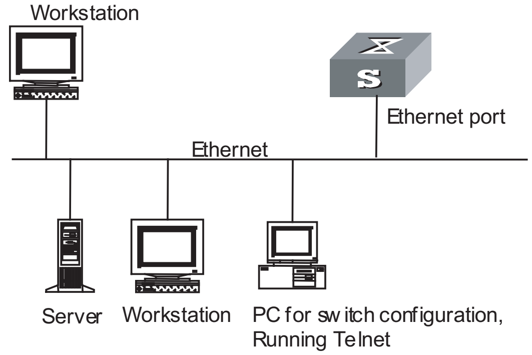
By default, password authentication is required for Telnet. In the case that no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface vty 0
[SW8800-ui-vty0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Telnet user.

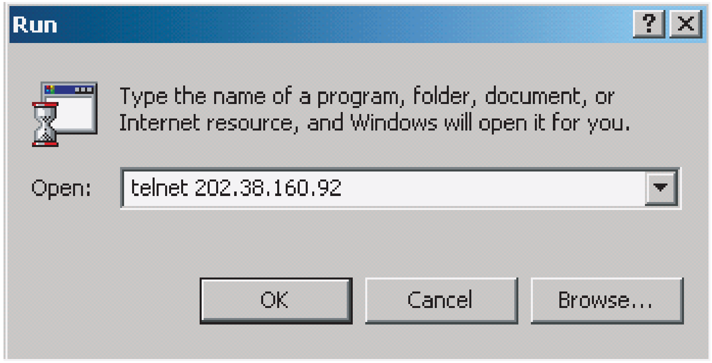
2 Connect the Ethernet port of the PC with that of the switch through a LAN, as shown in Figure 5.

Figure 5 Set up the local configuration environment through a LAN



3 Enter the IP address of the VLAN to which the Ethernet port connecting the PC belongs, and then run the Telnet program, as shown in Figure 6.

Figure 6 Run the Telnet program


Run
Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.
Open:
OK
Cancel
Browse...

```
telnet 202.38.160.92
```

4 The system displays "Login authentication" on the terminal and prompts you to enter a password. The system displays command line prompt (such as <SW8800>) if the password is correct. If it displays "All user interfaces are used, please try later! The connection was closed by the remote host!", you are recommended to try it later (this indicates that the number of login users has reached the maximum value, which is 5 for 3Com series switches).

5 Use the corresponding commands to configure the switch or view its running status. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

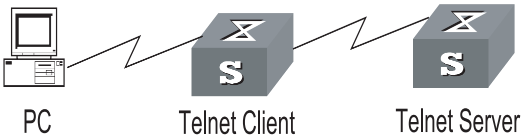
- *When configuring a switch through Telnet, do not delete or modify the IP address of the VLAN interface on the switch. Otherwise, Telnet connection fails.*
- *By default, Telnet users logging into the switch through password authentication can access the commands at level 0.*

Telnetting Another Switch from the Current Switch

After you Telnet a switch, you can Telnet another switch from it for configuration. The local switch functions as a Telnet client and the remote switch functions as the Telnet server. If the two switch ports are in a same LAN, their IP addresses must be configured in a same network segment. Otherwise, a route must exist between the two switches.

The configuration environment is shown in Figure 7.

Figure 7 Provide Telnet Client service



1 Set the Telnet username and password through the Console port on the switch functioning as the Telnet server.

By default, password authentication is required for Telnet. If no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface vty 0
[SW8800-ui-vty0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Telnet user.

2 Telnet the switch functioning as the Telnet client.

3 Perform the following operation on the client:

```
<SW8800> telnet xxxx
```

xxxx indicates the host name or IP address of the server, and the host name must be the one configured by the **ip host** command or resolved by the DNS client.

4 Enter the set login password, and the command line prompt (such as <SW8800>) appears if the password is correct. If it displays "All user interfaces are used, please try later! The connection was closed by the remote host!", you are recommended to try it later (this indicates that the number of login users has reached the maximum value, which is five for 3Com series switches).

5 Use the corresponding commands to configure the switch or view its running status. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

Setting up the Configuration Environment Through a Modem

1 Authenticate Modem users through the Console port on the switch.

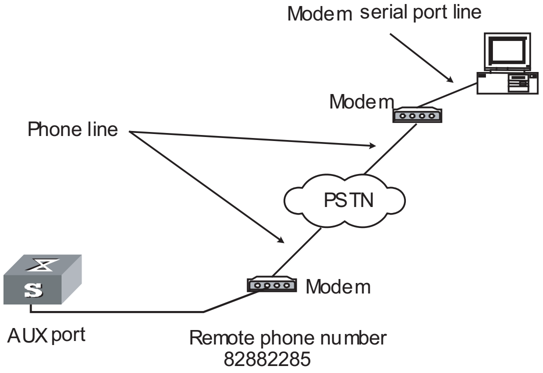
By default, password authentication is required for Telnet. If that no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface aux 0
[SW8800-ui-aux0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Modem user.

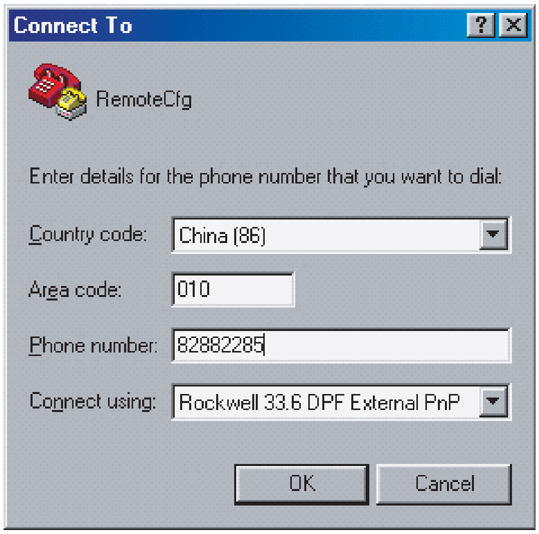
2 Set up the remote configuration environment. Connect the two Modems to the serial interface of the computer (or the terminal) and the AUX port of the switch respectively, as shown in Figure 8.

Figure 8 Set up the remote configuration environment



3 Dial up to connect to the switch through the terminal emulation program and Modem at the remote side (the dialed number should be the telephone number of the Modem connected to the switch), as shown in Figure 9 and Figure 10.

Figure 9 Set the dialed telephone number



Connect To

RemoteCfg

Enter details for the phone number that you want to dial:

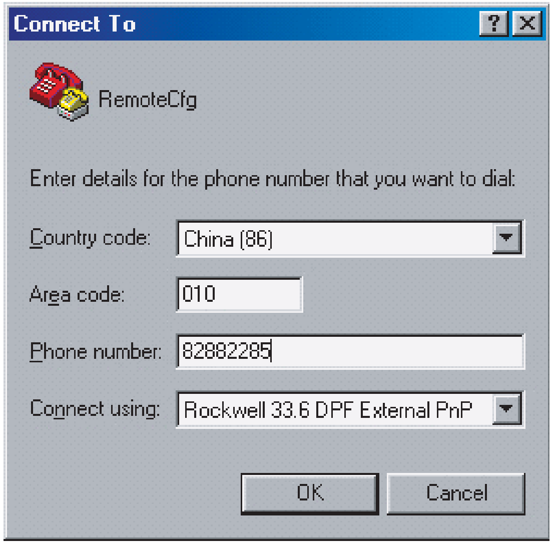
Country code: China (86)

Area code: 010

Phone number: 82882285

Connect using: Rockwell 33.6 DPF External PnP

Figure 10 Dial up on a remote computer



Connect To

RemoteCfg

Enter details for the phone number that you want to dial:

Country code: China (86)

Area code: 010

Phone number: 82882285

Connect using: Rockwell 33.6 DPF External PnP

4 Enter the login password through the remote terminal emulation program, and the system displays command line prompt (such as <SW8800>) to configure or manage the switch. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

By default, Modem users logging in successfully through a Modem can access the commands at level 0.

2 BASIC CONFIGURATIONS

While performing basic configurations of the system, go to these sections for information you are interested in:

- "Basic Configurations" on page 27
- "CLI Features" on page 33

Basic Configurations

This section covers the following topics:

- "Entering/Exiting System View" on page 27
- "Configuring the Device Name" on page 27
- "Configuring the System Clock" on page 27
- "Configuring a Banner" on page 28
- "Configuring CLI Hotkeys" on page 29
- "Configuring User Levels and Command Levels" on page 30
- "Displaying and Maintaining Basic Configurations" on page 32

Entering/Exiting System View

To do...	Use the command...	Remarks
Enter system view from user view	**system-view**	-
Return to user view from system view	**quit**	-

*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

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the device name	**sysname** *sysname*	Optional The default device name is 3Com.

Configuring the System Clock

To do...	Use the command...	Remarks
Set the standard time	**clock datetime** *time date*	Optional
Set the time zone	**clock timezone** *zone-name* { **add** | **minus** } *time*	Optional

To do...	Use the command...	Remarks
Set a daylight summer time scheme	**clock summer-time** *zone-name* { **one-off** | **repeating** } *start-time start-date end-time end-date offset-time*	Optional

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 TTY Modem user.
- **login** banner, welcome information at login authentications, displayed when password and scheme authentications are configured.
- **motd** 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 (TTY Modem login)	**header incoming** *text*	Optional
Configure the authorization information before login	**header legal** *text*	Optional
Configure the banner to be displayed at login authentication	**header login** *text*	Optional
Configure the banner to be displayed when a user enters user view (Non-TTY Modem login)	**header shell** *text*	Optional

Configuring 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 for hotkeys reserved by the system.

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 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.

Table 1 Hotkeys reserved by the system

Hotkey	Function
<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.
<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 the Enter key)
<Esc+P>	Moves the cursor up by one line (available before you press the Enter key)
<Esc+<>	Specifies the cursor as the beginning of the clipboard.
<Esc+>>	Specifies the cursor as the ending of the clipboard.

These hotkeys are defined by the system. 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 Levels and Command Levels

All the commands are defaulted to different views and categorized into four levels: visit, monitor, system, and manage, identified respectively by 0 through 3. If you want to acquire a higher privilege, you must switch to a higher user level, and it requires password to do so for AUX and VTY user interfaces for the security's sake.

The following table describes the default level of the commands.

Table 2 Default command levels

Level	Privilege	Command
0	Visit	ping, tracert, telnet
1	Monitor	refresh, reset, send
2	System	All configuration commands except for those at manage level
3	Manage	FTP, TFTP, Xmodem, and file system operation commands

Follow these steps to configure user level and command level:

To do...	Use the command...	Remarks
Switch the user level	**super** [*level*]	Optional
Enter system view	**system-view**	-
Configure the password for switching the user level	**super password** [**level** *user-level*] { **simple** | **cipher** } *password*	Optional By default, no password is needed for switching the user level.
Configure the command level in system view	**command-privilege level** *level* **view** *view command*	Optional

The commands available depend on your user level when you log onto a device. For example, if your user level is 3 and the command level of VTY 0 interface is 1, you can use commands below level 3 (inclusive).

CAUTION:

- *When you configure the password for switching user level with the **super password** command, the user level is defaulted to 3 if no user level is specified.*
- *You can switch to a lower user level unconditionally. To switch to a higher user level, however, you need to enter the password needed (The password can be set with the **super password** command.). If the entered password is incorrect or no password is configured, the switch fails. Therefore, before switching to a higher user level, you should configure the password needed.*
- *You are recommended to use the default user level; otherwise the change of user level may bring inconvenience to your maintenance and operation.*

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 configuration files saved in the device storage medium.	**display saved-configuration** [**by-linenum**]	
Display the current configurations	**display current-configuration** [**interface** [*interface-type* [*interface-number*]] | **configuration** [*configuration*] | [**by-linenum**] | [| { **begin** | **exclude** | **include** } *regular-expression*]] *	
Display debugging information	**display debugging** [**interface** *interface-type interface-number*] [*module-name*]	
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**	
Display the usage of the current system memory	**display memory**	

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 statistics of each module's running status. Use the **display diagnostic-information** command to collect at one time the information displayed by each of the following commands:

- **display clock**
- **display version**
- **display device**
- **display current-configuration**
- **display saved-configuration**
- **display interface**
- **display controller**
- **display fib**
- **display ip interface**
- **display ip statistics**
- **display memory**

- **display task**
- **display logbuffer**
- **display history all**

- *For the detailed description of the* ***display users*** *command, refer to "Displaying and Maintaining User Interface(s)" on page 50.*
- *The* ***display*** *commands discussed above are for the global configuration. Refer to the corresponding section for the* ***display*** *command for specific protocol and interface.*
- *If no configuration file is enabled when the device is started, no information is displayed by the* ***display saved-configuration*** *command; otherwise, the information of the configuration file is displayed. For the detailed information of the* ***display saved-configuration*** *command, refer to "Displaying and Maintaining Device Configuration" on page 1030.*
- *You are recommended to execute the* ***display diagnostic-information*** *command for at least two consecutive times, so that you can compare the differences between output running information to locate the fault. However, you should use this command only when necessary because execution of the command will continuously print lots of information, affecting the system operation.*

CLI Features

This section covers the following topics:

- "Online Help with Command Lines" on page 33
- "Display Features" on page 34
- "History Command" on page 34
- "Command Line Error Information" on page 35
- "Edit Features" on page 35

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            Display current 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> language-mode ?
  chinese  Chinese environment
  english  English environment
```

3 Enter a command and a <?> separated by a space. If <?> is at the position of a parameter, the description about this parameters 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> pi?
   ping
```

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.

Display Features

CLI offers the following feature:

When the information displayed exceeds one screen, you can pause using one of the methods shown in Table 3.

Table 3 Display functions

Action	Function
Enter <Ctrl+C> when information display pauses	Stops the display and the command execution.
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.
<Ctrl+E>	Moves the cursor to the end of the current line.

History Command

The CLI can automatically save the commands that have been used. You can 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 log buffer for the current user interface (For the detailed description of the **history-command max-size** command, refer to *"User Interface Configuration" on page 43).* The following table lists the operations that you can perform.

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.

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 use <Ctrl+P> and <Ctrl+N> instead.

Command Line Error Information

The commands are executed only if they have no syntax error. Otherwise, error information is reported. Table 4 lists some common errors.

Table 4 Common command line errors

Error information	Cause
Unrecognized command	The command was not found.
	The keyword was not found.
	Parameter type error
	The parameter value is beyond the allowed range.
Incomplete command	Incomplete command
Ambiguous command	Ambiguous command
Too many parameters	Too many parameters
Wrong parameter	Wrong parameter

Edit Features

The CLI provides the basic command edit functions and supports multi-line editing. The maximum length of each command is 256 characters. Table 5 lists these functions.

Table 5 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> key	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.

Table 5 Edit functions

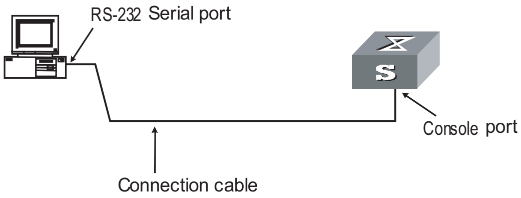
Key	Function
Up-arrow key or <Ctrl+P> Down-arrow key or <Ctrl+N>	Displays history commands
<Tab> key	Pressing <Tab> after entering part of a keyword enables the fuzzy help function. If finding a unique match, the system substitutes the complete keyword for the incomplete one and displays it in the next line. If there are several matches or no match at all, the system does not modify the incomplete keyword and displays it again in the next line.

3 LOGGING IN TO A SWITCH

Setting up the Configuration Environment Through the Console Port

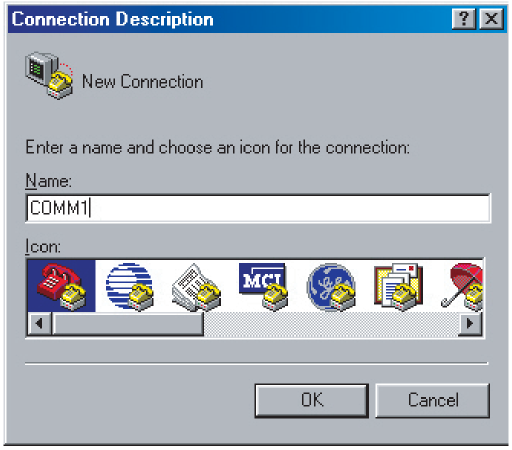
1 Set up the local configuration environment by connecting the serial port of the computer (or a terminal) with the Console port of the switch through a cable, as shown in Figure 11.

Figure 11 Set up the local configuration environment through the Console port



2 Run the terminal emulation program (Terminal in Windows 3.X or HyperTerminal in Windows 9X, etc.), and configure the terminal communication parameters as follows: baud rate as 9,600 bit/s, data bits as 8, stop bit as 1, parity and flow control as none, and terminal type as VT100, as shown in Figure 12, Figure 13, and Figure 14.

Figure 12 Create a connection

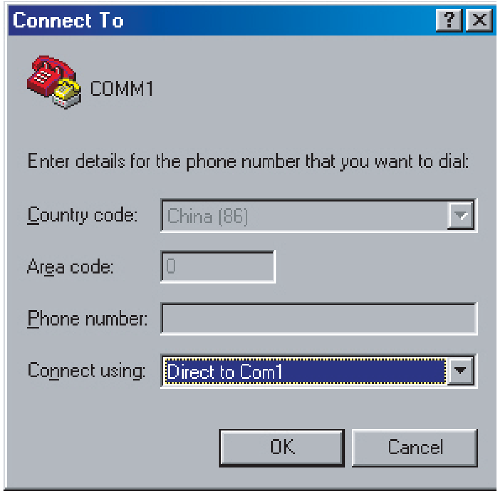

Connection Description

New Connection

Enter a name and choose an icon for the connection:

Name: COMM1

Figure 13 Configure the connection port



Connect To

COMM1

Enter details for the phone number that you want to dial:

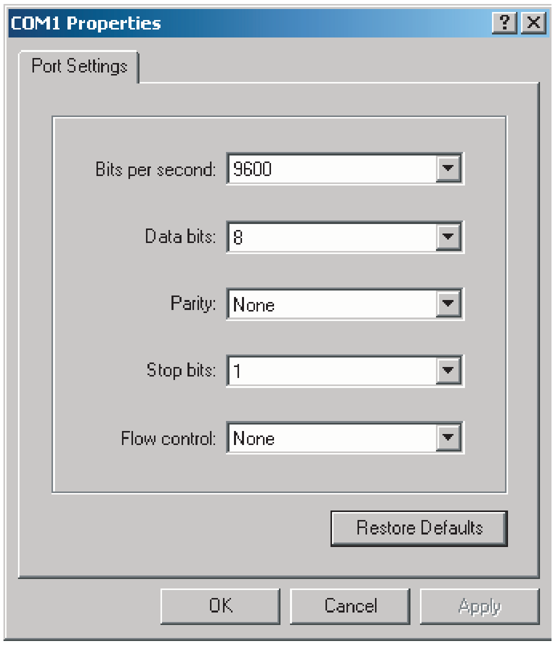
Country code: China (86)

Area code: 0

Phone number:

Connect using: Direct to Com1

Figure 14 Configure communication parameters of the port



COM1 Properties

Port Settings

Bits per second: 9600

Data bits: 8

Parity: None

Stop bits: 1

Flow control: None

3 Power on the switch to display the POST (power-on self test) information on the terminal. After the POST, the system will prompt you to press the <Enter> key and display the command line prompt (such as <SW8800>).

4 Enter the commands, configure the switch or view the running status of the switch. Enter "?" for help at any time, or refer to other chapters in this manual for specific commands.

Setting up the Configuration Environment Through Telnet

Telnetting a Switch from a PC (Terminal)

If you have properly configured the IP address of a VLAN interface through the Console port (using the **ip address** command in VLAN interface view) and specified the Ethernet port connecting the terminal to the VLAN (using the **port** command in VLAN view), you can log in to the switch through Telnet and configure the switch.

1 Before logging in to the switch through Telnet, set the username and password through the Console port.

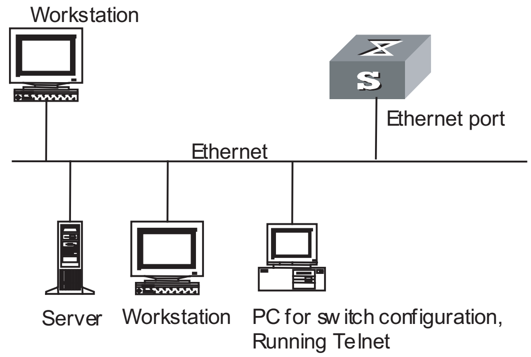
By default, password authentication is required for Telnet. In the case that no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface vty 0
[SW8800-ui-vty0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Telnet user.

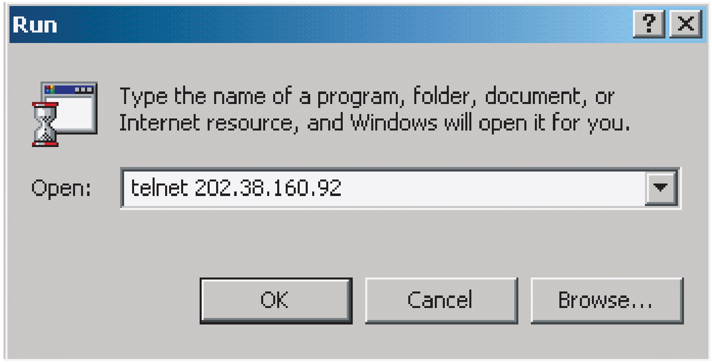
2 Connect the Ethernet port of the PC with that of the switch through a LAN, as shown in Figure 15.

Figure 15 Set up the local configuration environment through a LAN



3 Enter the IP address of the VLAN to which the Ethernet port connecting the PC belongs, and then run the Telnet program, as shown in Figure 16.

Figure 16 Run the Telnet program



Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

Open: telnet 202.38.160.92

4 The system displays "Login authentication" on the terminal and prompts you to enter a password. The system displays command line prompt (such as <SW8800>) if the password is correct. If it displays "All user interfaces are used, please try later! The connection was closed by the remote host!", you are recommended to try it later (this indicates that the number of login users has reached the maximum value, which is 5 for 3Com series switches).

5 Use the corresponding commands to configure the switch or view its running status. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

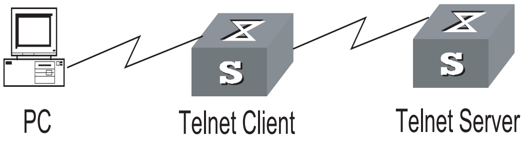
- *When configuring a switch through Telnet, do not delete or modify the IP address of the VLAN interface on the switch. Otherwise, Telnet connection fails.*
- *By default, Telnet users logging into the switch through password authentication can access the commands at level 0.*

Telnetting Another Switch from the Current Switch

After you Telnet a switch, you can Telnet another switch from it for configuration. The local switch functions as a Telnet client and the remote switch functions as the Telnet server. If the two switch ports are in a same LAN, their IP addresses must be configured in a same network segment. Otherwise, a route must exist between the two switches.

The configuration environment is shown in Figure 17.

Figure 17 Provide Telnet Client service



1 Set the Telnet username and password through the Console port on the switch functioning as the Telnet server.

By default, password authentication is required for Telnet. If no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface vty 0
[SW8800-ui-vty0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Telnet user.

2 Telnet the switch functioning as the Telnet client.

3 Perform the following operation on the client:

```
<SW8800> telnet xxxx
```

xxxx indicates the host name or IP address of the server, and the host name must be the one configured by the **ip host** command or resolved by the DNS client.

4 Enter the set login password, and the command line prompt (such as <SW8800>) appears if the password is correct. If it displays "All user interfaces are used, please try later! The connection was closed by the remote host!", you are recommended to try it later (this indicates that the number of login users has reached the maximum value, which is five for 3Com series switches).

5 Use the corresponding commands to configure the switch or view its running status. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

Setting up the Configuration Environment Through a Modem

1 Authenticate Modem users through the Console port on the switch.

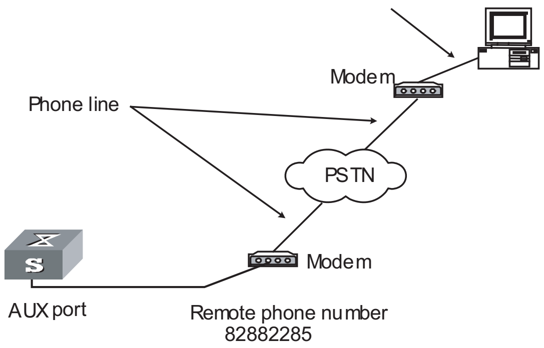
By default, password authentication is required for Telnet. If that no password is set, the system prompts "Login password has not been set!".

```
<SW8800> system-view
System View: return to User View with Ctrl+Z.
[SW8800] user-interface aux 0
[SW8800-ui-aux0] set authentication password simple xxxx
```

xxxx indicates the password to be set for the Modem user.

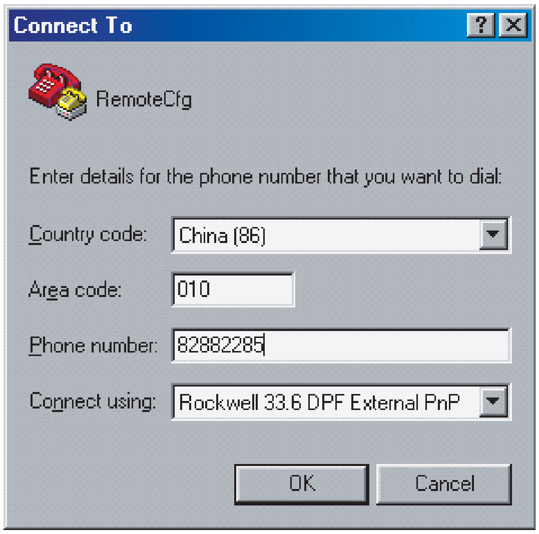
2 Set up the remote configuration environment. Connect the two Modems to the serial interface of the computer (or the terminal) and the AUX port of the switch respectively, as shown in Figure 18.

Figure 18 Set up the remote configuration environment



3 Dial up to connect to the switch through the terminal emulation program and Modem at the remote side (the dialed number should be the telephone number of the Modem connected to the switch), as shown in Figure 19 and Figure 20.

Figure 19 Set the dialed telephone number



Enter details for the phone number that you want to dial:

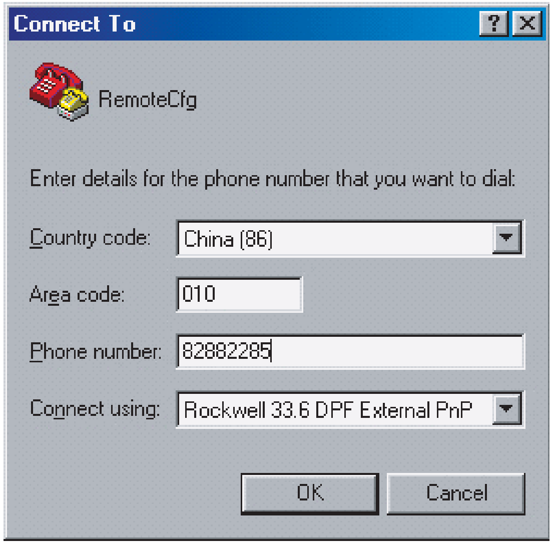
Country code: China (86)

Area code: 010

Phone number: 82882285

Connect using: Rockwell 33.6 DPF External PnP

Figure 20 Dial up on a remote computer



Enter details for the phone number that you want to dial:

Country code: China (86)

Area code: 010

Phone number: 82882285

Connect using: Rockwell 33.6 DPF External PnP

4 Enter the login password through the remote terminal emulation program, and the system displays command line prompt (such as <SW8800>) to configure or manage the switch. Enter "?" for help at any time, and refer to the corresponding chapters in this manual for specific commands.

By default, Modem users logging in successfully through a Modem can access the commands at level 0.

4 USER INTERFACE CONFIGURATION

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

User Interface Overview

Brief Introduction

User interface view is a feature that allows you to manage asynchronous serial interfaces that work in flow mode. By operating under user interface view, you can centralize the management of various configurations.

At present, the system supports the following three configuration modes:

- Local configuration via the Console port
- Local/Remote configuration via the AUX port (Auxiliary port)
- Local/Remote configuration through Telnet or SSH

The three modes correspond to four types of user interfaces. They are:

- Console port: A view which you log in from the console port. Console port is a line device port. The device has only one console port, with the port type as EIA/TIA-232 DCE.

- AUX port: A view which you log in from the AUX port. AUX port is also a line device port. The device has only one AUX port of EIA/TIA-232 DTE type. This port is usually used for dialup access via modem.
- VTY (Virtual Type Terminal): A view which you log in through VTY. VTY port is a logical terminal line used when you access the device by means of Telnet or SSH. Currently, each device supports up to five VTY users to access simultaneously.

User Interface Numbering

User interfaces can be numbered in two ways: absolute numbering and relative numbering.

Absolute numbering

Absolute numbering allows you to uniquely specify a user interface or a group of user interfaces. The numbering system starts from number 0 (representing the Console port), and followed by 1 (representing the AUX port), then 2 to represent VTY 0, and so on.

The numbering approach numbers the three types of user interfaces in the sequence of: Console port, AUX port and VTY. The Console interface and the AUX interface each occupy a number, and the four VTY user interfaces are numbered from 0 to 4.

You can use the **display user-interface** command to view the number of the user interfaces.

Relative numbering

Relative numbering numbers a user interface in the form of "user interface type + number". In this way, it can specify a user interface or a group of user interfaces of a specific type. This form of number is valid only when used under that type of user interface. It makes no sense when used under other types of user interfaces. The rules of relative numbering are as follows:

- CON is numbered CON 0.
- AUX is numbered AUX 0.
- VTYs are numbered from 0 in ascending order, with a step of 1.

Configuring User Interface

Complete these tasks to configure user interface:

Task	Remarks
“Configuring Asynchronous Serial Interface Attributes” on page 45	Optional
“Configuring Terminal Attributes” on page 45	Optional
“Configuring Modem Attributes” on page 46	Optional
“Configuring the auto-execute Command” on page 47	Optional
“Configuring User Privilege Level” on page 47	Optional

Task	**Remarks**
"Configuring Access Restriction on VTY User Interface(s)" on page 48	Optional
"Configuring Supported Protocols on VTY User Interface(s)" on page 48	Optional
"Configuring Authentication Mode at Login" on page 49	Optional
"Sending Messages to the Specified User Interface(s)" on page 50	Optional
"Releasing the Connection Established on the User Interface(s)" on page 50	Optional
"Displaying and Maintaining User Interface(s)" on page 50	Optional

Configuring Asynchronous Serial Interface Attributes

Follow these steps to configure asynchronous attributes of a serial interface:

To do...	**Use the command...**	**Remarks**
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--
Configure transmission speed	**speed** *speed-value*	Optional 9600 bps by default
Configure flow control mode	**flow-control** { **none** | **software** | **hardware** }	Optional **none** by default
Set parity bits	**parity** { **none** | **even** | **odd** | **mark** | **space** }	Optional **none** by default
Set stop bits	**stopbits** { **1.5** | **1** | **2** }	Optional 1 by default Currently, stop bits 1.5 cannot be configured.
Set data bits	**databits** { **5** | **6** | **7** | **8** }	Optional 8 by default Currently, data bits 5 and 6 cannot be configured.

The above configuration takes effect only when the asynchronous serial interface is working in asynchronous flow mode.

Configuring Terminal Attributes

To do...	**Use the command...**	**Remarks**
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--

To do...	Use the command...	Remarks
Start the terminal service	**shell**	Optional The terminal service is enabled on all user interfaces by default.
Set the idle-timeout disconnection function for terminal users	**idle-timeout** *minutes* [*seconds*]	Optional 10 minutes by default.
Set the screen-length of the terminal screen	**screen-length** *screen-length*	Optional The screen displays 24 lines of data by default.
Set the display type of a terminal	**terminal type** { **ansi** | **vt100** }	Optional ANSI by default.
Set the number of the history commands that can be stored in the history buffer	**history-command max-size** *size-value*	Optional The history buffer can store 10 commands by default.
Return to user view	**return**	--
Lock user interface, preventing unauthorized users from using this interface	**lock**	Optional Disabled by default.

The system supports two types of terminal display: ANSI and VT100. If the terminal display of the device and the client (for example, hyper terminal or Telnet terminal) is inconsistent or is set to ANSI, and if the total number of the characters of the currently using command line exceeds 80, anomalies such as cursor corruption or abnormal display of the terminal display may occur on the client. Therefore, you are recommended to set the display type of both the device and the client to VT100.

Configuring Modem Attributes

In the event of dial-in through a modem into an asynchronous interface, you can manage and configure the modem-concerned parameters in user interface view.

Follow these steps to configure the modem attributes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **vty** } *first-num2* [*last-num2*] }	--
Set the interval for a user from hookoff to dial-up when dial-in connection is established	**modem timer answer** *time*	Optional 30 seconds by default
Enable auto answer for the modem	**modem auto-answer**	Optional Manual answer by default
Enable the modem to dial in, dial out or both	**modem** { **both** | **call-in** | **call-out** }	Optional Disabled by default

The above configuration takes effect only for the AUX and VTY ports working in flow mode.

Configuring the auto-execute Command

With the **auto-execute command** command enabled, the system automatically executes the configured command when you log in. After the command is completed or after the tasks triggered by the command are completed, the connection breaks automatically.

This command is normally used to configure the Telnet command to enable you to connect to the specified host automatically.

Follow these steps to configure auto-execute command:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **vty** } *first-num2* [*last-num2*] }	--
Configure the command to be automatically executed	**auto-execute command** *command*	Required No command is set to be automatically executed by default.

The **auto-execute command** command is supported on all types of user interfaces except the Console port and the AUX port functioning as the console port.

CAUTION: *The* ***auto-execute command*** *command may disable you from configuring the system through the terminal line to which the command is applied. Therefore, before configuring the command and saving the configuration (using the save command), make sure that you can access the system by other means to remove the configuration in case a problem occurs.*

Configuring User Privilege Level

You can restrict a user to use only a subset of all the system commands through settings on two aspects: user interface level and user level.

- If username and password are needed in the configured authentication mode, the user privilege level is defined by the user level. For SSH users, when they use RSA public key authentication, their privilege level is defined by the level configured on the user interface.
- If no authentication is adopted or the password authentication is adopted, the user privilege level is defined by the user interface level used when login.
- If the setting of user interface level is inconsistent with that of the user level, the user level applies. For example, if user1 can use level 3 commands, and the user interface VTY0 can use level 2 commands, then user1 can use commands of level 3 or a lower level when logging onto the system through VTY0.

Setting of the user level: Use the **local-user** command in system view to create a user and enter local user view, in which use the **level** command to specify the user level. For the detailed description of the **local-user** and **level** commands, refer to "Configuring Local User Attributes" on page 889.

Follow these steps to configure the user privilege level under a user interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--
Configure user's privilege level under the current user interface	**user privilege level** *level*	Optional By default, users logging in from Console port have a privilege level of 3; users logging in from other user interfaces have a privilege level of 0.

Configuring Access Restriction on VTY User Interface(s)

You can configure access restriction on the VTY user interface through referencing an ACL. For details regarding ACL, refer to *"ACL Overview" on page 801.*

Follow these steps to configure access restriction on VTY user interfaces:

To do...		Use the command...	Remarks
Enter system view		**system-view**	--
Enter VTY user interface view		**user-interface** { *first-num1* [*last-num1*] | **vty** *first-num2* [*last-num2*] }	--
Configure the access restriction on the VTY user interface	By referencing basic/advanced ACL	**acl** [**ipv6**] *acl-number* { **inbound** | **outbound** }	Use either command No restriction is set by default.
	By referencing Layer 2 ACL	**acl** *acl-number* **inbound**	

Configuring Supported Protocols on VTY User Interface(s)

Currently, only the VTY user interface allows configuration on the supported protocols.

Follow these steps to configure supported protocols on the active VTY user interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter VTY user interface view	**user-interface** { *first-num1* [*last-num1*] | **vty** *first-num2* [*last-num2*] }	--
Configure the supported protocol(s) on the active user interface	**protocol inbound** { **all** | **ssh** | **telnet** }	Optional Both Telnet and SSH are supported by default.

CAUTION:

- *If SSH is configured, you must set the authentication mode to **scheme** using the **authentication-mode scheme** command to guarantee a successful login.*

*The **protocol inbound ssh** command fails if the authentication mode is **password** or **none**. For the corresponding configuration, refer to the **authentication-mode** command in the Switch 8800 Command Reference Guide.*

- *The protocol(s) configured through the protocol inbound command takes effect next time you log in from that user interface.*

Configuring Authentication Mode at Login

With the configuration of user interface authentication mode, you can decide whether to authenticate users when they log on through the specified user interface, thus enhancing the security of the device. The supported authentication modes on the device are **none**, **password**, and **scheme**.

- If you specify the authentication mode as **none**, then no username and password are needed when users log on through the specified user interface, which may be insecure.
- If you specify the authentication mode as **password**, then password authentication is needed when users log on through the specified user interface. Input of empty or wrong password may result in login failure. Before terminating the redirected Telnet connection, set the password of the specified user interface.
- If you specify the authentication mode as **scheme**, then username and password authentication is needed when users log on through the specified user interface. Input of empty or wrong password may result in login failure. Before terminating the redirected Telnet connection, set the username and password.

Follow these steps to configure authentication mode at login as **none**:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--
Set authentication mode at login to **none**	**authentication-mode none**	Required By default, the authentication mode is **password** for VTY and AUX user interfaces and is **none** for Console interface.

Follow these steps to configure authentication mode at login as **password**:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--
Set authentication mode at login to **password**	**authentication-mode password**	Required By default, the authentication mode is **password** for VTY and AUX user interfaces and is **none** for Console interface.

To do...	Use the command...	Remarks
Set local authentication password	**set authentication password** { **cipher** | **simple** } *password*	Required No local authentication password is set by default.

Follow these steps to configure authentication mode at login as **scheme**:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter user interface view	**user-interface** { *first-num1* [*last-num1*] | { **aux** | **console** | **vty** } *first-num2* [*last-num2*] }	--
Set authentication mode at login to **scheme**	**authentication-mode scheme** [**command-authorization**]	Required By default, the authentication mode is password for VTY and AUX user interfaces and is none for Console interface.
Set authentication username and enter local user view	**local-user** *user-name*	Required No local user is set on the device by default.
Set authentication password	**password** { **cipher** | **simple** } *password*	Required

*For the detailed description of the **local-user** and **password** commands, refer to "AAA, RADIUS and HWTACACS Configuration" on page 873.*

Sending Messages to the Specified User Interface(s)

To do...	Use the command...	Remarks
Send messages to the specified user interface(s)	**send** { **all** | *num1* | { **aux** | **console** | **vty** } *num2* }	Required

You cannot use this command to release the connection that a user is using.

Releasing the Connection Established on the User Interface(s)

To do...	Use the command...	Remarks
Release the connection established on the specified user interface(s)	**free user-interface** { *num1* | { **aux** | **console** | **vty** } *num2* }	Required

Displaying and Maintaining User Interface(s)

To do...	Use the command...	Remarks
Display the information on the use of the user interface(s)	**display users** [**all**]	Available in any view
Display the information about the specified or all user interface(s)	**display user-interface** [*num1* | { **aux** | **console** | **vty** } *num2*] [**summary**]	Available in any view

To do...	Use the command...	Remarks
Display the history commands that the current user has configured	**display history-command**	Available in any view

5 MANAGEMENT ETHERNET PORT CONFIGURATION

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

- "Management Ethernet Port Overview" on page 53
- "Management Ethernet Port Configuration" on page 53

Management Ethernet Port Overview

Each Fabric on a Switch 8800 series switch provides a 10/100Base-TX management Ethernet port (M-Ethernet) which has the functions listed below:

- Connected with a PC, the port implements downloading and debugging of system programs.
- Connected with a remote network management workstation, the port implements remote system management.

Management Ethernet Port Configuration

You can perform the following operations in management Ethernet port view:

- Configure the IP address for an interface
- Bring up/shut down an interface
- Set the description for an interface
- Display the current system information
- Test network connectivity (**ping**, **tracert**)

For detailed information, refer to "Ethernet Interface Configuration" on page 55 and "System Maintenance and Debugging" on page 1017.

CAUTION:

- *A management Ethernet port is available only when being configured with an IP address.*
- *Management Ethernet ports do not support dynamic routing protocols.*

6 ETHERNET INTERFACE CONFIGURATION

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

- "Ethernet Interface Configuration" on page 55
- "Maintaining and Displaying an Ethernet Interface" on page 60

Ethernet Interface Configuration

Configuration Task List

Complete the following tasks to configure an Ethernet interface:

Task	Remarks
"Basic Ethernet Interface Configuration" on page 55	Optional
"Configuring Flow Control on an Ethernet Interface" on page 56	Optional
"Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Interface" on page 57	Optional
"Configuring Loopback Testing on an Ethernet Interface" on page 57	Optional
"Configuring a Port Group" on page 58	Optional
"Setting the Storm Suppression Ratio for an Ethernet Interface" on page 58	Optional
"Setting the Interval for Collecting Ethernet Interface Statistics" on page 59	Optional
"Enabling the Forwarding of Jumbo Frames" on page 59	Optional
"Configuring the Cable Type for an Ethernet Interface" on page 60	Optional
"Configuring the Source MAC Address for an Interface" on page 60	Optional

Basic Ethernet Interface Configuration

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.

Follow these steps to perform basic Ethernet interface configurations:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Set the description string for the Ethernet interface	**description** *text*	Optional By default, the description string is "interface index + Interface".
Set the duplex mode	**duplex** { **auto** | **full** | **half** }	Optional **auto** by default.
Set the transmission rate	**speed** { **10** | **100** | **1000** | **auto** }	Optional **auto** by default.
Shut down the Ethernet interface	**shutdown**	Optional By default, an Ethernet interface is in up state.

- *The **speed 1000** command is only applicable to GigabitEthernet interfaces.*
- *GigabitEthernet electric interfaces cannot operate in half-duplex mode when the transmission rate is set to 1,000 Mbps.*
- *Ethernet optical interfaces cannot operate in half-duplex mode.*
- *When optoelectric transducers are used, make sure the interfaces operate in auto-negotiation mode. Otherwise, the interfaces may operate improperly.*

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 packets when it receives the Pause frame. In this way, flow controls helps to avoid the dropping of packets. Note that only after both the ingress and the egress interfaces have turned on their flow control will this be possible.

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 Turned off 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 up/down suppression time 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 Defaults to 1 second.

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: a loopback testing carried out within the device, if data packets sent from an Ethernet interface can be received by the same interface, the internal loopback testing is successful indicating that the interface is functioning properly.
- External loopback testing: a loopback plug needs to be plugged into an Ethernet interface, if data packets sent from the interface is received by the same interface through the loopback plug, the external loopback testing is successful indicating that the interface is functioning properly.

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*	-
Enable loopback testing	**loopback** { **external** | **internal** }	Optional Disabled by default.

- *Currently, Switch 8800s do not support the external loopback testing.*
- *Internal loopback testing can only be enabled on interfaces in down state. That is, it is unavailable to interfaces that are shut down manually.*
- *The **speed, duplex, mdi**, and **shutdown** commands are not applicable during a loopback testing.*
- *With the loopback testing enabled, the Ethernet interface works in the full duplex mode at highest speed. With the loopback testing enabled, the original configurations will be restored.*

Configuring a Port Group

To make the configuration task easier for users, certain devices allow users to configure on a single port as well as on multiple ports in a port group. In port group view, the user only needs to input the configuration command once on one port and that configuration will apply to all ports in the port group. This effectively reduces redundant configurations.

A Port group belongs to one of the following two categories:

- Manual port group: manually created by users. Multiple Ethernet interfaces can be added to the same port group;
- Dynamic port group: dynamically created by system, currently mainly applied in link aggregation port groups. A link aggregation port group is automatically created together with the creation of a link aggregation group and cannot be created by users through command line input. Adding or deleting of ports in a link aggregation port group can only be achieved through operations on the link aggregation group.

Follow these steps to enter port group view:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter port group view	Enter manual port group view	**port-group manual** *port-group-name*	-
	Enter aggregation port group view	**port-group aggregation** *agg-id*	-

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

Setting the Storm Suppression Ratio for an Ethernet Interface

You can suppress the broadcast traffic passing through an Ethernet interface in the following way.

In interface configuration mode, the suppression ratio indicates the maximum broadcast traffic allowed to pass through an interface. When the broadcast traffic over the interface exceeds the threshold, the system will discard the extra packets so that the broadcast traffic ratio can drop below the limit to ensure that the network functions properly.

Follow these steps to configure the storm suppression ratio for an Ethernet interface:

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 interface only; if configured in port group view, this feature takes effect on all the interfaces in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure 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.

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

Complete the following configuration tasks to configure the time interval for collecting interface statistics. Use the **display interface** command to display the interface statistics within the current interval.

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

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Configure the time interval for collecting interface statistics	**flow-interval** *interval*	Required Defaults to 300 seconds.

Enabling the Forwarding of Jumbo Frames

Due to tremendous amount of traffic occurring in Ethernet, it is likely that some frames might have a frame size greater than the standard Ethernet frame size. By allowing such frames (called jumbo frames) to pass through Ethernet interfaces, you can forward frames with a size greater than the standard Ethernet frame size and yet still within the specified parameter range.

Follow the following 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	**jumboframe enable** [*jumboframe-value*] **slot** *slot-number*	-

Configuring the Cable Type for an Ethernet Interface

Follow these steps to configure the cable type 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 cable type for an Ethernet interface	**mdi** { **across** | **auto** | **normal** }	Optional Defaults to auto, that is, the system automatically detects the type of cable in use.

Configuring the Source MAC Address for an Interface

Normally, when a packet is forwarded by a Switch 8800 on Layer 3, its source MAC address is that of the VLAN interface which the outbound interface corresponds to. However, It is required in some cases that the source MAC addresses of packets forwarded through different interfaces be different for forwarding policies to take effect on the peer devices.

By configuring the source MAC address for an interface, you can set the least octet of the source MAC addresses of the packets forwarded through the interface for forwarding policies to take effect on the peer devices.

Follow these steps to configure the source MAC address for an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Configure the source MAC address	**source-mac-tail** *last-byte*	Required By default, the source MAC address of an interface is that of the corresponding VLAN interface.

CAUTION: *After you configure the source MAC address for an interface, packets forwarded on Layer 3 through the interface use the MAC address of the corresponding VLAN interface as their source MAC addresses, with the least octet being replaced with value specified by the* ***source-mac-tail*** *command.*

Currently, this command is not supported on the 3C17532, 3C17538, and 3C17526 modules.

Maintaining and Displaying an Ethernet Interface

To do...	Use the command...	Remarks
Display the information about a interface	**display interface** [*interface-type* [*interface-number*]]	Available in any view
Display the information about an interface in brief	**display brief interface** [*interface-type* [*interface-number*]] [| { **begin** | **include** | **exclude**} *text*]	Available in any view

To do...	Use the command...	Remarks
Clear the statistics on a interface	**reset counters interface** [*interface-type* [*interface-number*]]	Available in user view
Display the ports that are of a specified type	**display port** { **hybrid** | **trunk** }	Available in any view
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 statistics on the packets passing through interfaces of specific type	**display counters** { **inbound** | **outbound** } **interface** [*interface-type*]	Available in any view
Display the rate statistics on the packets passing through the interfaces within the last sample time interval	**display counters rate** { **inbound** | **outbound** } **interface** [*interface-type*]	Available in any view

*The **display counters** and **display counters rate** commands only display the statistics on the packets passing through RPR logical ports.*

7 MAC ADDRESS TABLE MANAGEMENT CONFIGURATION

When configuring MAC table management, go to these sections for information you are interested in:

- "Introduction to MAC Address Table" on page 63
- "Configuring MAC Address Table Management" on page 64
- "Displaying MAC Address Table Management" on page 67
- "MAC Address Table Management Configuration Example" on page 67

The term router and router icons mentioned in the following routing protocol refer to the routers in a generic sense and the switches running routing protocols.

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, to which interface this device is connected and to which VLAN the interface belongs.

A MAC address table consists of two types of entries: static and dynamic. Static entries are manually configured and never age out. Dynamic entries can be manually configured or dynamically learned and may age out.

The following is how your device learns a MAC address after it receives a frame from a port, Port 1 for example:

1 Check the frame for the source MAC address (MAC A for example).

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 and indicate from which interface the frame is received.

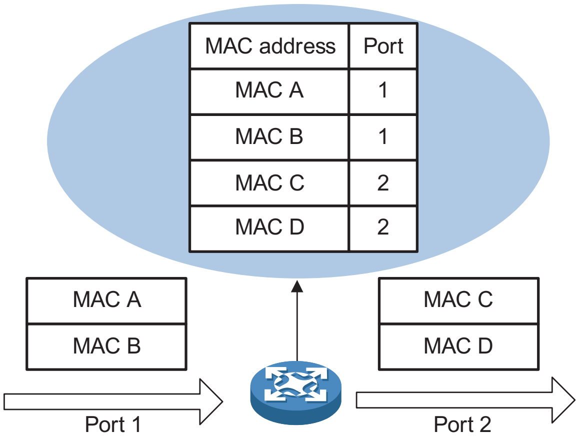
When receiving a frame destined for MAC A, the device then looks up the MAC address table and forwards it from port 1.

Dynamically learned MAC addresses cannot overwrite static MAC address entries, but the latter can overwrite the former.

As shown in Figure 21, when forwarding a frame, the device looks up the MAC address table. If an entry is available for the destination MAC address, the device forwards the frame directly from the hardware. If not, it does the following:

1. Broadcast the frame.
2. After the frame reaches the destination, the destination sends back a response with its MAC address. (If no response is received, the frame will be dropped.)
3. Upon receipt of the response, the device adds an entry in the MAC address table, indicating from which interface the frames destined for the MAC address should be sent.
4. Forward subsequent frames destined for the same MAC address directly from the hardware.
5. Discard the frames which cannot reach the destination MAC address.

Figure 21 Forward frames using the MAC address table



MAC address	Port
MAC A	1
MAC B	1
MAC C	2
MAC D	2

Configuring MAC Address Table Management

Configuring MAC Address Entries

Follow these steps to add, modify, or remove entries in the MAC address table:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Add/modify a MAC address entry	**mac-address** { **dynamic** | **static** } *mac-address* **interface** *interface-type interface-number* **vlan** *vlan-id* **mac-address blackhole** *mac-address* **vlan** *vlan-id*	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Add/modify MAC address entries under the specified interface view	**mac-address** { **static** | **dynamic** } *mac-address* **vlan** *vlan-id*	Required

Disabling Global MAC Address Learning

You may need to disable MAC address learning sometimes to prevent the MAC address table from being saturated, for example, when your device is being attacked by a great deal of packets with different source MAC addresses.

Disabling the global MAC address learning disables the learning function on all ports.

Follow these steps to disable MAC address learning:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Disable global MAC address learning	**mac-address mac-learning disable**	Required Enabled by default

Disabling MAC Address Learning on an Ethernet Port or Port Group

After enabling global MAC address learning, you may disable the MAC address learning function on a port as needed.

Follow these steps to disable MAC address learning on a port or port group:

<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><b>system-view</b></td><td>-</td></tr>
<tr><td colspan="2">Enable global MAC address learning</td><td><b>undo mac-address mac-learning disable</b></td><td>Optional
Enabled by default.</td></tr>
<tr><td rowspan="2">Enter Ethernet interface view or port group view</td><td>Enter Ethernet interface view</td><td><b>interface</b> <i>interface-type interface-number</i></td><td rowspan="2">Use either command
In Ethernet interface view, the following configurations only take effect for the current port; in port group view, the configurations take effect for all ports.</td></tr>
<tr><td>Enter port group view</td><td><b>port-group</b> { <b>aggregation</b> <i>agg-id</i> | <b>manual</b> <i>port-group-name</i>}</td></tr>
<tr><td colspan="2">Disable MAC address learning on an Ethernet interface or port group</td><td><b>mac-address mac-learning disable</b></td><td>Required
Enabled by default</td></tr>
</table>

Configuring MAC Address Aging Timer

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 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 MAC address aging timer:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Set the aging timer for dynamic MAC address entries	**mac-address timer** { **aging** *seconds* | **no-aging** }	Optional 300 seconds by default.

The aging time of the MAC address is available on all ports. The MAC address aging timer takes effect only on dynamic MAC address entries (learned or administratively configured) only.

Configuring Maximum Number of MAC Addresses an Ethernet Port or a Port Group Can Learn

To prevent a MAC address table from so large that it may degrade forwarding performance, you may restrict the number of MAC addresses that can be learned. One approach is to do this on a per-port or port group basis.

Follow these steps to configure the maximum number of MAC addresses that an Ethernet port or port group can learn:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter Ethernet interface or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command In the Ethernet interface view, the following configurations only take effect on the current port; in the port group view, the configurations take effect on all ports.
	Enter port group view	**port-group** { | **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the maximum number of MAC addresses that can be learned on an Ethernet port or port group or configure whether frames with unknown destination MAC addresses can be forwarded or not after the upper limit is reached		**mac-address max-mac-count** { *count* | **disable-forwarding** }	Required The default maximum number of MAC addresses that can be learned is 14336. After the upper limit is reached, frames with unknown destination MAC addresses are forwarded by default.

Configuring Maximum Number of MAC Addresses a VLAN Can Learn

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. One approach is to do this on a per-VLAN basis.

Follow these steps to configure the maximum number of MAC addresses that a VLAN can learn:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure the maximum number of MAC addresses that can be learned on a VLAN	**mac-address max-mac-count** *count*	Required 172032 by default

Since there are no layer 2 physical interfaces in the Super VLAN, and the number of the learned MAC addresses is always 0, it is meaningless to configure the maximum number of MAC addresses that the Super VLAN can learn.

Displaying MAC Address Table Management

To do...	Use the command...	Remarks
Display MAC address table information	**display mac-address** [*mac-address* [**vlan** *vlan-id*] | [**dynamic** | **static**] [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**count**]]	Available in any view
	display mac-address blackhole [**vlan** *vlan-id*] [**count**]	
Display the aging timer for dynamic MAC address entries	**display mac-address aging-time**	
Display the capability of system and port to learn MAC addresses dynamically	**display mac-address mac-learning** [*interface-type interface-number*]	

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 00EA-FC35-DC71 for port GigabitEthernet 1/2/1 in VLAN 25.

Configuration procedure

# Add a static MAC address entry (showing the VLAN to which it belongs, port and status).

```
<Sysname> system-view
[Sysname] mac-address static 00ea-fc35-dc71 interface GigabitEtherne
t 1/2/1 vlan 25
```

# Set the aging timer for dynamic MAC address entries to 500 seconds.

```
[Sysname] mac-address timer aging 500
```

# Display the MAC address entry in any view.

```
[Sysname] display mac-address interface GigabitEthernet 1/2/1
MAC ADDR          VLAN ID  STATE            PORT INDEX            AGING TIME(s)

00ea-fc35-dc71   25       Config static   GigabitEthernet1/2/1         NOAGED

  ---  1 mac address(es) found on port GigabitEthernet1/2/1 ---
```

8 LINK AGGREGATION OVERVIEW

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

- "Link Aggregation" on page 69
- "Approaches to Link Aggregation" on page 71
- "Load Sharing in a Link Aggregation Group" on page 73
- "Service Loop Group" on page 74
- "Link Aggregation Port Group" on page 75

Link Aggregation

Link aggregation aggregates multiple physical Ethernet ports into one logical link, also called a logical group, to increase reliability and bandwidth. Link aggregation allows you to increase bandwidth by distributing incoming/outgoing traffic on the member ports in an aggregation group. In addition, it provides reliable connectivity because these member ports can dynamically back up each other.

This section covers these topics:

- "LACP" on page 69
- "Consistency Considerations for Ports in an Aggregation Group" on page 70

Note the following when employing link aggregation on a Switch 8800:

- Up to 31 link aggregation groups are supported on a switch.
- Up to 31 link aggregation groups are supported on 3C17526 4-port 10 Gig module. However, ports on modules of this type cannot be added to aggregation groups.
- Up to 7 link aggregation groups are supported on Advanced Modules, or in system containing an Advanced Module.
- For modules other than the above mentioned, up to 31 link aggregation groups are supported.
- To enable load being properly shared among ports in an aggregation group, make sure the number of the aggregation groups created on a module is not larger than that the module can accommodate.

LACP

The link aggregation control protocol (LACP), as defined in IEEE 802.3ad, dynamically aggregates ports and removes aggregations.

LACP interacts with its peer by sending link aggregation control protocol data units (LACPDUs).

When aggregating ports, link aggregation control automatically assigns each port an operational key based on its rate, duplex mode, and other basic configurations. In a manual or static LACP aggregation, the selected ports share the same operational key.

Consistency Considerations for Ports in an Aggregation Group

To participate in traffic sharing, member ports in an aggregation group must use the same configurations with respect to STP, QoS, GVRP, QinQ, BPDU tunnel, VLAN, port attributes, MAC address learning, and so on, as shown in the following table.

Table 6 Consistency considerations for ports in an aggregation

Category	Considerations
STP	■ State of port-level STP (enabled or disabled) ■ Attribute of the link (point-to-point or otherwise) connected to the port ■ Port path cost ■ STP priority ■ Maximum transmission rate ■ Loop protection ■ Root protection ■ Port type (whether the port is an edge port)
QoS	■ Traffic policing ■ Traffic shaping ■ Congestion avoidance ■ Strict priority (SP) queuing ■ Port priority ■ Policy setting on the port ■ Flow template
GVRP	■ GVRP state on ports (enabled or disabled) ■ GVRP registration type ■ GARP timers
QinQ	■ State of QinQ (enabled or disabled) ■ Added outer VLAN tag
BPDU tunnel	■ BPDU tunnel state on ports (enabled or disabled) ■ BPDU tunnel state for STP on ports (enabled or disabled)
VLAN	■ VLANs carried on the port ■ Default VLAN ID on the port ■ Link type of the port, which can be trunk, hybrid, or access
Port attribute	■ Port rate ■ Duplex mode ■ Up/down state of the link ■ Isolation group membership of the port

Table 6 Consistency considerations for ports in an aggregation

Category	Considerations
MAC address learning	■ MAC address learning capability ■ Setting of maximum number of MAC addresses that can be learned on the port ■ Forwarding of frames with unknown destination MAC addresses after the upper limit of the MAC address table is reached

Approaches to Link Aggregation

The options available for implementing link aggregation are described in the sections entitled "Manual Link Aggregation" on page 71 and "Static LACP link aggregation" on page 72.

Manual Link Aggregation

Overview

In the manual aggregation approach, aggregation groups are created administratively and automatic port adding/removal is not available.

On the ports in a manual aggregation, LACP is disabled and cannot be administratively enabled.

Port states in a manual aggregation

In a manual aggregation group, ports are either selected or unselected. Selected ports can receive and transmit data frames whereas unselected ones cannot. Among all selected ports, the one with the lowest port number is the master port and others are member ports.

When setting the state of ports in a manual aggregation group, the system considers the following:

- Select a port from the ports in up state, if any, in the order of full duplex/high speed, full duplex/low speed, half duplex/high speed, and half duplex/low speed, with the 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. Then, place those ports in up state with the same speed/duplex pair, link state and basic configuration in selected state and all others in unselected state.
- When all ports in the group are down, select the port with the lowest port number as the master port and set all ports (including the master) in unselected state.
- Place the ports that cannot aggregate with the master in unselected state, for example, as the result of hardware restriction.

Manual aggregation limits the number of selected ports in an aggregation group. When the limit is exceeded, the system changes the state of selected ports with greater port numbers to unselected until the number of selected ports drops under the limit.

In addition, unless the master port should be selected, a port that joins the group after the limit is reached will not be placed in selected state even if it should be in normal cases. This is to prevent the ongoing service on selected ports from being interrupted. You need to avoid the situation however as the selected/unselected state of a port may become different after a reboot.

Currently, the number of the selected ports in a manual aggregation group created on a Switch 8800 can be up to eight.

Port Configuration Considerations in manual aggregation

As mentioned above, in a manual aggregation group, only ports with configurations consistent with those of the master port can become selected. These configurations include port rate, duplex mode, link state and other basic configurations described in "Consistency Considerations for Ports in an Aggregation Group" on page 70.

You need to maintain the basic configurations of these ports manually to ensure consistency. As one configuration change may involve multiple ports, this can become troublesome if you need to do that port by port. As a solution, you may add the ports into an aggregation port group where you can make configuration for all member ports.

When the configuration of some port in a manual aggregation group changes, the system does not remove the aggregation as it does in a dynamic aggregation group; instead, it re-sets the selected/unselected state of the member ports and re-selects a master port.

Static LACP link aggregation

Overview

In the static aggregation approach, aggregation groups are created administratively and the system cannot automatically add or remove ports.

On the ports in the group, LACP is enabled and cannot be administratively disabled. After the group is removed, all the member ports in up state form one or multiple dynamic aggregations with LACP enabled.

Port states in static aggregation

In a static aggregation group, ports can be selected or unselected, where both can receive and transmit LACPDUs but only selected ports can receive and transmit data frames. The selected port with the lowest port number is the master port and all others are member ports.

All member ports that cannot aggregate with the master are placed in unselected state. These ports include those using the basic configurations different from the master port or those located on a module different from the master port because of hardware restriction.

Member ports in up state can be selected if they have the configuration same as that of the master port. The number of selected ports however, is limited in a static aggregation group. When the limit is exceeded, the local and remote systems negotiate the state of their ports as follows:

1 Compare the actor and partner system IDs that each comprises a system LACP priority plus a system MAC address as follow:

- First compare the system LACP priorities. The system with lower system LACP priority wins out.

- If they are the same, compare the system MAC addresses. The system with the smaller ID has higher priority. (the lower the LACP priority, the smaller the MAC address, and the smaller the device ID)

2 Compare the port IDs that each comprises a port LACP priority and a port number on the system with higher ID as follows:

- Compare the port LACP priorities. The port with lower port LACP priority wins out.
- If two ports with the same port LACP priority are present, compare their port numbers. The state of the ports with lower IDs then change to selected and the state of the ports with higher IDs to unselected, so does the state of their corresponding remote ports. (the lower the LACP priority, the smaller the port number, and the smaller the port ID)

Currently, the number of the selected ports in a static LACP aggregation group created on a Switch 8800 can be up to eight.

Port configuration considerations in static aggregation

Like in a manual aggregation group, in a static LACP aggregation group, only ports with configurations consistent with those of the master port can become selected. These configurations include port rate, duplex mode, link state and other basic configurations described in "Consistency Considerations for Ports in an Aggregation Group" on page 70.

You need to maintain the basic configurations of these ports manually to ensure consistency. As one configuration change may involve multiple ports, this can become troublesome if you need to do that port by port. As a solution, you may add the ports into an aggregation port group where you can make configuration for all member ports.

When the configuration of some port in a static aggregation group changes, the system does not remove the aggregation as it does in a dynamic aggregation group; instead, it re-sets the selected/unselected state of the member ports and re-selects a master port.

Load Sharing in a Link Aggregation Group

Link aggregation groups fall into load sharing aggregation groups and non-load sharing aggregation groups.

A link aggregation group is a load sharing aggregation group if it contains two or more selected ports. A link aggregation group is a non-load sharing aggregation group if it contains only one selected port.

CAUTION: *A load sharing link aggregation group remains a load sharing link aggregation group even if the number of the selected ports in it decreases to one. (The number of the selected ports in a link aggregation group decreases if the selected ports are removed from it.)*

Service Loop Group

As a Switch 8800 can accommodate different types of I/O Modules, service loop ports are needed to redirect services between I/O Modules. Through service loop ports, packets reaching an I/O Module can be passed to another one for being processed. Service loop group is used to increase the throughput for redirecting packets among I/O Modules.

Service loop group is implemented by creating link aggregation group for service loop ports. To create a service loop group, create a manual link aggregation group and then specify the services to be redirected for it. At present, the services can be IPv6 unicast services, IPv6 multicast services, and tunnel services. Service Groups IPv6 and IPv6mc are needed when mixing IPv4 modules with IPv6 modules and IPv6 traffic needs to be forwarded thru the IPv4 module. For more information, see "Centralized Mode for IPv6" on page 709.

When adding a port to a service loop group, make sure that the port supports the services specified for the service loop group and meets the following requirements:

- The port is configured only with the physical configuration (such as speed and duplex mode), QoS, and ACL. Other conflicting configurations, such as STP, are not configured.
- The port belongs to VLAN 1.

For ports that are already in a service loop group, you can perform configurations that do not conflict with the service loop group for them, such as QoS.

Table 7 Switch 8800 Options

SKU	Description	FA ports	MPLS?	IPv6?
3C17511	1-port 10GBASE-X (XENPAK)	1	No	No
3C17512	2-port 10GBASE-X (XFP)	2	No	No
3C17513	12-port 1000BASE-X (SFP)	2	No	No
3C17514	24-port 1000BASE-X (SFP)	1	No	No
3C17516	24-port 10/100/1000BASE-T (RJ45)	2	No	No
3C17525	1-port 10GBASE-X (XENPAK) Advanced	2	Yes	No
3C17526	4-port 10GBASE-X (XFP)	1	No	No
3C17527	2-port 10GBASE-X (XFP) Advanced	2		No
3C17528	48-port 10/100/1000BASE-T (RJ45)	1	No	Yes
3C17530	24-port 1000BASE-X (SFP) Advanced	2	Yes	No
3C17531	24-port 10/100/1000BASE-T (RJ45) Advanced	2	Yes	No
3C17532	48-port 10/100/1000BASE-T (RJ45) Access	1	No	Yes
3C17533	24-port 1000BASE-X (SFP) IP6	2	No	Yes
3C17534	24-port 10/100/1000BASE-T (RJ45) IP6	2	No	Yes
3C17536	4-port 10GBASE-X (XFP) IP6	4	No	Yes
3C17537	2-port 10GBASE-X (XFP) IP6	2	No	Yes
3C17538	48-port 1000BASE-X (SFP) IP6	1	No	Yes

Link Aggregation Port Group

As mentioned earlier, in a manual or static aggregation group, a port can be selected only when its configuration is the same as that of the master port in terms of duplex/speed pair, link state, and other basic configurations. Their configuration consistency requires administrative maintenance, which is troublesome after you change some configuration.

To simplify configuration, port-groups are provided allowing you to configure for all ports in individual groups at one time. One example of port-groups is aggregation port group.

Upon creation or removal of a link aggregation group, an aggregation port-group which cannot be administratively created or removed is automatically created or removed. In addition, you can only assign/remove a member port to/from an aggregation port-group by assigning/removing it from the corresponding link aggregation group.

For more information about port-groups, refer to *"Configuring a Port Group" on page 58*.

CAUTION: *Ports in a link aggregation group with their configuration not consistent with that of the master port are unselected ports. Unselected port may affect the ongoing services. So if you want to modify the configuration of aggregation ports, do it in aggregation port view.*

9 LINK AGGREGATION CONFIGURATION

When performing link aggregation configuration, go to these sections for information you are interesting in:

- "Configuring Link Aggregation" on page 77
- "Displaying and Maintaining Link Aggregation" on page 80
- "Link Aggregation Configuration Example" on page 80

Configuring Link Aggregation

CAUTION: *If an abnormal operating state of a port in a dynamic aggregation group caused by the existence of an empty aggregation group, you can try the following steps to correct the error: (a) delete the empty aggregation group; (b) disable LACP on the port in question, and (c) enable LACP on the port in question again.*

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

- "Configuring a Manual Link Aggregation Group" on page 77
- "Configuring a Static LACP Link Aggregation Group" on page 78
- "Configuring an Name for a Link Aggregation Group" on page 79
- "Configuring a Service Loop Group" on page 79
- "Entering Aggregation Port Group View" on page 80

Configuring a Manual Link Aggregation Group

Follow these steps to configure a manual aggregation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Create a manual aggregation group	**link-aggregation group** *agg-id* **mode manual**	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	--
Assign the Ethernet port to the aggregation group	**port link-aggregation group** *agg-id*	Required

Note that:

- You may create a manual aggregation group by changing the type of a static or dynamic aggregation group that has existed. If the specified group contains ports, its group type changes to manual with LACP disabled on its member ports; if not, its group type directly changes to manual.

- An aggregation group cannot include ports with static MAC addresses, 802.1x-enabled ports, MAC address authentication-enabled ports, or POS interfaces. Besides, ports operating as upstream ports of isolation groups cannot be added to manual or static aggregation groups.
- After you assign an LACP-enabled port to a manual aggregation group, its LACP will be disabled.
- You can remove all ports in a manual aggregation group by removing the group.
- If this group contains only one port, you can remove the port only by removing the group.

To guarantee a successful aggregation, ensure that the ports at the two ends of each link to be aggregated are consistent in selected/unselected state.

Configuring a Static LACP Link Aggregation Group

Follow these steps to configure a static aggregation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Configure the system LACP priority	**lacp system-priority** *system-priority*	Optional 32768 by default.
Create a static LACP aggregation group	**link-aggregation group** *agg-id* **mode static**	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	--
Configure the port LACP priority	**lacp port-priority** *port-priority-value*	Optional 32768 by default.
Assign the Ethernet port to the aggregation group	**port link-aggregation group** *agg-id*	Required

Note that:

- You can create a static aggregation group by changing the type of an existing link aggregation group. An LACP aggregation group with ports contained in it can only be changed to a static LACP aggregation group. In this case, ports in it are still LACP-enabled. For a manual aggregation group containing no port, you can change it to a static LACP aggregation group.
- Changing LACP priority globally or on a port may change the state (selected/unselected) of the ports in a static LACP aggregation group.
- An aggregation group cannot include ports with static MAC addresses, 802.1x-enabled ports, MAC address authentication-enabled ports, or POS interfaces. Besides, ports operating as upstream ports of isolation groups cannot be added to manual or static aggregation groups.
- After you assign an LACP-disabled port to a static LACP aggregation group, it becomes LACP-enabled.

When making configuration, be aware that after a load-balancing aggregation group changes to a non-load balancing group due to resources exhaustion, either of the following may happen:

- *Forwarding anomaly resulted from inconsistency of the two ends in the number of selected ports.*
- *Some protocols such as GVRP malfunction because the state of the remote port connected to the master port is unselected.*

Configuring an Name for a Link Aggregation Group

Follow these steps to configure a name for an aggregation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Configure a name for a link aggregation group	**link-aggregation group** *agg-id* **description** *agg-name*	Required Not configured by default.

CAUTION: *When configuring a name for a link aggregation group, make sure the ID of the link aggregation group is available. You can obtain the ID of a link aggregation group using the* ***display link-aggregation summary*** *command or the* ***display link-aggregation interface*** *command.*

Configuring a Service Loop Group

Follow these steps to configure a service loop group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Create a manual aggregation group	**link-aggregation group** *agg-id* **mode manual**	Required
Specify the aggregation group as a service loop group	**link-aggregation group** *agg-id* **service-type** { { **ipv6** | **ipv6mc** } * | **tunnel** }	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	--
Assign the Ethernet port to the service loop group	**port link-aggregation group** *agg-id*	Required

When a service loop group contains only one port, you can remove the port only by removing the group.

- *There can only be up to one service loop group for each service loop group type.*
- *There can only be up to eight Ethernet ports valid for each service loop group type.*
- *An Ethernet port can be added to a service loop group only when STP is not enabled on it.*
- *You can change the type of an existing service loop group. The operation fails if it is currently referenced by a module or the service loop group contains ports whose attributes conflict with the intended service type.*
- *You can use the* ***undo link-aggregation group*** *command to remove an service loop group. The operation fails if it is currently referenced by a module.*

- *For a service loop group containing only one port, you can only remove the port from the service loop group by removing the service loop group.*

Entering Aggregation Port Group View

In aggregation port group view, you can make configuration for all the member ports in a link aggregation group at one time.

Follow these steps to enter aggregation port group view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter aggregation port group view	**port-group aggregation** *agg-id*	--

CAUTION: *In aggregation port group view, you can configure aggregation related settings such as STP, VLAN, QoS, GVRP, QinQ, BPDU tunnel, MAC address learning, but cannot add or remove member ports.*

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 detailed information about link aggregation for the specified port or ports	**display link-aggregation interface** *interface-type interface-number* [**to** *interface-type interface-number*]	Available in any view
Display information about the specified or all service loop groups	**display link-aggregation service-type** [*agg-id*]	Available in any view
Display summaries for all link aggregation groups	**display link-aggregation summary**	Available in any view
Display detailed information about specified or all link aggregation groups	**display link-aggregation verbose** [*agg-id*]	Available in any view
Clear the statistics about LACP for specified or all ports	**reset lacp statistics** [**interface** *interface-type interface-number* [**to** *interface-type interface-number*]]	Available in user view

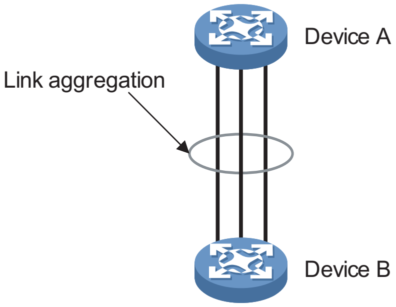
Link Aggregation Configuration Example

Network requirements

Switch A aggregates ports Ethernet 1/1/1 through Ethernet 1/1/3 to form one link connected to Switch B and performs load sharing among these ports.

Network diagram

Figure 22 Network diagram for link aggregation



Configuration procedure

- *This example only describes how to configure link aggregation on Switch A. To achieve link aggregation, do the same on Switch B.*
- *Manual aggregation group, static aggregation group, and dynamic aggregation group can all be used here.*

1 In manual aggregation approach

# Create manual aggregation group 1.

```
<SwitchA> system-view
[SwitchA] link-aggregation group 1 mode manual
```

# Add ports Ethernet 1/1/1 through Ethernet 1/1/3 to the aggregation group.

```
[SwitchA] interface ethernet 1/1/1
[SwitchA-Ethernet1/1/1] port link-aggregation group 1
[SwitchA-Ethernet1/1/1] quit
[SwitchA] interface ethernet 1/1/2
[SwitchA-Ethernet1/1/2] port link-aggregation group 1
[SwitchA-Ethernet1/1/2] quit
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] port link-aggregation group 1
```

2 In static aggregation approach

# Create static aggregation group 1.

```
<SwitchA> system-view
[SwitchA] link-aggregation group 1 mode static
```

# Add ports Ethernet 1/1/1 through Ethernet 1/1/3 to the aggregation group.

```
[SwitchA] interface ethernet 1/1/1
[SwitchA-Ethernet1/1/1] port link-aggregation group 1
[SwitchA-Ethernet1/1/1] quit
[SwitchA] interface ethernet 1/1/2
[SwitchA-Ethernet1/1/2] port link-aggregation group 1
[SwitchA-Ethernet1/1/2] quit
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] port link-aggregation group 1
```

3 In dynamic aggregation approach

Enable LACP on ports Ethernet 1/1/1 through Ethernet 1/1/3.

```
<SwitchA> system-view
[SwitchA] interface ethernet 1/1/1
[SwitchA-Ethernet1/1/1] lacp enable
[SwitchA-Ethernet1/1/1] quit
[SwitchA] interface ethernet 1/1/2
[SwitchA-Ethernet1/1/2] lacp enable
[SwitchA-Ethernet1/1/2] quit
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] lacp enable
```

To have the three ports form one dynamic aggregation group, you must make the same basic configurations for them.

10 PORT MIRRORING CONFIGURATION

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

Introduction to Port Mirroring

Classification of Port Mirroring

There are two kinds of port mirroring: local port mirroring and remote port mirroring.

- Local port mirroring copies packets passing through one or more ports (known as source ports) of a device to the monitor port (also destination port) for analysis and monitoring purpose. In this case, the source ports and the destination port are located on the same device.
- Remote port mirroring implements port mirroring between multiple devices. That is, the source ports and the destination port can be located on different devices in a network. Currently, remote port mirroring can only be implemented on Layer 2.

Implementing Port Mirroring

Port mirroring is implemented through port mirroring groups, which fall into these three categories: local port mirroring group, remote source port mirroring group, and remote destination port mirroring group.

Port Mirroring can be implemented as follows:

- Local port mirroring is implemented by local port mirroring groups. The source ports and the destination port are on the same device. In this case, packets passing through the source ports are duplicated and then are forwarded to the monitor port.
- Remote port mirroring can be implemented by remote source port mirroring groups and remote destination port mirroring groups. The source and destination ports are on different devices. In this case, packets passing through source ports are broadcast in remote port mirroring VLANs through reflector ports, and those with their VLAN IDs being the remote port mirroring VLAN IDs of the remote port mirroring groups are forwarded to the destination port of

the remote destination port mirroring group by the remote device receiving the packets.

- Port mirroring group supports inter-module mirroring, which means that the destination port and source ports can be located on different modules of a device. In addition, a destination port can monitor multiple source ports simultaneously.

- *Currently, Switch 8800s support the three types of port mirroring groups mentioned above as well as inter-module mirroring.*
- *As for the four Ten-GigabitEthernet ports on 3C17526 4-port 10 Gig module, port mirroring can only be implemented between port 1 and 2 (Ten-GigabitEthernet2/1/1 and Ten-GigabitEthernet2/1/2), and port 3 and 4 (Ten-GigabitEthernet2/1/3 and Ten-GigabitEthernet2/1/4.)*

CAUTION: *As port mirroring conflicts with STP, RSTP, and MSTP, do not enable STP, RSTP, or MSTP on destination mirroring ports.*

Configuring Local Port Mirroring

Follow these steps to configure local port mirroring:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Create a local mirroring group		**mirroring-group** *groupid* **local**	Required
Add ports to the port mirroring group as source ports	In system view	**mirroring-group** *groupid* **mirroring-port** *mirroring-port-list* { **inbound** | **outbound** | **both** }	You can add ports to a port mirroring group as source ports in either system view or interface view. They achieve the same purpose.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **mirroring-port** { **inbound** | **outbound** | **both** }	
		quit	
Add a port to the mirroring group as the destination port	In system view	**mirroring-group** *groupid* **monitor-port** *monitor-port-id*	You can add a destination port to a port mirroring group in either system view or interface view. They achieve the same purpose.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **monitor-port**	

- *A local mirroring group is effective only when it has both source ports and destination port configured.*
- *Member ports of existing port mirroring groups cannot be destination ports.*
- *A port mirroring group can contain multiple source ports and only one destination port.*
- *A port can belong to only one port mirroring group.*

Configuring Remote Port Mirroring

While configuring remote port mirroring, you need to configure the remote source port mirroring group and the remote destination port mirroring group on both devices.

Configuring a Remote Source Mirroring Group

You need to configure source ports, reflector ports, and remote port mirroring VLAN for a remote source mirroring group.

Follow these steps to configure a remote source port mirroring group:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Create a remote port mirroring VLAN		**vlan** *vlan-id*	Optional
Return to system view		**quit**	-
Create a remote source mirroring group		**mirroring-group** *groupid* **remote-source**	Required
Add ports to the mirroring group as source ports	In system view	**mirroring-group** *groupid* **mirroring-port** *mirroring-port-list* { **inbound** | **outbound** | **both** }	You can add ports to a source mirroring group in either system view or interface view. They achieve the same purpose.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **mirroring-port** { **inbound** | **outbound** | **both** }	
		quit	
Add a port to the mirroring group as the reflector port	In system view	**mirroring-group** *groupid* **reflector-port** *reflector-port-id*	You can add ports to a source port mirroring group in either system view or interface view. They achieve the same purpose.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		mirroring-group *groupid* **reflector-port**	
		quit	
Configure the remote port mirroring VLAN for the mirroring group		**mirroring-group** *groupid* **remote-probe vlan** *rprobe-vlan-id*	Required

- *All the ports (including the source ports and the reflector port) of a remote source mirroring group can only belong to the same single device.*
- *Do not add source ports to remote port mirroring VLANs for fear of interrupting device operation.*
- *Only the ports that are of the access type and belong to the default VLANs can be reflector ports. Member ports of existing port mirroring groups or destination stream mirroring ports cannot be reflector ports.*
- *Do not enable any of the following on a reflector port: 802.1x, QinQ, port loopback, and service loopback. Besides, to avoid interrupting device operation, do not enable static ARP or MAC address learning on reflector ports either.*
- *Outbound ports cannot be the reflector port.*

- *A remote source mirroring group can have only one reflector port.*
- *A port can be configured as a reflector port only when it operates with the following settings being the defaults: operation mode (half duplex/full duplex), port speed, and MDI setting.*
- *Use a remote port mirroring VLAN for remote port mirroring only.*
- *Only existing static VLANs can be configured as remote port mirroring VLANs. To remove a VLAN operating as a remote port mirroring VLAN, you need to restore it to a normal VLAN first. A remote port mirroring group gets invalid if the corresponding remote port mirroring VLAN is removed.*
- *A port can belong to only one port mirroring group. A VLAN can be the remote port mirroring VLAN of only one port mirroring group.*
- *Disable MAC address learning in remote port mirroring VLANs to ensure the functionality of remote port mirroring.*

Configuring a Remote Destination Port Mirroring Group

You need to configure destination ports and remote port mirroring VLAN for a remote destination 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 port mirroring VLAN and enter VLAN view		**vlan** *vlan-id*	Required
Disable MAC address learning in the remote port mirroring VLAN		**mac-address max-mac-count** *count*	Required With *count* being 0, MAC address learning is disabled in target remote port mirroring VLAN.
Return to system view		**quit**	-
Create a remote destination port mirroring group		**mirroring-group** *groupid* **remote-destination**	Required
Configure the remote port mirroring VLAN for the port mirroring group		**mirroring-group** *groupid* **remote-probe vlan** *rprobe-vlan-id*	Required
Add a port to the port mirroring group as the destination port	In system view	**mirroring-group** *groupid* **monitor-port** *monitor-port-id*	You can add a port to a remote port mirroring group as the destination port in either system view or interface view. They achieve the same purpose.
	In Ethernet interface view	**interface** *interface-type interface-number*	
		[**mirroring-group** *groupid*] **monitor-port**	
		quit	
Enter destination Ethernet interface view		**interface** *interface-type interface-number*	-

To do...		Use the command...	Remarks
Add the destination port to the remote port mirroring VLAN	The port is an access port	**port access vlan** *rprobe-vlan-id*	Perform one of these three operations according to the port type.
	The port is a trunk port	**port trunk permit vlan** *rprobe-vlan-id*	
	The port is a hybrid port	**port hybrid vlan** *rprobe-vlan-id* { **tagged** | **untagged** }	

- *Only existing static VLANs can be configured as remote port mirroring VLANs. To remove a VLAN operating as a remote port mirroring VLAN, you need to restore it to a normal VLAN first. A remote port mirroring group gets invalid if the corresponding remote port mirroring VLAN is removed.*
- *A port can belong to only one port mirroring group. A VLAN can be the remote port mirroring VLAN of only one port mirroring group.*
- *Member ports of existing port mirroring groups cannot be destination ports.*
- *When configuring remote port mirroring VLAN and destination ports for a remote destination mirroring group, add the destination ports to the remote port mirroring VLAN.*
- *Use a remote port mirroring VLAN for remote port mirroring only.*
- *Disable MAC address learning in remote port mirroring VLANs to ensure the functionality of remote port mirroring.*

Displaying Port Mirroring

To do...	Use the command...	Remarks
Display the configuration of a port mirroring group	**display mirroring-group** { *groupid* | **local** | **remote-source** | **remote-destination** | **all** }	Available in any view

Port Mirroring Configuration Example

Local Port Mirroring Configuration Example

Network requirements

The user's network is described as follows:

- Host A is connected to port Ethernet 1/1/1 of Switch C through Switch A.
- Host B is connected to port Ethernet 1/1/2 of Switch C through Switch B.
- The Server is connected to port Ethernet 1/1/3 of Switch C.

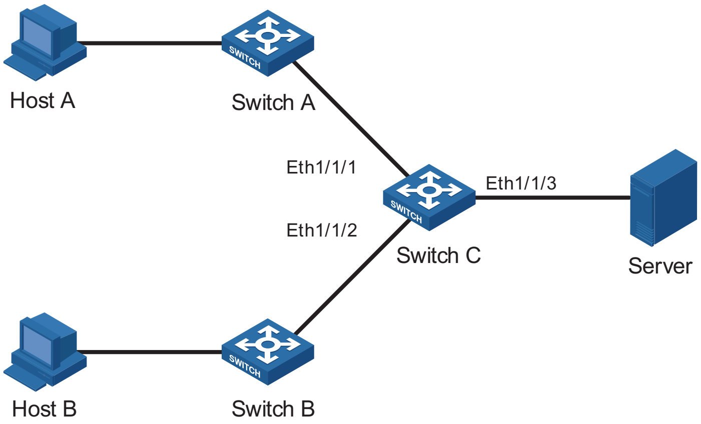
It is desired to monitor the packets of Host A and Host B on the Server.

This can be achieved by configuring a local port mirroring group. Perform the following configuration on Switch C.

- Configure port Ethernet 1/1/1 and Ethernet 1/1/2 as source mirroring ports.
- Configure port Ethernet 1/1/3 as the destination mirroring port.

Network diagram

Figure 23 Network diagram for local port mirroring configuration



Configuration procedure

1 Configure Switch C.

# Enter system view.

```
<Sysname> system-view
```

# Create a local port mirroring group.

```
[Sysname] mirroring-group 1 local
```

# Add port Ethernet 1/1/1 and Ethernet 1/1/2 to the port mirroring group as source ports. Add port Ethernet 1/1/3 to the port mirroring group as the destination port.

```
[Sysname] mirroring-group 1 mirroring-port ethernet 1/1/1 ethernet 1/1/2 both
[Sysname] mirroring-group 1 monitor-port ethernet 1/1/3
```

# Display the configuration of all the port mirroring groups.

```
[Sysname] display mirroring-group all
mirroring-group 1:
    type: local
    status: active
    mirroring port:
        Ethernet1/1/1  both
        Ethernet1/1/2  both
    monitor port: Ethernet1/1/3
```

After finishing the configuration, you can monitor all the packets received and sent by Host A and Host B on the Server.

Remote Port Mirroring Configuration Example

Network requirements

The user's network is described as follows:

- Host A is connected to port Ethernet 1/1/1 of Switch A.
- Host B is connected to port Ethernet 1/1/2 of Switch A.

- Port Ethernet 1/1/3 of Switch A and port Ethernet 1/1/1 of Switch B are two trunk ports. They are connected together.
- Port Ethernet 1/1/2 of Switch B and port Ethernet 1/1/1 of Switch C are two trunk ports. They are connected together.
- The Server is connected to port Ethernet 1/1/2 of Switch C.

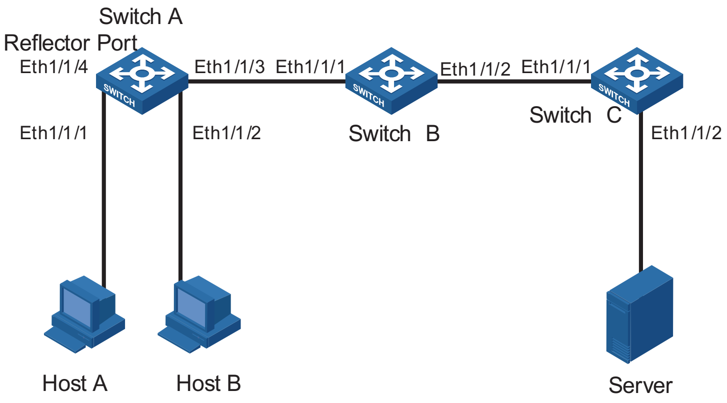
It is desired to monitor packets of Host A and Host B on the Server.

This can be achieved by configuring remote port mirroring groups, as described below.

- On Switch A, create a remote source mirroring group; create VLAN 2 and configure it as the remote port mirroring VLAN; add port Ethernet 1/1/1 and Ethernet 1/1/2 to the port mirroring group as two source ports. Configure port Ethernet 1/1/4 as the reflector port.
- Configure port Ethernet 1/1/3 of Switch A, port Ethernet 1/1/1 and Ethernet 1/1/2 of Switch B, and port Ethernet 1/1/1 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 Ethernet 1/1/2, to which the server is connected, as the destination port.

Network diagram

Figure 24 Network diagram for remote port mirroring configuration



Configuration procedure

1 Configure Switch A.

# Enter system view.

```
<Sysname> system-view
```

# Create a remote source port mirroring group.

```
[Sysname] mirroring-group 1 remote-source
```

# Create VLAN 2.

```
[Sysname] vlan 2
[Sysname-vlan2] quit
```

# Configure VLAN 2 as the remote port mirroring VLAN of the remote port mirroring group. Add port Ethernet 1/1/1 and Ethernet1/1/2 to the remote port mirroring group as source ports. Configure port Ethernet 1/1/4 as the reflector port.

```
[Sysname] mirroring-group 1 remote-probe vlan 2
[Sysname] mirroring-group 1 mirroring-port ethernet 1/1/1 ethernet 1/1/2 both
[Sysname] mirroring-group 1 reflector-port Ethernet ethernet 1/1/4
```

# Configure port Ethernet 1/1/3 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[Sysname] interface ethernet 1/1/3
[Sysname-Ethernet1/1/3] port link-type trunk
[Sysname-Ethernet1/1/3] port trunk permit vlan 2
```

2 Configure Switch B.

# Create VLAN 2 and disable MAC address learning in it.

```
<Sysname> system-view
[Sysname] vlan 2
[Sysname-vlan2] mac-address max-mac-count 0
[Sysname-vlan2] quit
```

# Configure port Ethernet 1/1/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port link-type trunk
[Sysname-Ethernet1/1/1] port trunk permit vlan 2
```

# Configure port Ethernet 1/1/2 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[Sysname-Ethernet1/1/1] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] port link-type trunk
[Sysname-Ethernet1/1/2] port trunk permit vlan 2
```

3 Configure Switch C.

# Enter system view.

```
<Sysname> system-view
```

# Configure port Ethernet 1/1/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port link-type trunk
[Sysname-Ethernet1/1/1] port trunk permit vlan 2
[Sysname-Ethernet1/1/1] quit
```

# Create a remote destination port mirroring group.

```
[Sysname] mirroring-group 1 remote-destination
```

# Create VLAN 2 and disable MAC address learning in it. Add port Ethernet1/1/2 to it.

```
[Sysname] vlan 2
[Sysname-vlan2] mac-address max-mac-count 0
[Sysname-vlan2] port ethernet 1/1/2
[Sysname-vlan2] quit
```

# Configure VLAN 2 as the remote port mirroring VLAN of the remote destination port mirroring group. Add port Ethernet 1/1/2 to the remote destination port mirroring group as the destination port.

```
[Sysname] mirroring-group 1 remote-probe vlan 2
[Sysname] mirroring-group 1 monitor-port ethernet 1/1/2
```

After finishing the configuration, you can monitor all the packets received and sent by Host A and Host B on the Server.

11 MSTP CONFIGURATION

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

MSTP Overview

Introduction to STP

Functions of STP

Spanning tree protocol (STP) aims to eliminate loops in a local area network (LAN). Devices running this protocol detect loops in the network by exchanging information with one another and eliminate the loops detected by blocking certain ports until the loop network is pruned into one with tree topology. As a network with tree topology is loop-free, it prevents packets in it from being duplicated and forwarded endlessly.

Protocol packets of STP

STP uses bridge protocol data units (BPDUs) for exchanging information. BPDU is also known as configuration packets (messages).

STP identifies the network topology by transmitting BPDUs between STP compliant network devices. BPDUs contain sufficient information for the network devices to complete the spanning tree computing.

In STP, BPDUs come in two types:

- Configuration BPDUs, used to maintain the spanning tree topology.
- Topology change notification (TCN) BPDUs, used to notify 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" has been introduced in STP.

An STP network has only one root bridge. The root bridge is globally significant in the entire network, and is the logical center of the network. However, it need not be the physical center of the network. The root bridge may change when the network topology changes. After the network converges, only the root bridge sends out protocol packets known as configuration BPDUs at a specific interval, and the other devices just relay these configuration BPDUs to keep the topology stable.

2 Root port

In an STP network, a root port is a port on a non-root bridge device. Among the ports on an STP-enabled device, the root port has the lowest path cost to the root bridge.

The root port takes charge of communicating with the root bridge. A non-root-bridge device has one and only one root port. The root bridge has no root port.

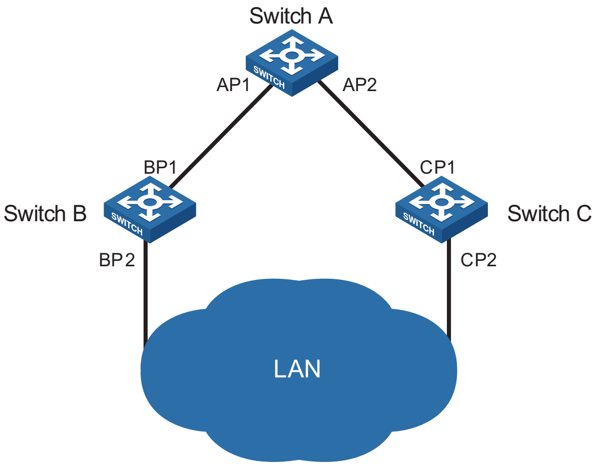
3 Designated bridge and designated port

Refer to Table 8 for the description of designated bridge and designated port.

Table 8 Description of designated bridge and designated port

Classification	Designated bridge	Designated port
For a device	The device directly connected with this device and responsible for forwarding configuration BPDUs	The port through which the designated bridge forwards configuration BPDUs to this device
For a LAN	The device responsible for forwarding configuration BPDUs to this LAN segment	The port through which the designated bridge forwards configuration BPDUs to this LAN segment

Figure 25 A schematic diagram of designated bridges and designated ports



As shown in Figure 25, AP1 and AP2, BP1 and BP2, and CP1 and CP2 are ports on Switch A, Switch B, and Switch C.

- If Switch A forwards configuration BPDUs to Switch B through AP1, the designated bridge for Switch B is Switch A, and the designated port is AP1 on Switch A.
- Two Switches are connected to the LAN: Switch B and Switch C. If Switch B forwards configuration BPDUs to the LAN, the designated bridge for the LAN is Switch B, and the designated port is BP2 on Switch B.

All the ports on the root bridge are designated ports.

How STP works

STP identifies the network topology by transmitting configuration BPDUs between network devices.

Configuration BPDUs contain sufficient information for network devices to complete the spanning tree computing. A configuration BPDU mainly contains the following information:

- Root bridge ID, formed by root bridge priority and MAC address
- Root path cost
- Designated bridge ID, formed by designated bridge priority and MAC address
- Designated port ID, formed by designated port priority and port name
- Message age: age of the configuration BPDU
- Max age: maximum age of the configuration BPDU
- Hello time: interval to send configuration BPDUs
- Forward delay: state transition delay of the port

1 A simplified STP computing model

For the convenience of description, the description and examples below involve only four parts of a configuration BPDU:

- *Root bridge ID (in the form of device priority)*

- *Root path cost*
- *Designated bridge ID (in the form of device priority)*
- *Designated port ID (in the form of port ID)*
- Initial state

Upon initialization of a device, each port generates a configuration BPDU with itself as the root, in which the root path cost is 0, designated bridge ID is the device ID, and the designated port is the local port.

- 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 9 Selection of the optimum configuration BPDU

Step	Description
1	Upon receiving a configuration BPDU on a port, the device performs the following processing: ■ If the received configuration BPDU has a lower priority than that of the configuration BPDU generated by the port, the device will discard the received configuration BPDU without doing any processing on 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 will replace 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.

Rules for configuration BPDU comparison:

- *The configuration BPDU that has the lowest root bridge ID has the highest priority.*
- *For configuration BPDUs with the same root bridge ID, they will be compared by their root path costs. If the root path cost in a configuration BPDU plus the path cost corresponding to this port is S, the configuration BPDU with the smallest S value has the highest priority.*
- *For configuration BPDUs with the same root bridge ID and the same root path cost, they will be compared by their designated bridge IDs, then their designated port IDs, and then the IDs of the ports through which they are received. The smaller the ID, the higher the message priority.*
- Selection of the root bridge

At network initialization, each STP-compliant 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 one another’s root bridge ID. The device with the smallest root bridge ID is elected as the root bridge.

- Selection of the root port and designated ports

The process of selecting the root port and designated ports is as follows:

Table 10 Selection of the root port and designated ports

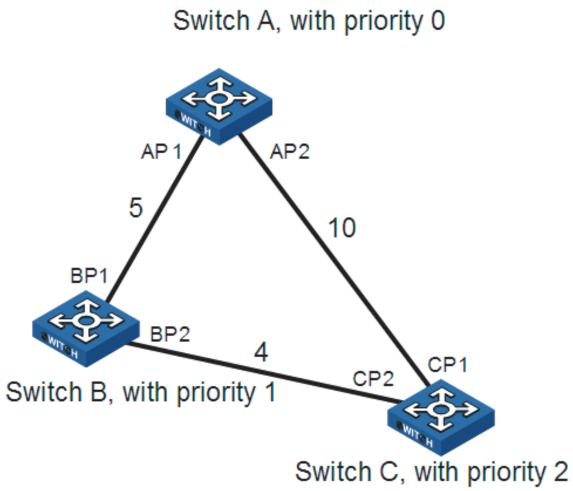
Step	Description
1	The root port is the port through which the optimum configuration BPDU was received.
2	Based on the configuration BPDU and the path cost of the root port, the device generates a designated port configuration BPDU for each of the rest ports as follows: ■ Using the root bridge ID of the configuration BPDU of the root port as the root bridge ID. ■ Using the sum of the root path cost of the configuration BPDU of the root port and the path cost corresponding to the root port as the root path cost. ■ Using the local device ID as the designated bridge ID. ■ Using the local port ID as the designated port ID.
3	The device compares the configuration BPDUs generated for its ports with the configuration BPDUs received on the corresponding port and acts as follows. ■ If the received configuration BPDU is superior, the device will block this port without changing its configuration BPDU, so that the port only receives configuration BPDUs, but does not forward packets or send configuration BPDUs. ■ If the generated configuration BPDU is superior, this port will serve as the designated port, and the device sends the generated configuration BPDU through the port periodically.

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 only receive STP packets but do not forward user traffic.

Once the root bridge, the root port on each non-root bridge and designated ports have been successfully elected, the entire tree-shaped topology has been constructed.

The following is an example of how the STP algorithm works. The specific network diagram is shown in Figure 26, where the priority of Switch A is 0, the priority of Switch B is 1, the priority of Switch C is 2, and the path costs of the links are 5, 10 and 4.

Figure 26 Network diagram for STP algorithm


Switch A, with priority 0
AP 1
AP2
5
10
BP1
BP2
4
CP2
CP1
Switch B, with priority 1
Switch C, with priority 2

- Initial state of each device

The following table shows the initial state of each device.

Table 11 Initial state of each device

Device	Port ID	BPDU of the port
Switch A	AP1	{0, 0, 0, AP1}
	AP2	{0, 0, 0, AP2}
Switch B	BP1	{1, 0, 1, BP1}
	BP2	{1, 0, 1, BP2}
Switch 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 12 Comparison process and result on each device

Device	Comparison process	BPDU of the port after comparison

Switch A	■ Port AP1 receives a configuration BPDU from Switch B (that is, {1, 0, 1, BP1}). As the configuration BPDU of the local port (that is, {0, 0, 0, AP1}) is superior to the received configuration BPDU, the received configuration BPDU is discarded.
	■ Port AP2 receives a configuration BPDU from Switch C (that is, {2, 0, 2, CP1}). As the BPDU of the local port (that is, {0, 0, 0, AP2}) is superior to the received configuration BPDU, the received configuration BPDU is discarded.
	■ Switch A finds that both the root bridge and designated bridge in the configuration BPDUs of all its ports are Switch A 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}

Switch B	■ Port BP1 receives a configuration BPDU from Switch A (that is, {0, 0, 0, AP1}). As the received configuration BPDU is superior to that of the local port (that is, {1, 0, 1, BP1}), Switch B uses the received configuration BPDU as the configuration BPDU of BP1.
	■ Port BP2 receives a configuration BPDU from Switch C (that is, {2, 0, 2, CP2}). As the configuration BPDU of the local port (that is, {1, 0, 1, BP2}) is superior to the received configuration BPDU, Switch B discards the received configuration BPDU.
		BP1: {0, 0, 0, AP1}
		BP2: {1, 0, 1, BP2}

	■ Switch B compares the configuration BPDUs of all its ports, and determines that the configuration BPDU of BP1 is the optimum one. So, BP1 acts as the root port, the configuration BPDUs of which remains unchanged.
	■ Based on the configuration BPDU of BP1 and the path cost of the root port (5), Switch B generates a designated port configuration BPDU for BP2 (that is, {0, 5, 1, BP2}).
	■ Switch B compares the generated configuration BPDU (that is, {0, 5, 1, BP2}) with the configuration BPDU of BP2. As the former is superior, BP2 acts as a designated port, and Switch B sends the generated configuration BPDU through BP2 periodically.
		Root port BP1:
		{0, 0, 0, AP1}
		Designated port BP2:
		{0, 5, 1, BP2}

Table 12 Comparison process and result on each device

Device	Comparison process	BPDU of the port after comparison

Switch C	■ Port CP1 receives a configuration BPDU from Switch A (that is, {0, 0, 0, AP2}). As the received configuration BPDU is superior to that of the local port (that is, {2, 0, 2, CP1}), Switch C uses the received configuration BPDU as the configuration BPDU of CP1.
	■ Port CP2 receives a configuration BPDU from Switch B (that is, {1, 0, 1, BP2}) before the configuration BPDU is updated on BP2. As the received configuration BPDU is superior to that of the local port (that is, {2, 0, 2, CP2}), Switch C uses the received configuration BPDU as the configuration BPDU of CP2.
		CP1: {0, 0, 0, AP2}
		CP2: {1, 0, 1, BP2}

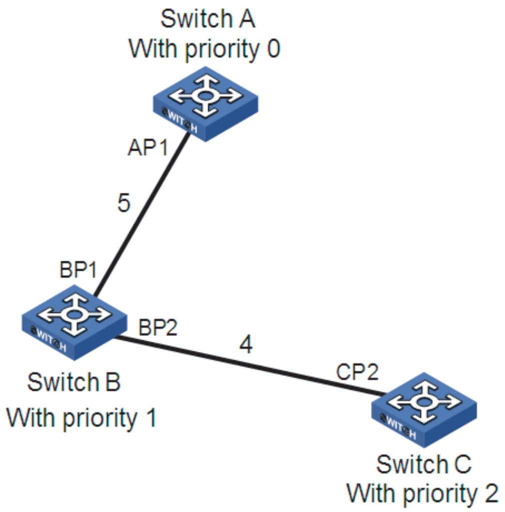
	By comparison:
	■ The configuration BPDUs of CP1 is the optimum configuration BPDU, so CP1 acts as the root port, the configuration BPDUs of which remains unchanged.
	■ Switch C generates a designated port configuration BPDU (that is, {0, 10, 2, CP2}) and compare it with the configuration BPDU of CP2. As the former is superior, CP2 acts as a designated port and Switch C sends the generated configuration BPDU through CP2 periodically.
		Root port CP1:
		{0, 0, 0, AP2}
		Designated port CP2:
		{0, 10, 2, CP2}

	■ Next, port CP2 receives the updated configuration BPDU of Switch B (that is, {0, 5, 1, BP2}). As the received configuration BPDU is superior to the local one, Switch C launches a BPDU update process.
	■ At the same time, port CP1 receives configuration BPDUs periodically from Switch A. Switch C does not launch an update process after comparison.
		CP1: {0, 0, 0, AP2}
		CP2: {0, 5, 1, BP2}

	By comparison:
	■ Because the root path cost of CP2 (which is 9) is smaller than the root path cost of CP1 (which is 10), the configuration BPDU of CP2 is the optimum BPDU, and CP2 acts as the root port, the configuration BPDUs of which remains unchanged.
	■ After the comparison between the configuration BPDU of CP1 and the generated designated port configuration BPDU, port CP1 is blocked, with the configuration BPDU of the port remaining unchanged, and the port will not receive data from Switch A until a spanning tree computing process is triggered by a new condition, for example, the link between Switch B and Switch C becomes down.
		Blocked port CP1:
		{0, 0, 0, AP2}
		Root port CP2:
		{0, 5, 1, BP2}

After the comparison processes described in the table above, a spanning tree with Switch A as the root bridge is stabilized, as shown in Figure 27.

Figure 27 A spanning tree with Switch A as the root bridge



To facilitate description, the spanning tree computing process in this example is simplified, while the actual process is more complicated.

2 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 interval of hello time.
- If it is the root port that received the configuration BPDU and the received configuration BPDU is superior to the configuration BPDU of the port, the device will increase message age carried in the configuration BPDU by a certain rule and start a timer to time the configuration BPDU while it sends out this configuration BPDU through the designated port.
- If the configuration BPDU received on the designated port is inferior to the configuration BPDU of the local port, the port will immediately send out its superior 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 BPDU. This triggers a new spanning tree computing process so that a new path is established to restore the network connectivity.

However, the newly computed 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 through the old path. If the new root port and designated port begin to forward data as soon as they are elected, a temporary loop may occur. For this reason, STP uses a state transition mechanism. Namely, a newly elected root port or designated port requires twice the forward delay time before transitioning to the forwarding state, when the new configuration BPDU has been propagated throughout the network.

Introduction to MSTP

Why MSTP

1 Disadvantages 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 transitioning to the forwarding state, even if it is a port on a point-to-point link or it is an edge port.

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 reach the final topology stability.

- *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 (a port is an edge port if it is not connected to the other devices directly or indirectly) 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 support 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 VLANs, 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 support for rapid network convergence, it also allows data flows of different VLANs to be forwarded along their own paths, thus providing a better load sharing mechanism for redundant links. For description about VLANs, refer to *"VLAN Overview" on page 155*.

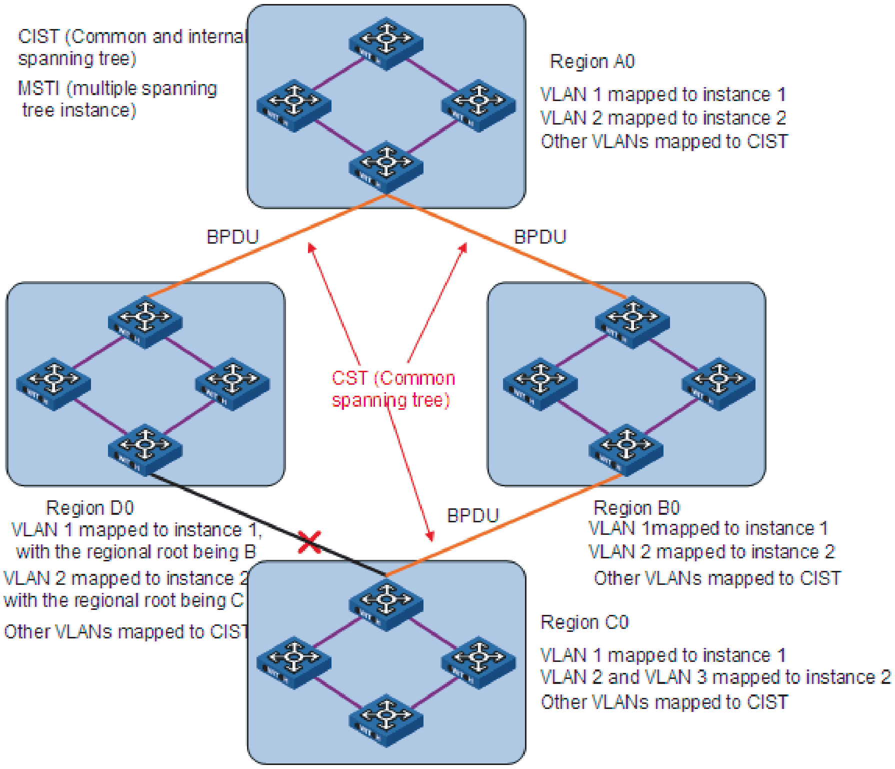
MSTP features the following:

- MSTP supports mapping VLANs to MST instances by means of a VLAN-to-instance mapping table;
- 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 network with tree topology. As a network of this type is loop-free, it prevents packets in it from being duplicated and forwarded endlessly. In addition, MSTP can provide multiple redundant paths for data forwarding , thus allowing for load balancing in VLANs;
- MSTP is compatible with STP and RSTP.

Some concepts in MSTP

As shown in Figure 28, there are four multiple spanning tree (MST) regions, each made up of four switches running MSTP. In light with the diagram, the following paragraphs will present some concepts of MSTP.

Figure 28 Basic concepts in MSTP


CIST (Common and internal spanning tree)
MSTI (multiple spanning tree instance)
Region A0
VLAN 1 mapped to instance 1
VLAN 2 mapped to instance 2
Other VLANs mapped to CIST
BPDU
BPDU
CST (Common spanning tree)
Region D0
VLAN 1 mapped to instance 1, with the regional root being B
VLAN 2 mapped to instance 2 with the regional root being C
Other VLANs mapped to CIST
BPDU
Region B0
VLAN 1mapped to instance 1
VLAN 2 mapped to instance 2
Other VLANs mapped to CIST
Region C0
VLAN 1 mapped to instance 1
VLAN 2 and VLAN 3 mapped to instance 2
Other VLANs mapped to CIST

1 MST region

An MST region is composed of multiple devices in a switched network and 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-instance mapping configuration,
- They have the same MSTP revision level configuration, and
- They are physically linked with one another.

In area A0 in Figure 28, for example, all the device have the same MST region configuration: the same region name, the same VLAN-to-instance mapping (VLAN 1 is mapped to MST instance 1, VLAN 2 to MST instance 2, and the rest to the common and internal spanning tree (CIST).), 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 group multiple devices to the same MST region.

2 VLAN-to-instance mapping table

As an attribute of an MST region, the VLAN-to-instance mapping table describes the mapping relationships between VLANs and MST instances. In Figure 28, for example, the VLAN-to-instance mapping table of region A0 describes that the same region name, the same VLAN-to-instance mapping (VLAN 1 is mapped to MST instance 1, VLAN 2 to MST instance 2, and the rest to CIST.

3 IST

Internal spanning tree (IST) is a spanning tree that runs in an MST region, with the instance number of 0. 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 an MST region. In Figure 28, for example, the CIST has a section in each MST region, and this section is the IST in each 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 computed by these devices through MSTP. For example, the red lines in Figure 28 describe 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 28, 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 28, for example, multiple spanning tree can exist in each MST region, each spanning tree corresponding to a VLAN. 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 MST or that MSTI. Based on the topology, different spanning trees in an MST region may have different regional roots. For example, in region D0 in Figure 28, the regional root of instance 1 is device B, while that of instance 2 is device C.

8 Common root bridge

The root bridge of the CIST is the common root bridge. In Figure 28, 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 configuration, or to a single spanning-tree region running STP, or to a single spanning-tree region running RSTP.

During MSTP computing, a boundary port assumes the same role on the CIST and on MST instances. Namely, if a boundary port is master port on the CIST, it is also the master port on all MST instances within this region. In Figure 28, 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.

Currently, the Switch 8800s are not capable of recognizing boundary ports. When a Switch 8800 is connected to a third party's device that supports boundary port recognition, the third party's device may malfunction in recognizing a boundary port.

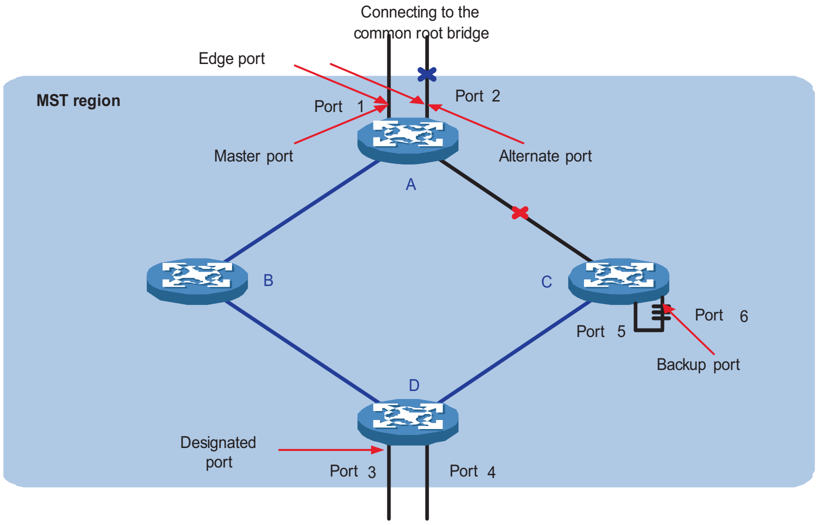
10 Roles of ports

In the MSTP computing process, port roles include designated port, root 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 entire region to the common root bridge, connect the MST region to the common root bridge.
- Alternate port: The standby port for a master port. If a master port is blocked, the alternate port becomes the new master port.
- Backup port: If a loop occurs when two ports of the same device are interconnected, the device will block either of the two ports, and the backup port is that port to be blocked.

A port can assume different roles in different MST instances.

Figure 29 Port roles



In Figure 29,

- 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 tree:

- Forwarding: the port learns MAC addresses and forwards user traffic;
- Learning: the port learns MAC addresses but does not forwards user traffic;
- Discarding: the port neither learns MAC addresses nor forwards user traffic.

When in different MST instances, a port can be in different states.

A port state is not exclusively associated with a port role. Table 13 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 13 Ports states supported by different port roles

Role\State	**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 computed CST. Inside an MST region, multiple spanning trees are generated through computing, each spanning tree called an MST instance. Among these MST instances, instance 0 is the IST, while all the others are MSTIs. Similar to STP, MSTP uses configuration BPDUs to compute spanning trees. The only difference between the two protocols being in that what is carried in an MSTP BPDU is the MSTP configuration on the device from which this BPDU is sent.

1 CIST computing

By comparison of configuration BPDUs, one device with the highest priority is elected as the root bridge of the CIST. MSTP generates an IST within each MST region through computing, and, at the same time, MSTP regards each MST region as a single device and generates a CST among these MST regions through computing. The CST and ISTs constitute the CIST of the entire network.

2 MSTI computing

Within an MST region, MSTP generates different MSTIs for different VLANs based on the VLAN-to-instance mappings.

MSTP performs a separate computing process, which is similar to spanning tree computing in STP, for each spanning tree. For details, refer to "How STP works" on page 95.

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 computing.

In addition to basic MSTP functions, many management-facilitating special functions are provided, as follows:

- Root bridge hold
- Root bridge backup
- Root guard
- BPDU guard
- Loop guard
- Support for hot swapping of interface cards and active/standby changeover.

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 configuration, you need to know the position of each device in each MST instance: root bridge or leave node. In each instance, one, and only one device acts as the root bridge, while all others as leaf nodes.

Task		Remarks
"Configuring the Root Bridge" on page 109	"Configuring an MST Region" on page 109	Required
	"Specifying the Root Bridge or a Secondary Root Bridge" on page 110	Optional
	"Configuring the Work Mode of MSTP Device" on page 112	Optional
	"Configuring the Priority of the Current Device" on page 113	Optional
	"Configuring the Maximum Hops of an MST Region" on page 113	Optional
	"Configuring the Network Diameter of a Switched Network" on page 114	Optional
	"Configuring Timers of MSTP" on page 114	Optional
	"Configuring the Timeout Factor" on page 116	Optional
	"Configuring the Maximum Transmission Rate of Ports" on page 116	Optional
	"Configuring Ports as Edge Ports" on page 117	Optional
	"Configuring Whether Ports Connect to Point-to-Point Links" on page 118	Optional
	"Configuring the MSTP Packet Format for Ports" on page 118	Optional
	"Enabling the MSTP Feature" on page 119	Required

Task		Remarks
"Configuring Leaf Nodes" on page 120	"Configuring an MST Region" on page 109	Required
	"Configuring the Work Mode of MSTP Device" on page 112	Optional
	"Configuring the Timeout Factor" on page 116	Optional
	"Configuring the Maximum Transmission Rate of Ports" on page 116	Optional
	"Configuring Ports as Edge Ports" on page 117	Optional
	"Configuring Path Costs of Ports" on page 120	Optional
	"Configuring Port Priority" on page 123	Optional
	"Configuring Whether Ports Connect to Point-to-Point Links" on page 118	Optional
	"Configuring the MSTP Packet Format for Ports" on page 118	Optional
	"Enabling the MSTP Feature" on page 119	Required
"Performing mCheck" on page 124		Optional
"Configuring the VLAN Ignore Feature" on page 125		Optional
"Configuring Digest Snooping" on page 126		Optional
"Configuring No Agreement Check" on page 128		Optional
"Configuring Protection Functions" on page 130		Optional

If both GVRP and MSTP are enabled on a device at the same time, GVRP packets will be forwarded along the CIST. Therefore, if both GVRP and MSTP are running on the same device and you wish to advertise a certain VLAN within the network through GVRP, make sure that this VLAN is mapped to the CIST (instance 0) when configuring the VLAN-to-instance mapping table. For detailed information of GVRP, refer to "GVRP Configuration" on page 139.

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 By default, the name of an MST region is the bridge MAC address.

To do...	Use the command...	Remarks
Configure the VLAN-to-instance mapping table	**instance** *instance-id* **vlan** *vlan-list* **vlan-mapping modulo** *modulo*	Optional By default, all VLANs in an MST region are mapped to MST instance 0.
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 the configuration of the current 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

Two device belong to the same MST region only if they are configure to have the same MST region name, the same VLAN-to-instance 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-instance mapping table, will cause MSTP to launch a new spanning tree computing 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 computing process when processing MST region-related configurations; instead, such configurations will take effect only if you:

- activate the MST region-related parameters suing 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 instance 1, and VLAN 20 through VLAN 30 to instance 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 computing. 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** [**bridge-diameter** *bridgenum*] [**hello-time** *centi-seconds*]	Required The device does not function as the root bridge by default

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** [**bridge-diameter** *bridgenum*] [**hello-time** *centi-seconds*]	Required By default, a device does not function as a secondary root bridge.

Note that:

- Upon 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 MST instance, 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 instances. 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 instance. However, the same device cannot be the root bridge and a secondary root bridge in the same instance 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 device.
- 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 instance. However, if you specify a new 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.

- When specifying the root bridge or a secondary root bridge, you can specify the network diameter and hello time. However, these two options are effective only for MST instance 0, namely the CIST. If you include these two options in your command for any other instance, the configuration can succeed, but they will not actually work. For the description of network diameter and hello time, refer to "Configuring the Network Diameter of a Switched Network" on page 114 and "Configuring Timers of MSTP" on page 114.
- Alternatively, you can also specify the current device as the root bridge by setting by priority of the device to 0. For the device priority configuration, refer to "Configuring the Priority of the Current Device" on page 113.

Configuration example

# Specify the current device as the root bridge of MST instance 1 and a secondary root bridge of MST instance 2.

```
<Sysname> system-view
[Sysname] stp instance 1 root primary
[Sysname] stp instance 2 root secondary
```

Configuring the Work Mode of 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 full interoperability with RSTP-compliant 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
```

# Configure MSTP to work in RSTP mode.

```
<Sysname> system-view
[Sysname] stp mode rstp
```

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 spanning tree. An MSTP-compliant device can have different priorities in different MST instances.

Configuration procedure

Follow these steps to configure the priority of the current device:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the priority of the current device	**stp** [**instance** *instance-id*] **priority** *priority*	Optional 32768 by default

CAUTION:

- *Upon specifying the current device as the root bridge or a secondary root bridge, you cannot configure 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 MST instance 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 setting configured on the regional root bridge will be used as the maximum hops of the MST region.

After a configuration BPDU leaves the root bridge of the spanning tree in the region, its hop count is decreased by 1 whenever it passes a device. When its hop count reaches 0, it will be discarded by the device that has received it. As a result, devices beyond the maximum hops are unable to take part in spanning tree computing, and thereby the size of the MST region is restricted.

Configuration procedure

Follow these steps to configure the maximum 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

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 specific paths, which are composed of a series of devices. Represented by the number of devices on a path, the network diameter is the path that comprises more devices than any other among these paths.

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** *bridgenum*	Optional 7 by default

- *Network diameter is a parameter that indicates 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.*

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.

- Forward delay: the time a device will wait before changing states. A link failure can trigger a spanning tree computing process, and the spanning tree structure will change accordingly. However, as a new configuration BPDU cannot be propagated throughout the network immediately, if the new root port and designated port begin to forward data as soon as they are elected, a temporary loop may occur. For this reason, the protocol uses a state transition mechanism. Namely, a newly elected root port or designated port must wait twice the forward delay time before transitioning to the forwarding state, when the new configuration BPDU has been propagated throughout the network.
- Hello time is sued to detect whether a link is faulty. A device sends a hello packet to the devices around it at a regular interval of hello time to check whether any link is faulty.
- Max time is used for determining whether a configuration BPDU has "expired". A BPDU that has expired will be discarded by the device.

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 time timer	**stp timer hello** *centi-seconds*	Optional 200 centiseconds (2 seconds) by default
Configuring 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 resume connectivity. 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 on a link for link failure and trigger a new spanning tree computing 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 computing and may take network congestion to 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 computing, thus reducing the auto-sensing capability of the network. We recommend that you use the default setting.*

The setting of hello time, forward delay and max age must meet the following formulae; otherwise network instability will frequently occur.

- 2 × (forward delay - 1 second) ƒ max age
- Ma x age ƒ 2 × (hello time + 1 second)

We recommend that you specify the network diameter and the hello time using the **stp root primary** command by preference and let MSTP automatically calculate an optimal setting of the other two 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

A device sends hello packets to the devices around it at a specific interval to check whether any link is faulty. Typically, if a device does not receive a BPDU from the upstream device within a period three times the hello time, it will assume that the upstream device has failed and start a new spanning tree computing process.

In a very stable network, this kind of spanning tree computing may occur because the upstream device is busy. In this case, you can avoid such unwanted spanning tree computing 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

- *Timeout time = timeout factor × hello time.*
- *Typically, we recommend that you set the timeout factor to 5, 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 Transmission Rate of Ports

The maximum transmission rate of a port refers to the maximum number of MSTP packets that the port can send within each hello time.

The maximum transmission rate of an Ethernet port is related to the physical status of the port and the network structure.

Configuration procedure

Following these steps to configure the maximum transmission rate of a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...		Use the command...	Remarks
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the maximum transmission rate of the port(s)		**stp transmit-limit** *packet-number*	Optional 10 by default

If the maximum transmission 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 Ethernet 1/1/1 to 5.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] 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. Therefore, if you specify a port as an edge port, this port can transition rapidly from the blocked state to the forwarding state without delay.

Configuration procedure

Following 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 Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the port(s) as edge port(s)		**stp edged-port enable**	Required All Ethernet ports are non-edge ports by default

- *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.*
- *If a port directly connects to a user terminal, configure it to be an edge port and enable BPDU guard for it. This enables the port to transition to the forwarding state while ensuring network security.*

Configuration example

# Configure Ethernet 1/1/1 to be an edge port.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp edged-port enable
```

Configuring Whether Ports Connect to Point-to-Point Links

A point-to-point link is a link directly connecting with two devices. If the roles of two ports directly connected by a point-to-point link meet specific requirements, the ports can rapidly transition to the forwarding state after a proposal-agreement handshake process.

Configuration procedure

Following these steps to configure whether a port or a group of ports connect to point-to-point links:

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*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure whether the port(s) connect to point-to-point links		**stp point-to-point** { **force-true** | **force-false** | **auto** }	Optional The default setting is **auto**; namely the device automatically detects whether an Ethernet port connects to a point-to-point link

- *As for aggregated ports, all ports can be configured as connecting to point-to-point links. 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 MST instances. 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 Ethernet 1/1/1 as connecting to a point-to-point link.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp point-to-point force-true
```

Configuring the MSTP Packet Format for Ports

A port support two types of MSTP packets:

- 802.1s-compliant standard format
- Compatible format

The default packet format setting is **auto**, namely a port recognizes the two MSTP packet formats automatically. You can configure the MSTP packet format to be used by a port as 802.1s-compliant standard format or compatible format using corresponding commands. After the configuration, when working in MSTP mode, the port sends and receives only MSTP packets of the format you have configured.

Configuration procedure

Follow these steps to configure the MSTP packet format for a port or a group of ports:

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*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the MSTP packet format for the port(s)		**stp compliance** { **auto** | **dot1s** | **legacy** }	Optional **auto** by default

- *If the port is configured not to detect the packet format automatically while it works in the MSTP mode, and if it receives a packet in the format other than as configured, that port will become a designated port, and the port will remain in the discarding state to prevent the occurrence of a loop.*
- *If a port receives MSTP packets of different formats frequently, this means that the MSTP packet formation configuration contains error. In this case, if the port is working in MSTP mode, it will be shut down for protection.*

Configuration example

# Configure port Ethernet 1/1/1 to receive and send standard-format MSTP packets.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp compliance dot1s
```

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**	Optional By default, the MSTP feature is enabled globally.

To do...		Use the command...	Remarks
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the MSTP feature for the port(s)		**stp enable**	Optional By default, MSTP is enabled for all ports after it is enabled for the device globally

- *You must enable MSTP for the device before any other MSTP-related configuration can take effect.*
- *To control MSTP flexibly, you can use the **stp disable** or **undo stp** command to disable the MSTP feature for specific ports so that they will not take part in spanning tree computing and thus to save the device's CPU resources.*

Configuration example

# Enable MSTP for the device and disable MSTP for port Ethernet 1/1/1.

```
<Sysname> system-view
[Sysname] stp enable
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp disable
```

Configuring Leaf Nodes

Perform the following configurations for a device operating as a leaf node.

Configuring an MST Region

Refer to "Configuring an MST Region" on page 109.

Configuring the Work Mode of MSTP

Refer to "Configuring the Work Mode of MSTP Device" on page 112.

Configuring the Timeout Factor

Refer to "Configuring the Timeout Factor" on page 116.

Configuring the Maximum Transmission Rate of Ports

Refer to "Configuring the Maximum Transmission Rate of Ports" on page 116.

Configuring Ports as Edge Ports

Refer to "Configuring Ports as Edge Ports" on page 117.

Configuring Path Costs of Ports

Path cost is a parameter related to the rate of port-connected links. On an MSTP-compliant device, ports can have different priorities in different MST instances. Setting an appropriate path cost allows VLAN traffic flows to be forwarded along different physical links, thus to enable per-VLAN load balancing.

The device can automatically calculate the path cost; alternatively, you can also configure the path cost for ports.

Specifying a standard that the device uses when calculating the path cost

You can specify a standard for the device to use in automatic calculation for the 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 cost of the link connected with the device	**stp pathcost-standard** { **dot1d-1998** | **dot1t** | **legacy** }	Optional The device uses the private standard by default.

Table 14 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
	Aggregated Link 2 Ports	100	1,000,000	1,800
	Aggregated Link 3 Ports	100	666,666	1,600
	Aggregated Link 4 Ports	100	500,000	1,400
100 Mbps	Single Port	19	200,000	200
	Aggregated Link 2 Ports	19	100,000	180
	Aggregated Link 3 Ports	19	66,666	160
	Aggregated Link 4 Ports	19	50,000	140

Table 14 Link speed vs. path cost

Link speed	Duplex state	802.1D-1998	802.1t	Private standard
1000 Mbps	Single Port	4	20,000	20
	Aggregated Link 2 Ports	4	10,000	18
	Aggregated Link 3 Ports	4	6,666	16
	Aggregated Link 4 Ports	4	5,000	14
10 Gbps	Single Port	2	2,000	2
	Aggregated Link 2 Ports	2	1,000	1
	Aggregated Link 3 Ports	2	666	1
	Aggregated Link 4 Ports	2	500	1

In the calculation of the path cost value of an aggregated link, 802.1D-1998 does not take into account the number of ports in the aggregated link. Whereas, 802.1T takes the number of ports in the aggregated link into account. The calculation formula 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 aggregated link.

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 Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
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

- *When the path cost of a port is changed, MSTP will re-compute 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.*
- *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 out of effect.*

Configuration example I

# Configure the path cost of Ethernet 1/1/1 in MST instance 1 to 2000.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp instance 1 cost 2000
```

Configuration example II

# Configure MSTP to automatically calculate the path cost of Ethernet 1/1/1 based on the IEEE 802.1D-1998 standard.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] undo stp instance 1 cost
[Sysname-Ethernet1/1/1] quit
[Sysname] stp pathcost-standard dot1d-1998
```

Configuring Port Priority

The priority of a port is an import basis that determines whether the port can be elected as the root port of device. If all other conditions are the same, the port with the highest priority will be elected as the root port.

On an MSTP-compliant device, a port can have different priorities in different MST instances, and the same port can play different roles in different MST instances, 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 Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure port priority		**stp** [**instance** *instance-id*] **port priority** *priority*	Optional 128 for all Ethernet ports by default

- *When the priority of a port is changed, MSTP will re-compute the role of the port and initiate a state transition.*
- *Generally, a lower configured value priority indicates a higher priority of the port. If you configure the same priority value for all the Ethernet ports on the a device, the specific priority of a port depends on the index number of that port. Changing the priority of an Ethernet port triggers a new spanning tree computing process.*

Configuration example

# Set the priority of port Ethernet 1/1/1 to 16 in MST instance 1.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp instance 1 port priority 16
```

Configuring Whether Ports Connect to Point-to-Point Links

Refer to "Configuring Whether Ports Connect to Point-to-Point Links" on page 118..

Configuring the MSTP Packet Format for Ports

Refer to "Configuring the MSTP Packet Format for Ports" on page 118.

Enabling the MSTP Feature

Refer to "Enabling the MSTP Feature" on page 119.

Performing mCheck

Ports on an MSTP-compliant 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, this 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.

Perform global mCheck

Follow these steps to perform global mCheck:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Perform mCheck	**stp mcheck**	Required

Perform mCheck in Ethernet interface view

Follow these steps to perform mCheck in Ethernet interface view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Perform mCheck	**stp mcheck**	Required

__CAUTION:__ The __stp mcheck__ command is meaningful only when the device works in the MSTP (or RSTP) mode, not in the STP-compatible mode.

Configuration example

# Perform mCheck on port Ethernet 1/1/1.

1 Perform mCheck globally

```
<Sysname> system-view
[Sysname] stp mcheck
```

2 Perform mCheck in Ethernet interface view

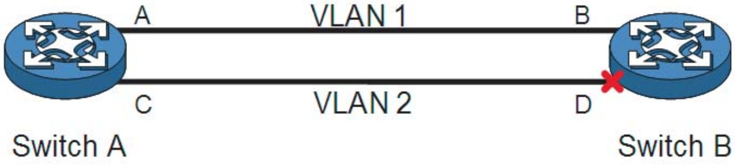
```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp mcheck
```

Configuring the VLAN Ignore Feature

Introduction to the VLAN Ignore Feature

Traffic on a VLAN in a complex network may be blocked by spanning tree.

Figure 30 VLAN connectivity blocked by MSTP



As shown in Figure 30, port A on Switch A allows VLAN 1 to pass, C allows VLAN 2 to pass; port B on Switch B allows VLAN 1 to pass, port D allows VLAN 2 to pass. Switch A and Switch B run MSTP. Switch A is the root bridge, and port A and port C on it are designated ports. Port B on Switch B is the root port, and port D is a blocked port. In this case, traffic on VLAN 2 is blocked.

Enabling the VLAN Ignore feature for a VLAN can make ports of the VLAN forward packets normally rather than comply with the calculated result of MSTP.

Configuration Procedure

Follow these steps to configure VLAN Ignore:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable VLAN Ignore for a VLAN	**stp ignored vlan** *vlan-list*	Required By default, VLAN Ignore is disabled in a VLAN.
Display VLAN Ignore enabled VLANs	**display stp ignored-vlan**	Available in any view

Configuration Examples

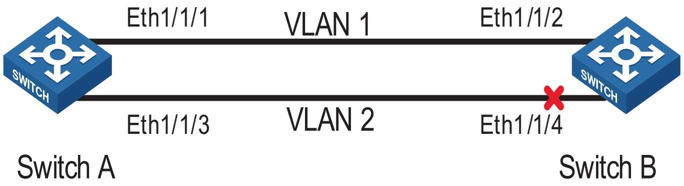
Network requirements

- Switch A and B are in direct connection;

- Ethernet 1/1/1 on Switch A and Ethernet 1/1/2 on Switch B allow VLAN 1 to pass. Ethernet 1/1/3 on Switch A and Ethernet 1/1/4 on Switch B allow VLAN 2 to pass.
- Switch A is the root bridge, and both Switch A and Switch B run MSTP. Ethernet 1/1/4 on Switch B is blocked, causing traffic block on VLAN 2.
- Configure VLAN Ignore to keep the ports in VLAN 2 on Switch B in the forwarding state.

Network diagram

Figure 31 VLAN Ignore configuration



Configuration procedure

1 Enable VLAN Ignore on Switch B.

# Enable VLAN Ignore on VLAN 2.

```
<SysnameB> system-view
[SysnameB] stp ignored vlan 2
```

2 Verify the configuration

# Display the VLAN Ignore-enabled VLANs.

```
[SysnameB] display stp ignored-vlan
STP-Ignored VLAN: 2
```

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-instance 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 computed via the HMAC-MD5 algorithm based on VLAN-to-instance mappings.

In practical networking implementations, since MSTP implementations differ with vendors, the configuration digest computed 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 associated port can make a device communicate with another vendor's device in the same MST region.

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 Ethernet interface or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Choose either
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
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 only enable the Digest Snooping feature on the device connected to another vendor's device that use private key to compute 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-instance mappings must be the same on associated ports.*
- *With global Digest Snooping enabled, modification of VLAN-to-instance 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 the MST region edge port to avoid loops.*
- *Do not enable Digest Snooping when the network works well to avoid traffic interruption.*

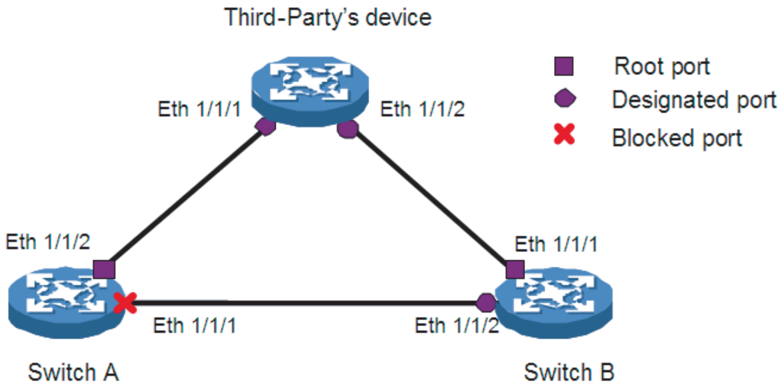
Configuration Examples

Network requirements

- Switch A and Switch B connect to a third-party's device and all the devices are in the same region.
- Enable Digest Snooping on Switch A and Switch B so that the three devices can communicate with one another.

Network diagram

Figure 32 Digest Snooping configuration



Configuration procedure

1 Enable Digest Snooping on Switch A

# Enable Digest Snooping on Ethernet 1/1/2.

```
<SysnameA> system-view
[SysnameA] interface ethernet 1/1/2
[SysnameA-Ethernet1/1/2] stp config-digest-snooping
```

# Enable global Digest Snooping.

```
[SysnameA-Ethernet1/1/2] quit
[SysnameA] stp config-digest-snooping
```

2 Enable Digest Snooping on Switch B (the same as the configuration procedure of Switch A, omitted)

Configuring No Agreement Check

Two types of packet are used for rapid state transition on designated RSTP and MSTP ports:

- Proposal: Packets sent by designated ports to request rapid transition
- Agreement: Packets used to acknowledge rapid transition requests

Both RSTP and MSTP switches can perform rapid transition operation on a designated port only when the port receives an agreement packet from the downstream switch. The differences between RSTP and MSTP switches 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 33 and Figure 34 show the rapid state transition mechanism on MSTP and RSTP designated ports.

Figure 33 Rapid state transition mechanism on the MSTP designated port

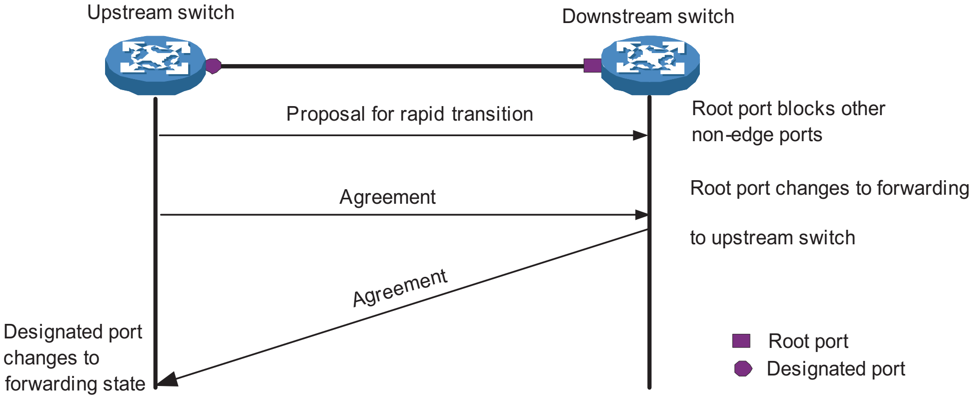
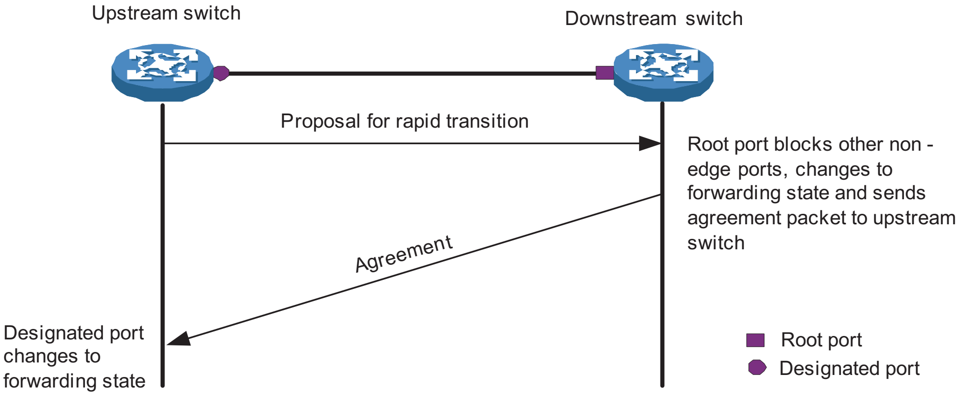

Upstream switch
Downstream switch
Proposal for rapid transition
Root port blocks other non-edge ports
Agreement
Root port changes to forwarding
to upstream switch
Agreement
Designated port changes to forwarding state
Root port
Designated port

Figure 34 Rapid state transition mechanism on the RSTP designated port


Upstream switch
Downstream switch
Proposal for rapid transition
Root port blocks other non - edge ports, changes to forwarding state and sends agreement packet to upstream switch
Agreement
Designated port changes to forwarding state
Root port
Designated port

If the upstream device comes from another vendor, the rapid state transition implementation may be limited. For example, when the upstream device adopts RSTP, the downstream device adopts MSTP and does not support 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 switch 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 perform rapid state transition.

Configuration Prerequisites

- A device is the upstream one that is connected to another vendor's MSTP supported device via a point-to-point link.
- Configure the same region name, revision level and VLAN-to-instance mappings on the two devices, making them in the same region.

Configuration Procedure

Following these steps to configure No Agreement Check:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...		Use the command...	Remarks
Enter Ethernet interface or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Choose either
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable No Agreement Check		**stp no-agreement-check**	Required Not enabled by default

The No Agreement Check feature can take effect only when it is enabled on the root port.

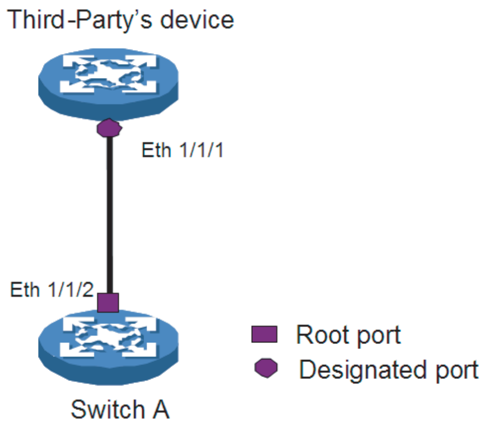
Configuration Examples

Network requirements

- Switch A connects to a third-party's device that has different MSTP implementation. Both switches are in the same region.
- Another vendor's device is the regional root bridge, and Switch A is the downstream device.

Network diagram

Figure 35 No Agreement Check configuration



Configuration procedure

# Enable No Agreement Check on Ethernet 1/1/2 of Switch A.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] stp no-agreement-check
```

Configuring Protection Functions

An MSTP-compliant device supports the following protection functions:

- BPDU guard
- Root guard
- Loop guard
- TC-BPDU attack guard

Among loop guard, root guard and edge port setting, only one function can take effect on the same port at the same time.

These protection functions function as follows:

- BPDU guard

For access layer devices, the access ports generally have user terminals (such as PCs) or file servers directly connected to them. These ports are usually configured as edge ports to allow rapid transition. However, these ports become non-edge ports when they receive configuration BPDUs, which triggers a new round of spanning tree computing process and causes changes of network topology. Under normal conditions, these ports are not supposed to receive configuration BPDUs. However, if someone forges configuration BPDUs maliciously to attack the devices, network may become instable.

MSTP provides the BPDU guard function to protect the system against such attacks. With the BPDU guard function enabled on the devices, edge ports receiving configuration BPDUs are shut down and the NMS is informed. Those ports closed thereby can be restored only by the network administrators.

- Root guard

The root bridge and its secondary root bridges of a spanning tree must reside in the same MST region. Especially for the CIST, the root bridge and its secondary root bridges are generally put in a high-bandwidth core region during network design. However, due to possible configuration errors or attacks in the network, the 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 undesired change of the network topology. As a result of this kind of illegal topology change, the traffics that are to travel along high-speed links may be led 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. Ports with root guard function enabled can only be designated ports in all MST instances. Once a port of this type receives a configuration BPDU with a higher priority from an MST instance, it turns to the listening state in the MST instance and stops forwarding packets (as if it is disconnected from the link). If the port receives no BPDUs with higher priorities within twice the forwarding delay, the port reverts to its original state.

- Loop guard

  A device maintains the states of its root port and blocked ports by receiving and processing BPDUs from the upstream device. However, due to link congestion or unidirectional link failures, these ports may fail to receive BPDUs from the upstream device. In this case, the downstream device will reselect the port roles (for example, ports failing to receive upstream BPDUs become designated ports and the blocked ports transition to the forwarding state), resulting in loops in the switched network. The loop guard function can suppress the occurrence of such loops.

A loop guard-enabled port that fails to receive BPDUs from the upstream device remains in the discarding state in all the MST instances in the process of STP computing, regardless of the role it plays.

- TC-BPDU attack guard

A device removes the corresponding forwarding entries upon receiving a TC-BPDU (a PDU notifying of a topology change). If a malicious user forges large amount of TC-BPDUs and sends them to a device in a short period, the device may be busy removing the forwarding entries, decreasing the performance of the switch and introducing potential stability risks.

The TC-BPDU attack guard function can relieve a switch from this dilemma. With this function enabled, the device removes the forwarding address entries only once within a specific period (10 seconds) after it receives a TC-BPDU. At the same time, the system monitors whether other TC-BPDUs are received within that period. If so, the device will perform another removing operation after the period elapses. This prevents removing forwarding address entries frequently.

Configuration Prerequisites

MSTP has been correctly configured on the device.

Enabling the BPDU Guard Function

We recommend that you enable the BPDU guard function.

Configuration procedure

Following these steps to enable the BPDU guard function:

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

Configuration example

# Enable the BPDU guard function.

```
<Sysname> system-view
[Sysname] stp bpdu-protection
```

Enabling the Root Guard Function

We recommend that you enable the root guard function.

Configuration procedure

Follow these steps to enable the root guard function:

To do...	Use the command...	Remarks
Enter system view	**System-view**	-

To do...		Use the command...	Remarks
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all the ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the root guard function for the ports(s)		**stp root-protection**	Required Disabled by default

Configuration example

# Enable the root guard function for Ethernet 1/1/1.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp root-protection
```

Enabling the Loop Guard Function

We recommend that you enable the loop guard function.

Configuration procedure

Follow these steps to enable the loop guard function:

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*	User either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all the ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the loop guard function for the ports(s)		**stp loop-protection**	Required Disabled by default

Configuration example

# Enable the loop guard function for Ethernet 1/1/1.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] stp loop-protection
```

Enabling the TC-BPDU Attack Guard Function

We recommend that you keep this function enabled.

Configuration procedure

Follow these steps to enable the TC-BPDU attack guard function

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

Configuration example

# Enable the TC-BPDU attack guard function.

```
<Sysname> system-view
[Sysname] stp tc-protection enable
```

Displaying and Maintaining MSTP

To do...	Use the command...	Remarks
View the MSTP status information and statistics information	**display stp** [**instance** *instance-id*] [**interface** *interface-list* | **slot** *slot-num*] [**brief**]	Available in any view
View the MST region configuration information	**display stp region-configuration**	Available in any view
View the list of the VLANs with VLAN Ignore enabled	**display stp ignored-vlan**	Available in any view
Clear the MSTP statistics information	**reset stp** [**interface** *interface-list*]	Available in user view

MSTP Configuration Examples

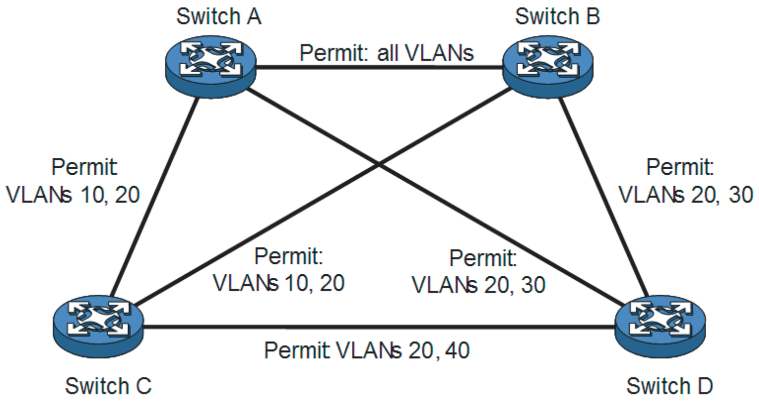
Network requirements

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 MST instance 1, those of VLAN 30 are forwarded along MST instance 3, those of VLAN 40 are forwarded along MST instance 4, and those of VLAN 20 are forwarded along MST instance 0.
- Switch A and Switch B are convergence layer devices, while Switch C and Switch D are access layer devices. VLAN 10 and VLAN 30 are terminated on the convergence layer devices, and VLAN 40 is terminated on the access layer devices, so the root bridges of MST instance 1 and MST instance 3 are Switch A and Switch B, while the root bridge of MST instance 4 is Switch C.

Network diagram

Figure 36 Network diagram for MSTP configuration



"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 Switch A

# Configure an MST region.

```
<SysnameA> system-view
[SysnameA] stp region-configuration
[SysnameA-mst-region] region-name example
[SysnameA-mst-region] instance 1 vlan 10
[SysnameA-mst-region] instance 3 vlan 30
[SysnameA-mst-region] instance 4 vlan 40
[SysnameA-mst-region] revision-level 0
```

# Activate MST region configuration manually.

```
[SysnameA-mst-region] active region-configuration
[SysnameA-mst-region] quit
```

# Configure Switch A as the root bridge of MST instance 1.

```
[SysnameA] stp instance 1 root primary
```

# View the MST region configuration information that has taken effect.

```
[SysnameA] 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 Switch B

# Configure an MST region.

```
<SysnameB> system-view
[SysnameB] stp region-configuration
[SysnameB-mst-region] region-name example
[SysnameB-mst-region] instance 1 vlan 10
[SysnameB-mst-region] instance 3 vlan 30
[SysnameB-mst-region] instance 4 vlan 40
[SysnameB-mst-region] revision-level 0
```

# Activate MST region configuration manually.

```
[SysnameB-mst-region] active region-configuration
[SysnameB-mst-region] quit
```

# Configure Switch B as the root bridge of MST instance 3.

```
[SysnameB] stp instance 3 root primary
```

# View the MST region configuration information that has taken effect.

```
[SysnameB] 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 Switch C

# Configure an MST region.

```
<SysnameC> system-view
[SysnameC] stp region-configuration
[SysnameC-mst-region] region-name example
[SysnameC-mst-region] instance 1 vlan 10
[SysnameC-mst-region] instance 3 vlan 30
[SysnameC-mst-region] instance 4 vlan 40
[SysnameC-mst-region] revision-level 0
```

# Activate MST region configuration manually.

```
[SysnameC-mst-region] active region-configuration
[SysnameC-mst-region] quit
```

# Configure Switch C as the root bridge of MST instance 4.

```
[SysnameC] stp instance 4 root primary
```

# View the MST region configuration information that has taken effect.

```
[SysnameC] 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 Switch D

# Configure an MST region.

```
<SysnameD> system-view
[SysnameD] stp region-configuration
[SysnameD-mst-region] region-name example
[SysnameD-mst-region] instance 1 vlan 10
[SysnameD-mst-region] instance 3 vlan 30
[SysnameD-mst-region] instance 4 vlan 40
[SysnameD-mst-region] revision-level 0
```

# Activate MST region configuration manually.

```
[SysnameD-mst-region] active region-configuration
[SysnameD-mst-region] quit
```

# View the MST region configuration information that has taken effect.

```
[SysnameD] 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
```

12 GVRP CONFIGURATION

GARP VLAN registration protocol (GVRP) is a GARP application. Based on the operating mechanism of GARP, GVRP maintains and propagates dynamic VLAN registration information for the GVRP devices on a network.

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

- “Introduction to GVRP” on page 139
- “Configuring GVRP” on page 142
- “Displaying and Maintaining GVRP” on page 143
- “GVRP Configuration Example” on page 144

Introduction to GVRP

This section covers these topics:

- “GARP” on page 139
- “Configuring GVRP” on page 142
- “Protocols and Standards” on page 142

GARP

The generic attribute registration protocol (GARP) provides a mechanism that allows GARP participants in a LAN to distribute, propagate, and register with other participants some attributes such as VLAN IDs or multicast addresses.

GARP itself does not exist on a device as an entity. GARP-compliant application entities are called GARP applications. One example is GVRP. When a GARP application entity is present on a port on your device, this port is regarded a GARP application entity.

This section covers these topics:

- “GARP messages and timers” on page 139
- “Operating mechanism of GARP” on page 140
- “GARP message format” on page 141

GARP messages and timers

1 GARP messages

A GARP participant exchanges information with other GARP participants mainly by sending the following three types of messages: Join, Leave, and LeaveAll.

- Join to register some attribute with other participants.

- Leave to deregister some attribute with other participants. Together with Join messages, Leave messages help GARP participants complete attribute reregistration and deregistration.
- LeaveAll to deregister all attributes. A LeaveAll message is sent upon expiration of the LeaveAll timer, which starts upon the startup of a GARP application entity.

Through message exchange, all attribute information to be registered propagates to all GARP participants throughout the LAN.

2 GARP timers

There are four GARP timers:

- Hold timer - When a GARP application entity receives the first registration request, it starts the Hold timer and collects succeeding requests. When the timer expires, the entity sends all these requests in one Join message, thus saving bandwidth.
- Join timer -- A GARP application entity sends each Join message twice for reliability sake and uses the Join timer to set the interval between the two sending operations.
- Leave timer -- Starts upon receipt of a Leave message from another GARP application entity for deregistering some attribute information. If no Join message is received before this timer expires, the GARP application entity removes the attribute information as requested.
- LeaveAll timer - Starts when a GARP application entity starts. When this timer expires, the entity sends a LeaveAll message so that other entities can re-register all its attribute information, and, at the same time, it restarts the LeaveAll timer.

- *The settings of GARP timers apply to all GARP applications, such as GVRP, on a LAN.*
- *Unlike other three timers, which are set on a port basis, the LeaveAll timer is set in system view and takes effect globally on all ports.*
- *Different devices on a network may have different LeaveAll timer values. Each time a device on the network receives a LeaveAll message, it resets its LeaveAll timer. Therefore, each GARP application entity will send LeaveAll messages based on the shortest LeaveAll timer in the network. As a result, only the shortest LeaveAll timer in the network will take effect.*

Operating mechanism of GARP

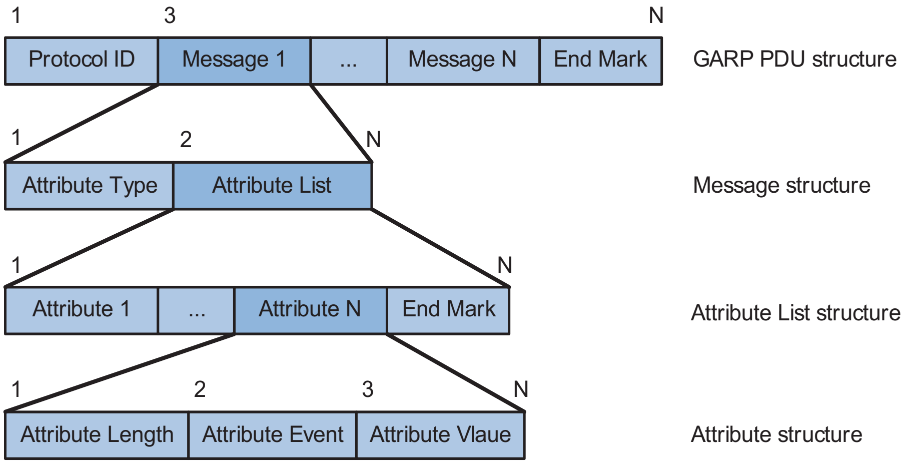
The GARP mechanism allows the configuration of a GARP participant to propagate throughout a LAN quickly. In GARP, a GARP participant registers or deregisters its attributes with other participants by making or withdrawing declarations of attributes and at the same time, based on received declarations or withdrawals handles attributes of other participants.

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

GARP message format

The following figure illustrates the format of GARP messages, which are carried in GARP PDUs.

Figure 37 Figure 11-1 GARP message format



The following table describes the GARP message fields.

Table 15 Table 11-1 Description on the GARP message fields

Field	Description	Value
Protocol ID	Protocol identifier for GARP	1
Message	Each message contains 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 1: JoinEmpty 2: JoinIn 3: LeaveEmpty 4: LeaveIn 5: Empty
Attribute Value	Attribute value	VLAN ID for GVRP If the Attribute Event is LeaveAll, Attribute Value is omitted.
End Mark	Indicates the end of PDU	--

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 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 from dynamically registering VLANs and from propagating information about dynamic VLANs, but allows the port to propagate information about static VLANs. On a trunk port with fixed registration type, GVRP can only propagate manually configured VLANs' information even though the port is configured to allow all VLANs to pass.
- Forbidden -- Disables the port from dynamically registering VLANs and from propagating any VLAN information except information about VLAN 1. On a trunk port with forbidden registration type, GVRP can only propagate the information of VLAN 1 (that is, the default VLAN) even though the port is configured to allow all VLANs to pass.

Protocols and Standards

IEEE 802.1Q specifies GVRP.

Configuring GVRP

GVRP configuration includes configuring GVRP functions and configuring GARP timers.

Configuring GVRP Functions

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

<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">Enable global GVRP</td><td>**gvrp**</td><td>Required
Disabled by default</td></tr>
<tr><td rowspan="2">Enter Ethernet interface view or port-group view</td><td>Enter Ethernet interface view</td><td>**interface** *interface-type interface-number*</td><td rowspan="2">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.</td></tr>
<tr><td>Enter port-group view</td><td>**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }</td></tr>
<tr><td colspan="2">Enable GVRP on the port</td><td>**gvrp**</td><td>Required
Disabled by default</td></tr>
<tr><td colspan="2">Configure the GVRP registration mode on the port</td><td>**gvrp registration** { **fixed** | **forbidden** | **normal** }</td><td>Optional
The default is **normal**.</td></tr>
</table>

- *Because GVRP is not compatible with the BPDU tunneling feature, you must disable BPDU tunneling before enabling GVRP on a BPDU tunneling-enabled Ethernet interface.*
- *Because global GVRP is not compatible with Isolate-user-VLAN, make sure that no Isolate-user-vlan has been created on the switch before enabling GVRP.*
- *You should enable GVRP globally before enabling it on a port.*
- *The port on which you want to enable GVRP must be a trunk port.*

Configuring GARP Timers

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.
Enter Ethernet interface view or port-group view	Enter Ethernet 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* | **aggregation** *agg-id* }	
Configure the Hold timer, Join timer, or Leave timer		**garp timer** { **hold** | **join** | **leave** } *timer-value*	Optional The default is 10 centiseconds for the Hold timer, 20 centiseconds for the Join timer, and 60 centiseconds for the Leave timer.

When configuring GARP timers, note that their values are dependent on each other and must be a multiplier of five centiseconds. If the value range for a timer is not desired, you may change it by tuning the value of another related timer as shown in the following table:

Table 16 Table 11-2 Dependencies of GARP timers

Timer	Lower limit	Upper limit
Hold	10 centiseconds	Not greater than half of the Join timer setting
Join	Not 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

To do...	Use the command...	Remarks
Display GARP timers for specified or all ports	**display garp timer** [**interface** *interface-list*]	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
Clear the GARP statistics	**reset garp statistics** [**interface** *interface-list*]	Available in user view

GVRP Configuration Example

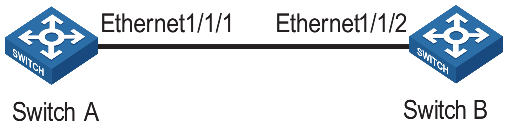
GVRP Configuration Example I

Network requirements

Configure GVRP for dynamic VLAN information registration and update between the switches. The trunk ports are in the default "normal" registration mode.

Network diagram

Figure 38 Figure 11-2 Network diagram for GVRP configuration



Configuration procedure

1 Configure Switch A

# Enable GVRP globally.

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

# Configure port Ethernet1/1/1 as a trunk port, allowing all VLANs to pass.

```
[SysnameA] interface ethernet 1/1/1
[SysnameA-Ethernet1/1/1] port link-type trunk
[SysnameA-Ethernet1/1/1] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/1, the trunk port.

```
[SysnameA-Ethernet1/1/1] gvrp
[SysnameA-Ethernet1/1/1] quit
```

# Create VLAN 2 (a static VLAN).

```
[SysnameA] vlan 2
[SysnameA-vlan2] return
```

2 Configure Switch B

# Enable GVRP globally.

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

# Configure port Ethernet1/1/2 as a trunk port, allowing all VLANs to pass.

```
[SysnameB] interface ethernet 1/1/2
[SysnameB-Ethernet1/1/2] port link-type trunk
[SysnameB-Ethernet1/1/2] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/2, the trunk port.

```
[SysnameB-Ethernet1/1/2] gvrp
[SysnameB-Ethernet1/1/2] quit
```

# Create VLAN 3 (a static VLAN).

```
[SysnameB] vlan 3
[SysnameB-vlan3] return
```

3 Verify the configuration

# Display dynamic VLAN information on Switch A.

```
<SysnameA> display vlan dynamic
Total 1 dynamic VLAN exist(s).
 The following dynamic VLANs exist:
  3
```

# Display dynamic VLAN information on Switch B.

```
<SysnameB> display vlan dynamic
Total 1 dynamic VLAN exist(s).
 The following dynamic VLANs exist:
  2
```

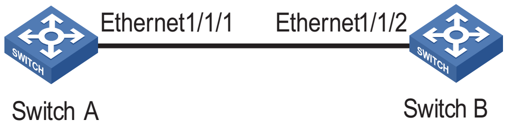
GVRP Configuration Example II

Network requirements

Configure GVRP for dynamic VLAN information registration and update between the switches. Set the "fixed" GVRP registration mode on the trunk port of Switch A and keep the default "normal" GVRP registration mode on the trunk port of Switch B.

Network diagram

Figure 39 Figure 11-3 Network diagram for GVRP configuration



Configuration procedure

1 Configure Switch A

# Enable GVRP globally.

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

# Configure port Ethernet1/1/1 as a trunk port, allowing all VLANs to pass.

```
[SysnameA] interface ethernet 1/1/1
[SysnameA-Ethernet1/1/1] port link-type trunk
[SysnameA-Ethernet1/1/1] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/1.

```
[SysnameA-Ethernet1/1/1] gvrp
```

# Set the GVRP registration type to fixed on the port.

```
[SysnameA-Ethernet1/1/1] gvrp registration fixed
[SysnameA-Ethernet1/1/1] quit
```

# Create VLAN 2 (a static VLAN).

```
[SysnameA] vlan 2
[SysnameA-vlan2] return
```

2 Configure Switch B

# Enable GVRP globally.

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

# Configure port Ethernet1/1/2 as a trunk port, allowing all VLANs to pass.

```
[SysnameB] interface ethernet 1/1/2
[SysnameB-Ethernet1/1/2] port link-type trunk
[SysnameB-Ethernet1/1/2] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/2.

```
[SysnameB-Ethernet1/1/2] gvrp
[SysnameB-Ethernet1/1/2] quit
```

# Create VLAN 3 (a static VLAN).

```
[SysnameB] vlan 3
[SysnameB-vlan3] return
```

3 Verify the configuration

# Display dynamic VLAN information on Switch A.

```
<SysnameA> display vlan dynamic
No dynamic VLAN exists!
```

# Display dynamic VLAN information on Switch B.

```
<SysnameB> display vlan dynamic
Total 1 dynamic VLAN exist(s).
```

```
 The following dynamic VLANs exist:
  2
```

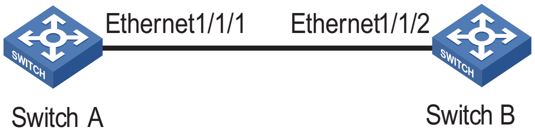
GVRP Configuration Example III

Network requirements

Configure GVRP for dynamic VLAN information registration and update between the switches. Set the "forbidden" GVRP registration mode on the trunk port of Switch A and keep the default "normal" mode on the trunk port of Switch B.

Network diagram

Figure 40 Figure 11-4 Network diagram for GVRP configuration



Configuration procedure

1 Configure Switch A

# Enable GVRP globally.

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

# Configure port Ethernet1/1/1 as a trunk port, allowing all VLANs to pass.

```
[SysnameA] interface ethernet 1/1/1
[SysnameA-Ethernet1/1/1] port link-type trunk
[SysnameA-Ethernet1/1/1] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/1.

```
[SysnameA-Ethernet1/1/1] gvrp
```

# Set the GVRP registration type to forbidden on the port.

```
[SysnameA-Ethernet1/1/1] gvrp registration forbidden
[SysnameA-Ethernet1/1/1] quit
```

# Create VLAN 2 (a static VLAN).

```
[SysnameA] vlan 2
[SysnameA-vlan2] return
```

2 Configure Switch B

# Enable GVRP globally.

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

# Configure port Ethernet1/1/2 as a trunk port, allowing all VLANs to pass.

```
[SysnameB] interface ethernet 1/1/2
[SysnameB-Ethernet1/1/2] port link-type trunk
[SysnameB-Ethernet1/1/2] port trunk permit vlan all
```

# Enable GVRP on Ethernet1/1/2.

```
[SysnameB-Ethernet1/1/2] gvrp
[SysnameB-Ethernet1/1/2] quit
```

# Create VLAN 3 (a static VLAN).

```
[SysnameB] vlan 3
[SysnameB-vlan3] return
```

3 Verify the configuration

# Display dynamic VLAN information on Switch A.

```
<SysnameA> display vlan dynamic
No dynamic VLAN exists!
```

# Display dynamic VLAN information on Switch B.

```
<SysnameB> display vlan dynamic
No dynamic VLAN exists!
```

13 BPDU TUNNELING CONFIGURATION

When configuring BPDU tunneling, refer to the following sections:

- "Introduction to BPDU Tunneling" on page 149
- "Configuring BPDU Isolation" on page 150
- "Configuring BPDU Transparent Transmission" on page 151
- "BPDU Tunneling Configuration Example" on page 152

Introduction to BPDU Tunneling

Why BPDU Tunneling

To avoid loops in your network, you can enable the spanning tree protocol (STP) on your device. However, STP gets aware of the topological structure of a network by means of bridge protocol data units (BPDUs) exchanged between different devices and the BPDUs are Layer 2 multicast packets, which can be received and processed by all STP-enabled devices on the network. This prevents each network from correctly calculating its spanning tree. As a result, when redundant links exist in a network, data loops will unavoidably occur.

By allowing each network has its own spanning tree while running STP, BPDU tunneling can resolve this problem. It has the following functions:

- It can isolate BPDUs of different customer networks, so that one network is not affected by others while calculating the topological structure.
- It enables BPDUs of the same customer network to be multicast over specific VLAN VPNs in the service provider network, so that the same, geographically dispersed customer network can implement consistent spanning tree calculation across the service provider network.

BPDU tunneling for the Switch 8800 Family only supports STP packets.

How BPDU Tunneling Works

The BPDU tunneling works implements the following two functions:

- BPDU isolation
- BPDU transparent transmission

The work process of IGMP is as follows:

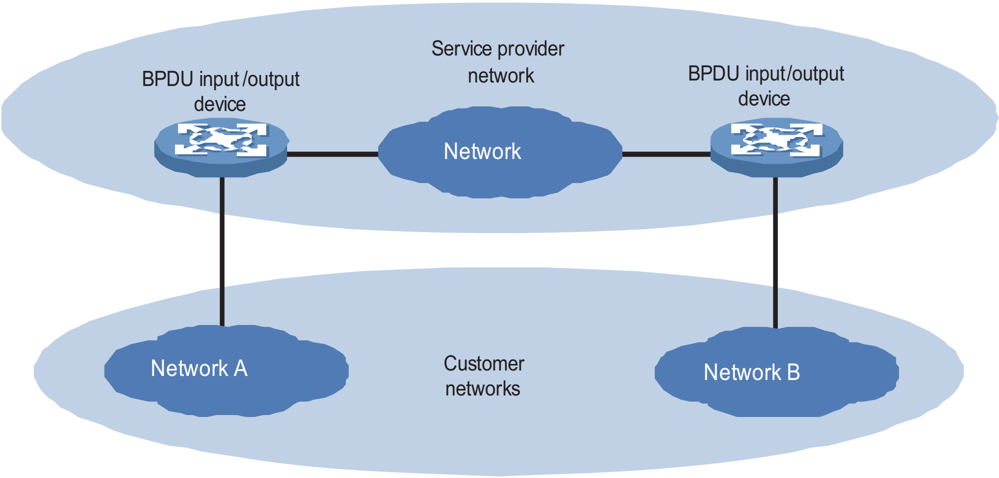
BPDU isolation

When a port receives BPDUs of other networks, the port will discard the BPDUs, so that they will not take part in spanning tree calculation. Refer to "Configuring BPDU Isolation" on page 150.

BPDU transparent transmission

As shown in Figure 41, 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 BPDU input/output 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 interfering each other. Refer to "Configuring BPDU Transparent Transmission" on page 151.

Figure 41 Network hierarchy of BPDU tunneling



The BPDU packet is processed in the operator network as follows:

- At the BPDU input side, the device changes the destination MAC address of a BPDU from a customer network from 0x0180-C200-0000 to a special multicast MAC address, 0x0100-0CCD-CDD0. In the service provider's network, the modified BPDUs are forwarded as data packets in the user VLAN.
- At the packet output side, the device recognizes the BPDU with the destination MAC address of 0x0100-0CCD-CDD0 and restores its original destination MAC address 0x0180-C200-0000. Then, the device removes the out-layer VLAN tag, and sends the BPDU to the destination customer network.

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 Isolation

Perform the following tasks to configure BPDU isolation:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable BPDU tunneling globally		**bpdu-tunnel dot1q enable**	Optional Enabled by default The configured BPDU tunneling on a port cannot take effect unless BPDU tunneling is enabled globally.
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable BPDU tunneling for the Ethernet port		**bpdu-tunnel dot1q enable**	Required Disabled by default

- *The BPDU tunneling feature is incompatible with the GVRP feature, so these two features cannot be enabled at the same time. For information about GVRP, refer to "GVRP Configuration" on page 139.*
- *The configured BPDU tunneling on a port cannot take effect unless BPDU tunneling is enabled globally..*

Configuring BPDU Transparent Transmission

Perform the following tasks to configure BPDU transparent transmission:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable BPDU tunneling globally		**bpdu-tunnel dot1q enable**	Optional Enabled by default
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable BPDU tunneling for the port		**bpdu-tunnel dot1q enable**	Required Disabled by default
Disable STP for the port		**stp disable**	Required
Enable STP BPDU tunneling for the port		**bpdu-tunnel dot1q stp**	Required

- *BPDU tunneling must be enabled globally before the BPDU TUNNEL configuration for a port can take effect.*

- *The BPDU tunneling feature is incompatible with the GVRP feature, so these two features cannot be enabled at the same time. For introduction to GVRP, refer to "GVRP Configuration" on page 139.*

BPDU Tunneling Configuration Example

Network requirements

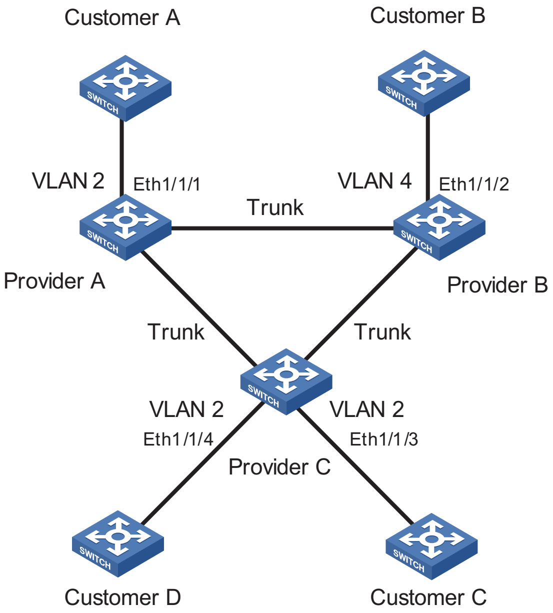
- Customer A, Customer B, Customer C, and Customer D are customer network access devices.
- Provider A, Provider B, and Provider C are service provider network access devices, which are interconnected through configured trunk ports.

The configuration is required to satisfy the following requirements:

- Geographically dispersed customer networks Customer A, Customer C and Customer D can implement consistent spanning tree calculation across the service provider network.
- BPDU packets are isolated for the customer network Customer B, so it does not take part in the spanning tree calculation.

Network diagram

Figure 42 Network diagram for BPDU tunneling configuration



Configuration procedure

1 Configuration on Provider A

# Configure BPDU transparent transmission on Ethernet 1/1/1.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port access vlan 2
[Sysname-Ethernet1/1/1] stp disable
[Sysname-Ethernet1/1/1] bpdu-tunnel dot1q enable
[Sysname-Ethernet1/1/1] bpdu-tunnel dot1q stp
```

2 Configuration on Provider B

# Configure BPDU isolation on Ethernet 1/1/2.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/2] port access vlan 4
[Sysname-Ethernet1/1/2] undo ntdp enable
[Sysname-Ethernet1/1/2] bpdu-tunnel dot1q enable
```

3 Configuration on Provider C

# Configure BPDU transparent transmission on Ethernet 1/1/3.

```
<Sysname> system-view
[Sysname] interface ethernet 1/1/3
[Sysname-Ethernet1/1/3] port access vlan 2
[Sysname-Ethernet1/1/3] stp disable
[Sysname-Ethernet1/1/3] bpdu-tunnel dot1q enable
[Sysname-Ethernet1/1/3] bpdu-tunnel dot1q stp
```

# Configure BPDU transparent transmission on Ethernet 1/4.

```
[Sysname-Ethernet1/1/3] quit
[Sysname] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] port access vlan 2
[Sysname-Ethernet1/1/4] stp disable
[Sysname-Ethernet1/1/4] undo ntdp enable
[Sysname-Ethernet1/1/4] bpdu-tunnel dot1q enable
[Sysname-Ethernet1/1/4] bpdu-tunnel dot1q stp
```

14 VLAN Configuration

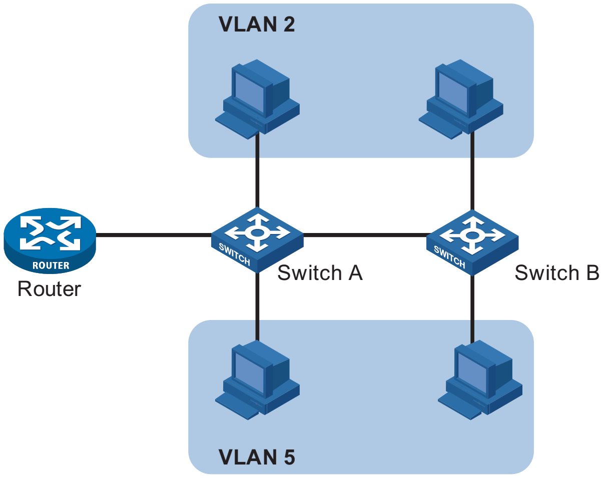
When configuring VLAN, go to these sections for information you are interested in:

Introduction to VLAN

VLAN Overview

The communication medium is shared in Ethernet. If the number of the hosts in the network reaches a certain level, problems caused by collisions, broadcasts, and so on emerge, resulting in improper network operation. Interconnecting LANs can suppress collisions but cannot isolate broadcast packets. Therefore, VLAN (virtual LAN) is developed to solve these problems. VLAN divides a LAN into multiple logical LANs with each being a broadcast domain. Hosts in the same VLAN can communicate with each other like in a LAN. However, hosts from different VLANs cannot communicate directly. In this way, broadcast packets are confined to a single VLAN, as illustrated in the following figure.

Figure 43 A VLAN diagram



A VLAN is not restricted by physical factors, that is to say, hosts that reside in different network segments may belong to the same VLAN; a VLAN can be with the same switch, or span across multiple switches or routers.

VLAN technology has the following advantages:

- Broadcast traffic is confined to each VLAN, reducing bandwidth utilization and improving network performance.
- LAN security is improved. Packets in different VLANs cannot communicate with each other directly. That is, users in a VLAN cannot interact directly with users in other VLANs, unless routers or Layer 3 switches are used.
- A more flexible way to establish virtual working groups. With VLAN technology, clients can be allocated to different working groups, and users from the same group do not have to be within the same physical area, making network construction and maintenance much easier and more flexible.

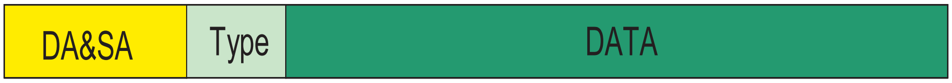
VLAN Fundamental

To enable network devices to identify packets of different VLANs, a field identifying VLANs is added to packets. As common switches operate on the data link layer, the field thus needs to be inserted to the data link layer encapsulation.

The format of the packets carrying the fields identifying VLANs is defined in IEEE 802.1Q, which is issued in 1999.

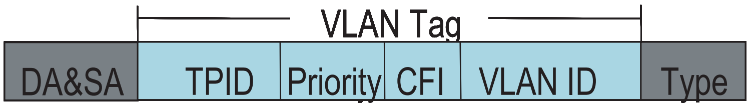
In the header of a traditional Ethernet packet, the field following the destination MAC address and the source MAC address is protocol type, which indicates the upper layer protocol type. Figure 44 illustrates the format of a traditional Ethernet packet, where DA stands for destination MAC address, SA stands for source MAC address, and Type stands for upper layer protocol type.

Figure 44 The format of a traditional Ethernet packet



IEEE 802.1Q defines a four-byte VLAN Tag field between the DA&SA field and the Type field to carry VLAN-related information, as shown in Figure 45.

Figure 45 The position and the format of the VLAN Tag field



The VLAN Tag field comprises four sub-fields: the TPID field, the Priority field, the CFI field, and the VLAN ID field.

- The TPID field, 16 bits in length, indicates that this data frame is VLAN-tagged. IEEE 802.1Q defines the value of this filed as 0x8100.
- The Priority field, three bits in length, indicates the priority of a packet. For information about packet priority, refer to *"Priority Mapping" on page 851*.

- The CFI field, one bit in length, specifies whether or not the MAC addresses are encapsulated in standard format when packets are transmitted across different medium. This field is not described here.
- The VLAN ID field, 12 bits in length and with its value ranging from 0 to 4095, identifies the ID of the VLAN a packet belongs to. As VLAN IDs of 0 and 4095 are reserved by the protocol, the actual value of this field ranges from 1 to 4094.

A network device determines the VLAN to which a packet belongs to by the VLAN ID field the packet carries. The VLAN Tag determines the way a packet is processed. For more information, refer to “Introduction to the Port-Based VLAN” on page 158.

VLAN Classification

VLANs can be classified into different categories. The following four types are the most commonly used:

- Port-based
- MAC address-based
- IP-subnet-based
- Protocol-based

At present, Switch 8800s support port-based VLANs and protocol-based VLANs.

The following contents introduce the configuration of port-based VLANs and protocol-based VLANs respectively.

Configuring Basic VLAN Attributes

Follow the following steps to configure basic VLAN attributes:

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.
Enter VLAN view	**vlan** *vlan-id*	Required The VLAN must be created first before entering its view; otherwise, using the command creates a VLAN and enters its view
Specify a description string for the VLAN	**description** *text*	Optional VLAN ID is used by default, for example, "VLAN 0001".

Configuring VLAN Interface Basic Attributes

VLAN interfaces are virtual interfaces used for communications between different VLANs. Each VLAN can have one VLAN interface. Packets of a VLAN can be forwarded on network layer through the corresponding VLAN interface. As each VLAN forms a broadcast domain, a VLAN can be an IP network segment and the VLAN interface can be the gateway to enable IP address-based Layer 3 forwarding.

Follow the following steps to configure VLAN interface basic attributes:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a VLAN interface or enter VLAN interface view	**interface vlan-interface** *vlan-interface-id*	Required This command leads you to VLAN interface view if the VLAN interface already exists.
Configure an IP address for the VLAN interface	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Optional Not configured by default
Specify the descriptive character string for the VLAN interface	**description** *text*	Optional VLAN interface name used by default
Bring up the VLAN interface	**undo shutdown**	Optional By default, the VLAN interface is down if all ports in the VLAN are down, as long as one port in the VLAN is up, the VLAN interface is up

Before creating a VLAN interface, ensure that the corresponding VLAN already exists. Otherwise, the specified VLAN interface will not be created.

Configuring the Port-Based VLAN

Introduction to the Port-Based VLAN

This is the simplest and yet the most effective way of classifying VLANs. It groups VLAN members by port. After added to a VLAN, a port can forward the packets of the VLAN.

Port link type

Based on the tag handling mode, a port's link type can be one of the following three:

- Access: an Access port only belongs to one VLAN, normally used to connect user device;
- Trunk: a Trunk port can belong to multiple VLANs, can receive and send packets of multiple VLANs, normally used to connect network devices;
- Hybrid: a Hybrid port can belong to multiple VLANs, can receive and send packets of multiple VLANs, used to connect either user or network devices;

The differences between Hybrid and Trunk ports:

- A Hybrid port allows packets of multiple VLANs to be sent without the VLAN tag;
- A Trunk port only allows packets from the default VLAN to be sent without the VLAN tag.

Default VLAN

You can configure the default VLAN for a port. By default, VLAN 1 is the default VLAN for all ports. However, this can be changed as needed.

- An Access port only belongs to one VLAN. Therefore, its default VLAN is the VLAN it belongs to and cannot be configured.
- You can configure the default VLAN for the Trunk port or the Hybrid port as they can both belong to multiple VLANs.
- If the VLAN removed through the **undo vlan** command is the default VLAN of a port, the default VLAN for an Access port reverts to VLAN 1, whereas that for the Trunk or Hybrid port keeps unchanged, meaning a Trunk or Hybrid port can use a nonexistent VLAN as the default VLAN.

Configured with the link type and default VLAN, a port handles packets in different ways, as described in the following table:

	Inbound packets handling		
Port type	**If no tag is carried in the packet**	**If a tag is carried in the packet**	**Outbound packets handling**
Access port	Tag the packet with the default VLAN ID	■ Receive the packet if its VLAN ID is the same as the default VLAN ID ■ Discard the packet if its VLAN ID is different from the default VLAN ID	Directly send the packet with the tag stripped as the VLAN ID is the default VLAN ID.
Trunk port		■ Receive the packet if the VLAN ID is the same as the default VLAN ID ■ Receive the packet if the VLAN ID is not the same as the default VLAN ID but is allowed to pass through the port	■ Strip the tag and send the packet if the VLAN ID is the same as the default VLAN ID. ■ Keep the tag and send the packet if the VLAN ID is not the same as the default VLAN ID.
Hybrid port		■ Discard the packet if the VLAN ID is neither the same as the default VLAN ID nor allowed to pass through the port	Send the packet if the VLAN ID is allowed on the port. You can use the **port hybrid vlan** command to configure whether the port keeps or strips the tags when sending the packets of the VLAN.

Configuring the Access-Port-Based VLAN

There are two ways to add an Access port to a VLAN: one way is to configure in VLAN view, the other way is to configure in Ethernet interface view or port group view.

Follow the following steps to configure the Access-port-based VLAN in VLAN view:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter VLAN view	**vlan** *vlan-id*	Required The VLAN must be created first before entering its view
Add an Access port to the current VLAN	**port** *interface-list*	Required By default, the system will add all ports to VLAN 1

Follow the following steps to configure the Access-port-based VLAN in Ethernet interface view/port group view:

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. Under Ethernet interface view, the subsequent configurations only apply to the current port; under port group view, the subsequent configurations apply to all ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the port link type as Access		**port link-type access**	Optional The link type of a port is Access by default.
Add the current Access port to a specified VLAN		**port access vlan** *vlan-id*	Optional By default, the system will add all ports to VLAN 1.

Ensure that you create a VLAN first before trying to add an Access interface to the VLAN.

Configuring the Trunk-Port-Based VLAN

A Trunk port may belong to multiple VLANs, and you can only perform this configuration in Ethernet interface view or port group view.

Follow the following steps to configure the Trunk-port-based VLAN:

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 Under Ethernet interface view, the subsequent configurations only apply to the current port; under port group view, the subsequent configurations apply to all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* **|** **aggregation** *agg-id* }	
Configure the port link type as Trunk		**port link-type trunk**	Required
Allow a specified VLAN to pass through the current Trunk port		**port trunk permit vlan** { *vlan-id-list* | **all** }	Required By default, all Trunk ports belong to VLAN 1 only

To do...	Use the command...	Remarks
Configure the default VLAN for the Trunk port	**port trunk pvid vlan** *vlan-id*	Optional VLAN 1 is the default by default

- *To convert a Trunk port into a Hybrid port (or vice versa), you need to use the Access port as a medium. For example, the Trunk port has to be configured as an Access port first and then a Hybrid port.*
- *The default VLAN ID on the Trunk ports of the local and peer devices must be the same. Otherwise, packets cannot be transmitted properly.*

Configuring the Hybrid-Port-Based VLAN

A Hybrid port may belong to multiple VLANs, and this configuration can only be performed in Ethernet interface view or port group view.

Follow the following steps to configure the Hybrid-port-based VLAN:

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; Under Ethernet interface view, the subsequent configurations only apply to the current port; under port group view, the subsequent configurations apply to all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the port link type as Hybrid		**port link-type hybrid**	Required
Allow a specified VLAN to pass through the current Hybrid port		**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required By default, all Hybrid ports belong to VLAN 1
Configure the default VLAN of the Hybrid port		**port hybrid pvid vlan** *vlan-id*	Optional VLAN 1 is the default by default

- *To configure a Trunk port into a Hybrid port (or vice versa), you need to use the Access port as a medium. For example, the Trunk port has to be configured as an Access port first and then a Hybrid port.*
- *Ensure that a VLAN already exists before configuring it to pass through a certain Hybrid port.*
- *The default VLAN ID on the Hybrid ports of the local and the peer devices must be the same. Otherwise, packets cannot be transmitted properly.*

Configuring the Protocol-Based VLAN

Introduction to the Protocol-Based VLAN

In this approach, inbound packets are assigned with different VLAN IDs based on their protocol type and encapsulation format. The protocols that can be used to categorize VLANs include: IP, IPX, and AppleTalk (AT). The encapsulation formats include: Ethernet II, 802.3, 802.3/802.2 LLC, and 802.3/802.2 SNAP.

A protocol-based VLAN can be defined by a protocol template, which is determined by the encapsulation format and protocol type. A port can be associated to multiple protocol templates. An untagged packet (that is, packet carrying no VLAN tag) reaching a port associated with a protocol-based VLAN will be processed as follows.

- If the packet matches a protocol template, the packet will be tagged with the VLAN ID of the protocol-based VLAN defined by the protocol template, and then sent to the specified VLAN.
- If the packet matches no protocol template, the packet will be tagged with the default VLAN ID of the port.

A tagged packet (that is, a packet carrying VLAN tags) reaching the port is processed in the same way as that of port-based VLAN.

- If the port is configured to permit packets with the VLAN tag, the packet is forwarded.
- If the port is configured to deny packets with the VLAN tag, the packet is dropped.

This feature is mainly used to bind the service type with VLAN for ease of management and maintenance.

Configuring the Protocol-Based VLAN

This feature is only applicable to the Hybrid port.

Follow the following steps to configure the protocol-based VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	Required For a nonexistent VLAN, this command will create a VLAN and enter its view.
Configure the protocol based VLAN and specify the protocol template	**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
Return to system view	**quit**	-

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. Under Ethernet interface view, the subsequent configurations only apply to the current port; under port group view, the subsequent configurations apply to all ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the port link type as Hybrid		**port link-type hybrid**	Required
Allow specified VLANs to pass through the current Hybrid port		**port hybrid vlan** *vlan-id-list* { **tagged** | **untagged** }	Required
Configure the association between the Hybrid port and the protocol-based VLAN		**port hybrid protocol-vlan vlan** *vlan-id* { *protocol-index* [**to** *protocol-end*] | **all** }	Required

CAUTION:

- *You cannot configure both dsap-id and ssap-id as 0xE0 or 0xFF; otherwise the matching packets will take the same encapsulation format as that of the **ipx llc** packets and the **ipx raw** packets respectively.*
- *When you use the **mode** keyword to configure a user-defined protocol template, do not set the etype-id argument for **ethernetii** packets to 0x0800, 0x86DD, 0x809B, or 0x8137; otherwise, the matching packets will take the same format as that of the IPv4, IPv6, IPX, and AppleTalk packets respectively.*

Displaying and Maintaining VLAN

To do...	Use the command...	Remarks
Display VLAN information	**display vlan** [*vlan-id1* [**to** *vlan-id2*] | **all** | **static** | **dynamic** | **reserved**]	Available in any view
Display VLAN interface information	**display interface vlan-interface** [*vlan-interface-id*]	Available in any view
Display information about the protocol-based VLAN configured on the specified VLANs	**display protocol-vlan vlan** { *vlan-id* [**to** *vlan-id*] | **all** }	Available in any view
Display the protocol information and protocol indexes configured on the specified interfaces	**display protocol-vlan interface** { *interface-type interface-number* [**to** *interface-type interface-number*] | **all** }	Available in any view

VLAN Configuration Examples

Port-Based VLAN Configuration Example

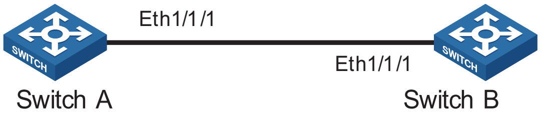
Network requirements

- Switch A connects to Switch B through Trunk port Ethernet 1/1/1;

- The default VLAN ID of the Trunk port is 100;
- The Trunk port allows packets from VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass.

Network diagram

Figure 46 Network diagram for port-based VLAN configuration



Configuration procedure

Configure Switch A:

# Create VLAN 100.

```
<SysnameA> system-view
[SysnameA] vlan 100
[SysnameA-vlan100] quit
```

# Enter Ethernet interface view of Ethernet 1/1/1.

```
[SysnameA] interface ethernet 1/1/1
```

# Configure Ethernet 1/1/1 as a Trunk port and configure its default VLAN ID as 100.

```
[SysnameA-Ethernet1/1/1] port link-type trunk
[SysnameA-Ethernet1/1/1] port trunk pvid vlan 100
```

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

```
[SysnameA-Ethernet1/1/1] port trunk permit vlan 2 6 to 50 100
```

Configuration on Switch B is the same as that on Switch A.

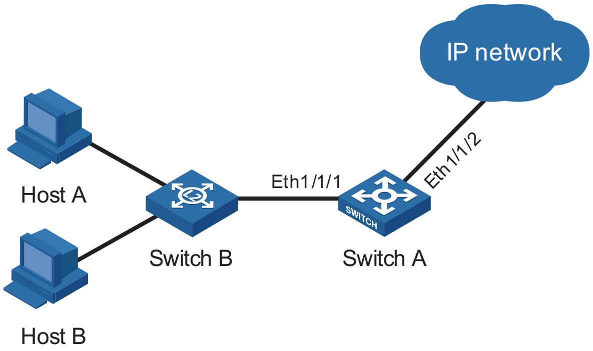
Protocol-Based VLAN Configuration Example

Network requirements

- Switch A connects to Switch B through Hybrid port Ethernet 1/1/1, and accesses to an IP network through port Ethernet 1/1/2.
- Switch B is a common switch, which connect with multiple hosts for different applications.
- Through protocol-based VLAN configuration, make Ethernet 1/1/1 to forward the received IPv4 packets to VLAN 2, and IPv6 packets to VLAN 6.

Network diagram

Figure 47 Network diagram for protocol-based VLAN configuration



Configuration procedure

# Create VLAN 2 and VLAN 6 and configure them as protocol-based VLANs.

```
<Sysname> system-view
[Sysname] vlan 2
[Sysname-vlan2] protocol-vlan ipv4
[Sysname-vlan2] quit
[Sysname] vlan 6
[Sysname-vlan6] protocol-vlan ipv6
[Sysname-vlan6] quit
```

# Configure Ethernet 1/1/1 and Ethernet 1/1/2 as Hybrid ports which permit packets of VLAN 2 and VLAN 6 to pass, and associate Ethernet 1/1/1 with the protocol-based VLANs.

```
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port link-type hybrid
[Sysname-Ethernet1/1/1] port hybrid vlan 2 6 tagged
[Sysname-Ethernet1/1/1] port hybrid protocol-vlan vlan 2 all
[Sysname-Ethernet1/1/1] port hybrid protocol-vlan vlan 6 all
[Sysname-Ethernet1/1/1] quit
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] port link-type hybrid
[Sysname-Ethernet1/1/2] port hybrid vlan 2 6 tagged
```

15 SUPER VLAN CONFIGURATION

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

- "Introduction to Super VLAN" on page 167
- "Configuring Super VLAN" on page 167
- "Displaying Super VLAN" on page 168
- "Super VLAN Configuration Example" on page 168

Introduction to Super VLAN

With the development of networks, network address resource has become more and more scarce. The concept of Super VLAN was introduced to save the IP address space. Super VLAN is also named as VLAN aggregation. A super VLAN involves multiple sub-VLANs and has a VLAN interface with an IP address. Sub-VLANs are configured with no VLAN interfaces and isolated from each other on Layer 2. If Layer 3 communication is needed from a sub-VLAN, it uses the IP address of the VLAN interface of the super VLAN as the gateway IP address. Thus, multiple sub-VLANs share the same gateway address and therefore saving the IP address resource.

The local proxy Address Resolution Protocol (ARP) function is used to realize Layer 3 communications between sub-VLANs and between sub-VLANs and other networks. It works as follows: after creating the super VLAN and the VLAN interface, enable the local proxy ARP function on the VLAN interface, so that the super VLAN can forward ARP responses and requests to realize the Layer 3 communication between sub-VLANs.

For introduction on proxy ARP, refer to "Proxy ARP Configuration" on page 201.

Configuring Super VLAN

Follow the following steps to configure super VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure the VLAN as a super VLAN	**supervlan**	Required
Correlate the super VLAN with its sub-VLAN(s)	**subvlan** *vlan-list*	Required The specified sub-VLANs must have existed.
Return to system view	**quit**	-

To do...	Use the command...	Remarks
Enter VLAN interface view	**interface vlan-interface** *vlan-interface-id*	-
Configure the IP address of the VLAN interface	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Required By default, the IP address of a VLAN interface is not configured
Enable local proxy ARP	**local-proxy-arp enable**	Required Disabled by default

- *The IP address of the VLAN interface configured in the above table is the IP address of the corresponding VLAN interface of the super VLAN.*
- *For more information about the* ***local-proxy-arp enable*** *command, refer to the Switch 8800 Command Reference Guide.*
- *A VLAN that is configured as a super VLAN cannot be configured as the Guest VLAN for a certain port, and vice versa. For more information, refer to "Configuring a Guest VLAN" on page 922.*

CAUTION:

- *If a port is already added to a VLAN, the VLAN cannot be configured as a super VLAN.*
- *Layer 2 multicast function can be configured for a super VLAN, but the function does not take effect.*
- *The functions of DHCP, Layer 3 multicast, dynamic routing, and NAT can be configured on the VLAN interface of a super VLAN, but only DHCP takes effect.*
- *You cannot configure VRRP on the VLAN interface of a super VLAN.*

Displaying Super VLAN

To do...	Use the command...	Remarks
Display the mapping between a super VLAN and its sub-VLAN(s)	**display supervlan** [*supervlan-id*]	Available in any view

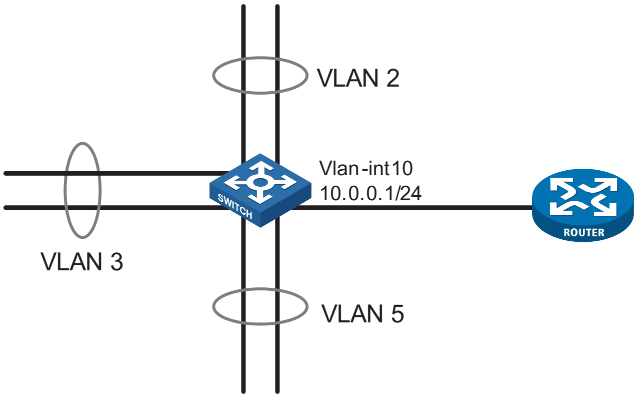
Super VLAN Configuration Example

Network requirements

- Create super VLAN 10, and configure the VLAN interface IP address of the super VLAN as 10.0.0.1/24.
- Create the sub-VLANs: VLAN 2, VLAN 3, and VLAN 5.
- Ports Ethernet 0/1/1 and Ethernet 0/1/2 belong to VLAN 2, Ethernet 0/1/3 and Ethernet 0/1/4 belong to VLAN 3, and Ethernet 0/1/5 and Ethernet 0/1/6 belong to VLAN 5.
- Through configuration, the sub-VLANs are isolated at Layer 2 but connected at Layer 3.

Network diagram

Figure 48 Network diagram for super-VLAN configuration



Configuration procedure

# Create VLAN 10, configure its VLAN interface address as 10.0.0.1/24.

```
<Sysname> system-view
[Sysname] vlan 10
[Sysname-vlan10] quit
[Sysname] interface vlan-interface 10
[Sysname-Vlan-interface10] ip address 10.0.0.1 255.255.255.0
```

# Enable local proxy ARP.

```
[Sysname-Vlan-interface10] local-proxy-arp enable
[Sysname-Vlan-interface10] quit
```

# Create VLAN 2, add ports Ethernet 0/1/1 and Ethernet 0/1/2 to it.

```
[Sysname] vlan 2
[Sysname-vlan2] port ethernet 0/1/1 ethernet 0/1/2
```

# Create VLAN 3, add ports Ethernet 0/1/3 and Ethernet 0/1/4 to it.

```
[Sysname-vlan2] vlan 3
[Sysname-vlan3] port ethernet 0/1/3 ethernet 0/1/4
```

# Create VLAN 5, add ports Ethernet 0/1/5 and Ethernet 0/1/6 to it.

```
[Sysname-vlan3] vlan 5
[Sysname-vlan5] port ethernet 0/1/5 ethernet 0/1/6
```

# Specify VLAN 10 as the super VLAN, and VLAN 2, VLAN 3, and VLAN 5 as the sub-VLANs.

```
[Sysname-vlan5] vlan 10
[Sysname-vlan10] supervlan
[Sysname-vlan10] subvlan 2 3 5
```

16 ISOLATE-USER VLAN CONFIGURATION

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

Introduction to Isolate-User-VLAN

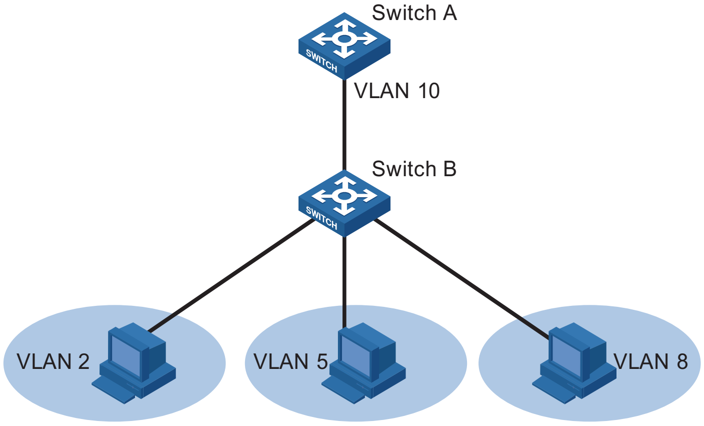
The isolate-user-VLAN adopts a two-tier VLAN structure. In this approach, two types of VLANs, isolate-user-VLAN and secondary VLAN, are configured on the same device.

- The isolate-user-VLAN is mainly used for upstream data exchange. An isolate-user-VLAN can have multiple secondary VLANs associated to it. The upstream device only knows the isolate-user-VLAN, how the secondary VLANs are working is not its concern. In this way, network configurations are simplified and VLAN resources are saved.
- Secondary VLANs are used for connecting users. Secondary VLANs are isolated from each other on Layer 2.
- One isolate-user-VLAN can have multiple secondary VLANs, which are invisible to the corresponding upstream device.

As illustrated in Figure 49, 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 that are mapped to VLAN 10 and invisible to Switch A. To realize the Layer 3 connectivity between the secondary VLANs (VLAN 2, VLAN 5, and VLAN 8) that are under the same isolate-user-VLAN (VLAN 10), the following two methods can be used:

- Configure a VLAN interface and the VLAN interface IP address for each secondary VLAN on Switch B.
- Configure the local proxy ARP function on the upper layer device (Switch A). For detailed information about proxy ARP, refer to *"Proxy ARP Configuration" on page 201*.

Figure 49 Network diagram for isolate-user-VLAN configuration



Configuring Isolate-User-VLAN

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

1. Create the isolate-user-VLAN;
2. Create the secondary VLAN;
3. Add ports to the isolate-user-VLAN (note that the ports cannot be Trunk ports) and ensure that at least one port has the isolate-user-VLAN as its default VLAN;
4. Add ports to the secondary VLAN (note that the ports cannot be Trunk ports) and ensure that at least one port has the secondary VLAN as its default VLAN;
5. Configure the mapping between the isolate-user-VLAN and the secondary VLAN.

Follow the following steps to configure isolate-user-VLAN:

To do...		Use the command	Remarks
Enter system view		**system-view**	-
Create a VLAN (or enter VLAN view)		**vlan** *vlan-id*	-
Configure the VLAN as an isolate-user-VLAN		**isolate-user-vlan enable**	Required
Quit to system view		**quit**	-
Add ports to the isolate-user-VLAN and ensure that at least one port has the isolate-user-VLAN as its default VLAN	Access port	Refer to "Configuring the Access-Port-Based VLAN" on page 159	Required to choose either
	Hybrid port	Refer to "Configuring the Hybrid-Port-Based VLAN" on page 161	
Quit to system view		**quit**	-
Create the secondary VLAN		**vlan** *vlan-id*	-
Quit to system view		**quit**	-

To do...		Use the command	Remarks
Add ports to the secondary VLAN and ensure that at least one port has the secondary VLAN as its default VLAN	Access port	Refer to "Configuring the Access-Port-Based VLAN" on page 159	Required to choose either
	Hybrid port	Refer to "Configuring the Hybrid-Port-Based VLAN" on page 161	
Quit to system view		**quit**	-
Configure the mapping between the isolate-user-VLAN and secondary VLAN		**isolate-user-vlan** *isolate-user-vlan-id* **secondary** *secondary-vlan-id* [**to** *secondary-vlan-id*]	Required

- *To create an isolate-user-VLAN, you need to disable the GVRP function first, and vice versa.*
- *After a mapping is configured, the system disallows adding ports to and removing ports or VLANs from the mapped isolate-user-VLAN and secondary VLAN.*

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

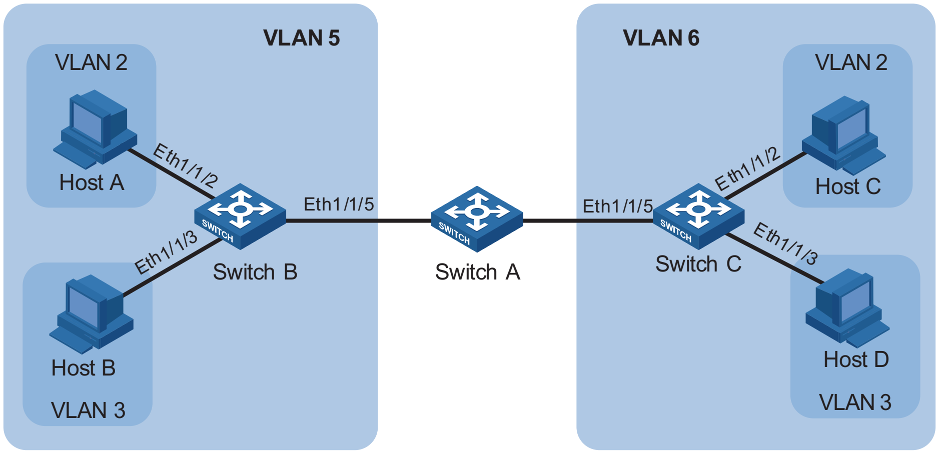
Isolate-User-VLAN Configuration Example

Network requirements

- Switch A is connected to Switch B and Switch C;
- On Switch B, VLAN 5 is configured as an isolate-user-VLAN, which includes an upstream port Ethernet 1/1/5 and two secondary VLANs VLAN 2 and VLAN 3. VLAN 2 has Ethernet 1/1/2 and VLAN 3 has Ethernet 1/1/3.
- On Switch C, VLAN 6 is configured as an isolate-user-VLAN, which includes an upstream port Ethernet 1/1/5 and two secondary VLANs VLAN 3 and VLAN 4. VLAN 3 has Ethernet 1/1/3 and VLAN 4 has Ethernet 1/1/2.
- Through the configuration, for Switch A, Switch B only has one VLAN (VLAN 5) and Switch C only has one VLAN (VLAN 6).

Network diagram

Figure 50 Isolate-User-VLAN configuration diagram



Configuration procedure

The following are the configuration procedures for Switch B and Switch C.

1 Configure Switch B

# Configure the isolate-user-VLAN.

```
<SysnameB> system-view
[SysnameB] vlan 5
[SysnameB-vlan5] isolate-user-vlan enable
[SysnameB-vlan5] port ethernet1/1/5
[SysnameB-vlan5] quit
```

# Configure the secondary VLANs.

```
[SysnameB] vlan 3
[SysnameB-vlan3] port ethernet1/1/3
[SysnameB-vlan3] quit
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet1/1/2
[SysnameB-vlan2] quit
```

# Establish the mapping between the isolate-user-VLAN and the secondary VLANs.

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

2 Configure Switch C

# Configure the isolate-user-VLAN.

```
<SysnameC> system-view
[SysnameC] vlan 6
[SysnameC-vlan6] isolate-user-vlan enable
[SysnameC-vlan6] port ethernet1/1/5
[SysnameC-vlan6] quit
```

# Configure the secondary VLANs.

```
[SysnameC] vlan 3
[SysnameC-vlan3] port ethernet1/1/3
[SysnameC-vlan3] quit
[SysnameC] vlan 2
[SysnameC-vlan2] port ethernet1/1/2
```

# Establish the mapping between the isolate-user-vlan and the secondary VLANs.

```
[SysnameC-vlan2] quit
[SysnameC] isolate-user-vlan 6 secondary 2 to 3
```

Verification

# Display the isolate-user-VLAN configuration on Switch B.

```
<SysnameB> 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
Tagged   Ports: none
 Untagged Ports:
    Ethernet1/1/2            Ethernet1/1/3            Ethernet1/1/5
 VLAN ID: 2
 VLAN Type: static
 Isolate-user-VLAN type : secondary
Route Interface: not configured
 Description: VLAN 0002
Tagged   Ports: none
 Untagged Ports:
    Ethernet1/1/2            Ethernet1/1/5

 VLAN ID: 3
 VLAN Type: static
 Isolate-user-VLAN type : secondary
 Route Interface: not configured
 Description: VLAN 0003
Tagged   Ports: none
 Untagged Ports:
    Ethernet1/1/3            Ethernet1/1/5
```

The isolate-user-VLAN configuration on Switch C is similar to that on Switch B.

17 PORT ISOLATION CONFIGURATION

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

Introduction to Port Isolation

To implement Layer 2 isolation, you can add different ports to different VLANs. However, this will waste the limited VLAN resource. With port isolation, the ports can be isolated within the same VLAN. Thus, you need only to add the ports to the isolation group to implement Layer 2 isolation. This provides you with more secure and flexible networking schemes.

To enable the interconnection between an isolation group and Layer 2 outside the isolation group, you must configure an uplink port for the isolation group. The last configuration will overwrite the previous configurations if you configure different ports as the uplink port.

- Layer 2 traffic can pass from the ports in the isolation group to the uplink port.
- To enable the Layer 2 traffic to pass from the uplink port to the port in a certain isolation group, you must configure these two ports to be in the same VLAN.

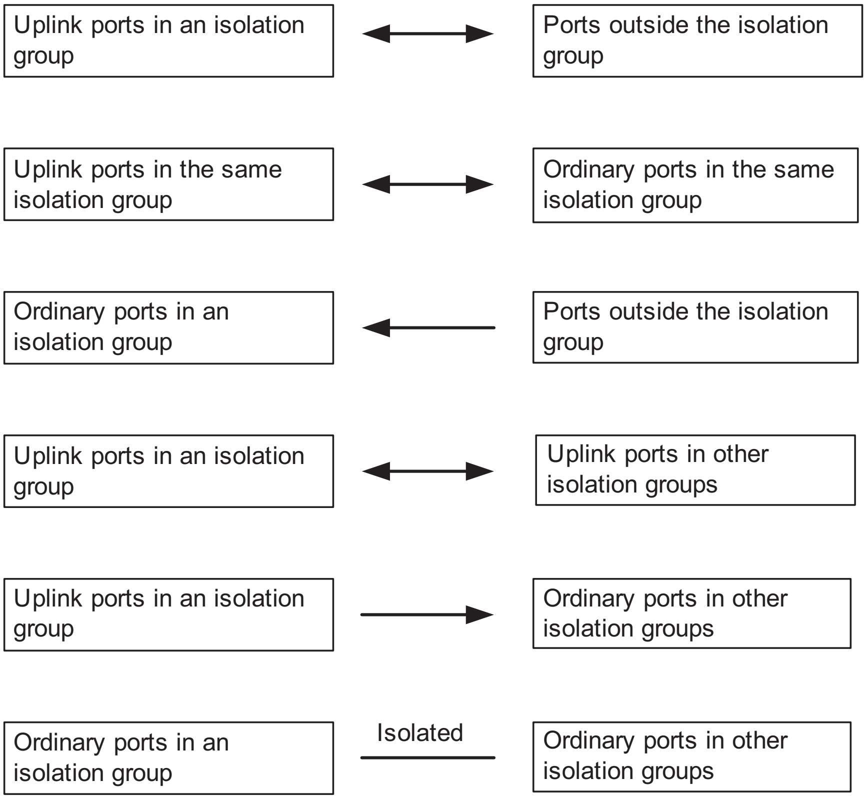
At present, for the Switch 8800 Families:

- A maximum of 64 isolation groups can be configured.
- There is no restriction on the number of ports to be added to an isolation group.

- *When a port in the summary group is configured as the ordinary port for some isolation group, the other ports of the summary group can be added to the isolation group as ordinary ports but cannot be configured as uplink ports.*
- *When a port in the summary group is configured as the uplink port for some isolation group, the other ports of the summary group cannot be added to the isolation group and the other ports of the device cannot be added to the summary group.*
- *The port isolation feature only isolates Layer 2 data instead of Layer 3 data.*

Port isolation is independent of the VLAN the port belongs to. For ports belonging to different VLANs, Layer 2 data can pass only from the ordinary port to the uplink port in the same isolation group unidirectionally. Within the same VLAN, there are two types of connectivity of Layer 2 data on ports within and outside the isolation group, as shown in Figure 1-1:

Figure 51 Connectivity of layer 2 data between ports inside and outside an isolation group on a device supporting uplink port


Uplink ports in an isolation group
Ports outside the isolation group
Uplink ports in the same isolation group
Ordinary ports in the same isolation group
Ordinary ports in an isolation group
Ports outside the isolation group
Uplink ports in an isolation group
Uplink ports in other isolation groups
Uplink ports in an isolation group
Ordinary ports in other isolation groups
Ordinary ports in an isolation group
Isolated
Ordinary ports in other isolation groups

- *The arrows in the above figure indicate the transmission direction of layer 2 data.*
- *Within the same VLAN, the ports outside the isolation group can access the Layer 2 data of the ordinary ports inside the isolation group, but the ordinary ports in the isolation group cannot access the Layer 2 data of the ones outside the isolation group.*

Configuring Isolation Groups on a Device

Adding 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**	-
Create an isolation group	**port-isolate group** *group-number*	Required

To do...		Use the command...	Remarks
Enter Ethernet interface view or port group view	Enter Ethernet interface view	**interface** *interface-type interface-number*	One of them is required. Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Add a specified port to the isolation group as an ordinary port		**port-isolate enable group** *group-number*	Required No ports are added to the isolation group by default.

Configuring an Uplink Port for the Isolation Group

Follow these steps to configure an uplink port for the isolation group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Configure the current port as the uplink port of the isolation group	**port-isolate uplink-port group** *group-number*	Required An isolation group has no uplink port by default.

- *An isolation group can have only one uplink port. When a user configures multiple ports as the uplink port, only the last one prevails.*
- *If a port has already been configured as an ordinary port for an isolation group, it cannot be configured as an uplink port, and vice versa.*

Displaying Isolation Groups

Follow these steps to display and maintain an isolation group:

To do...	Use the command...
Display an isolation group and its information	**display port-isolate group** [*group-number*]

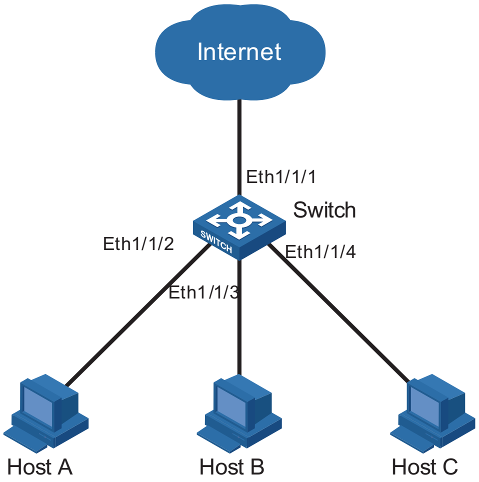
Port Isolation Configuration Example

Networking Requirement

- Users Host A, Host B, and Host C are connected to Ethernet 1/1/2, Ethernet 1/1/3, and Ethernet 1/1/4 of Device.
- A switch is connected to an external network through Ethernet 1/1/1.
- Ethernet 1/1/1, Ethernet 1/1/2, Ethernet 1/1/3, and Ethernet 1/1/4 belong to the same VLAN (VLAN 2). It is desired that Host A, Host B, and Host C cannot communicate with each other, but can access the external network.

Networking diagram

Figure 52 Networking diagram for port isolation configuration



Configuration procedure

# Create a VLAN, and add the ports to this VLAN.

```
<Sysname> system-view
[Sysname] vlan 2
[Sysname-vlan2] port ethernet 1/1/1 to ethernet 1/1/4
[Sysname-vlan2] quit
```

# Create Isolation Group 2.

```
[Sysname] port-isolate group 2
```

# Add Ethernet 1/1/2, Ethernet 1/1/3, and Ethernet 1/1/4 to Isolation Group 2.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] port-isolate enable group 2
[Sysname-Ethernet1/1/2] interface ethernet 1/1/3
[Sysname-Ethernet1/1/3] port-isolate enable group 2
[Sysname-Ethernet1/1/3] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] port-isolate enable group 2
```

# Configure port Ethernet1/1/1 as the uplink port of Isolation Group 2.

```
[Sysname-Ethernet1/1/4] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port-isolate uplink-port group 2
[Sysname-Ethernet1/1/1] return
```

# Display information of Isolation Group 2.

```
<Sysname> display port-isolate group 2
Port-isolate group information:
Uplink port support: YES
Group ID: 2
Uplink port: Ethernet1/1/1
   Ethernet1/1/2     Ethernet1/1/3     Ethernet1/1/4
```

18 QinQ Configuration

Introduction to QinQ

Understanding QinQ

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 4,094 VLANs. In actual applications, however, a large number of VLAN are required to isolate users, especially in metropolitan area networks (MANs), and 4,094 VLANs are far from satisfying such requirements.

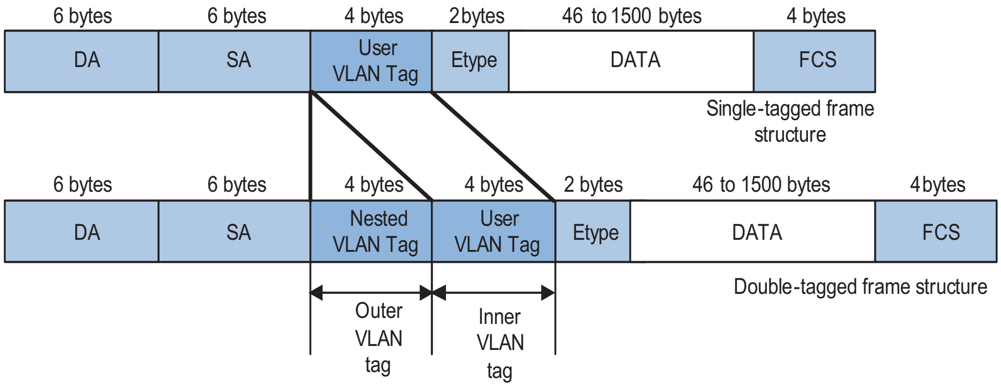
The port QinQ feature provided by the device tags a frame with double VLAN tags, allowing for 4094 × 4094 VLANs, thus satisfying the demand for large amount of VLANs in MANs.

The QinQ feature encapsulates the private network VLAN tag in the public network VLAN tag, and enables the frame to be transmitted through the service provider backbone network (the public network) with double VLAN tags. The inner VLAN tag is the customer network VLAN tag while the outer one is the VLAN tag assigned by the service provider to the customer. In the public network, frames are forwarded based on the outer VLAN tag only, while the customer network VLAN tag remains untouched.

- *To implement QinQ, configuration is required on devices in the service provider network only.*
- *The QinQ feature is implemented based on the 802.1q standard, so it is necessary that all the devices along the tunnel support the 802.1q standard.*

Figure 53 shows the structure of a single-tagged frame and a double-tagged frame.

Figure 53 Single-tagged frame structure vs. double-tagged Ethernet frame structure



Advantages of QinQ:

- Saves the public network VLAN ID resources.
- Enables customers to plan their own private network VLAN IDs, without running into conflicts with public network VLAN IDs.
- Provides a simple Layer 2 VPN solution for small-sized MANs or Enterprise networks.

Implementations of QinQ

For Switch 8800s, the QinQ feature is implemented through enabling the basic QinQ feature on ports.

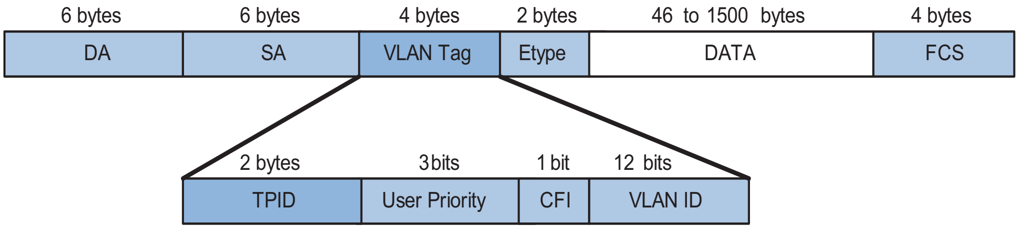
With the basic QinQ feature enabled on a port, when a frame arrives at the port, the port will tag it with the port's default VLAN tag, regardless of whether the frame is tagged or untagged. If the received frame is already tagged, this frame becomes a double-tagged frame; if it is an untagged frame, it is tagged with the port's default VLAN tag.

Adjustable TPID Value of QinQ Frames

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.

Figure 54 shows the structure of a VLAN-tagged Ethernet frame defined in IEEE 802.1Q.

Figure 54 Structure of a VLAN-tagged Ethernet frame



Refer to "VLAN Configuration" on page 155 for the description on the TPID, Priority, CFI, and VLAN ID fields.

On devices of different vendors, the TPID of the outer VLAN tag of QinQ frames may have different default values. You can set and/or modify this TPID value, so that the QinQ frames, when arriving at the public network, carries the TPID value of a specific vendor to allow interoperation with 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 chaotic packet forwarding and receiving, you cannot set the TPID value to any of the values in the table below.

Table 17 Protocol type values

Protocol type	Value
ARP	0x0806
PUP	0x0200
RARP	0x8035
IP	0x0800
IPv6	0x86DD

Table 17 Protocol type values

Protocol type	Value
PPPoE	0x8863/0x8864
MPLS	0x8847/0x8848
IPX/SPX	0x8137
IS-IS	0x8000
LACP	0x8809
802.1x	0x888E
Cluster	0x88A7
Reserved	0xFFFD/0xFFFE/0xFFFF

Configuring Basic QinQ

Follow these steps to configure basic QinQ:

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 Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the basic QinQ function for the Ethernet port or the port group		**qinq enable**	Required Disabled by default.

CAUTION:

- *The basic QinQ function must be enabled on ports (of devices in the service provider network) with customer networks connected to them.*
- *As basic QinQ function affects layer-3 packet forwarding and MPLS switching, do not enable layer-3 packet forwarding or MPLS switching on ports with basic QinQ function enabled.*

Setting TPID Value for QinQ Frames

Follow these steps to set the TPID value for QinQ frames:

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 Configured in Ethernet interface view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Set the TPID value		**qinq ethernet-type** *hex-value*	Optional 0x8100 by default

CAUTION:

- *Perform the above configuration on ports (of devices in the service provider network) with customer networks connected to them.*
- *The* ***qinq ethernet-type*** *command needs to be coupled with the* ***qinq enable*** *command.*

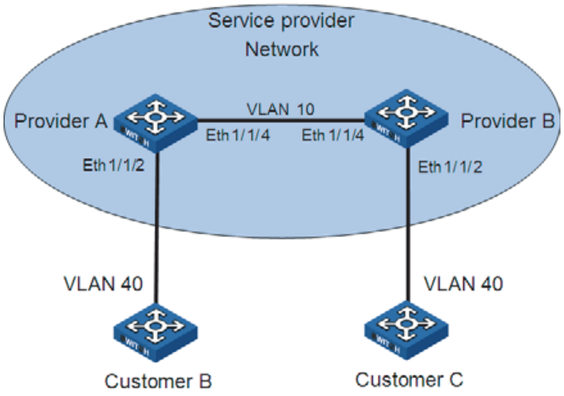
QinQ Configuration Examples

Network requirements

- Provider A and Provider B are service provider network access devices.
- Customer A and Customer B are customer network access devices.
- Provider A and Provider B, belonging to VLAN 10 created in the service provider network, are interconnected through trunk ports.
- Ethernet 1/1/2 on Provider A belongs to VLAN 10 and is QinQ-enabled.
- The TPID values of Ethernet 1/1/4 on Provider A and Ethernet 1/1/4 on Provider B are both 0x8200.

Network diagram

Figure 55 Network diagram for QinQ configuration



Configuration procedure

1 Configuration on Provider A

# Enter system view.

```
<Sysname> system-view
[Sysname] vlan 10
[Sysname-vlan10] quit
```

# Enter Ethernet 1/1/2 port view.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] port access vlan 10
```

# Enable the basic QinQ function on Ethernet 1/1/2.

```
[Sysname-Ethernet1/1/2] qinq enable
[Sysname-Ethernet1/1/2] quit
```

# Configure Ethernet 1/1/4 as a trunk port and configure the port to permit frames of VLAN 10.

```
[Sysname] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] port link-type trunk
[Sysname-Ethernet1/1/4] port trunk permit vlan 10
```

# Set the TPID value of Ethernet 1/1/4 to 0x8200.

```
[Sysname-Ethernet1/1/4] qinq ethernet-type 8200
```

2 Configuration on Provider B

# Enter system view.

```
<Sysname> system-view
[Sysname] vlan 10
[Sysname-vlan10] quit
```

# Enter Ethernet 1/1/2 port view.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] port access vlan 10
```

# Enable the basic QinQ function on Ethernet 1/1/2.

```
[Sysname-Ethernet1/1/2] qinq enable
[Sysname-Ethernet1/1/2] quit
```

# Configure Ethernet 1/1/4 as a trunk port and configure the port to permit frames of VLAN 10.

```
[Sysname] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] port link-type trunk
[Sysname-Ethernet1/1/4] port trunk permit vlan 10
```

# Set the TPID value of Ethernet 1/1/4 to 0x8200.

```
[Sysname-Ethernet1/1/4] qinq ethernet-type 8200
```

After the above configuration, the frames of VLAN 40 on Customer A are double-tagged when they are transmitted between the trunk ports of Provider A and Provide B.

19 IP ROUTING OVERVIEW

Go to these sections for information you are interested in:

- "IP Routing and Routing Table" on page 187
- "Routing Protocol Overview" on page 189
- "Displaying and Maintaining a Routing Table" on page 191

The term "router" or router icon 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 Through a 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.

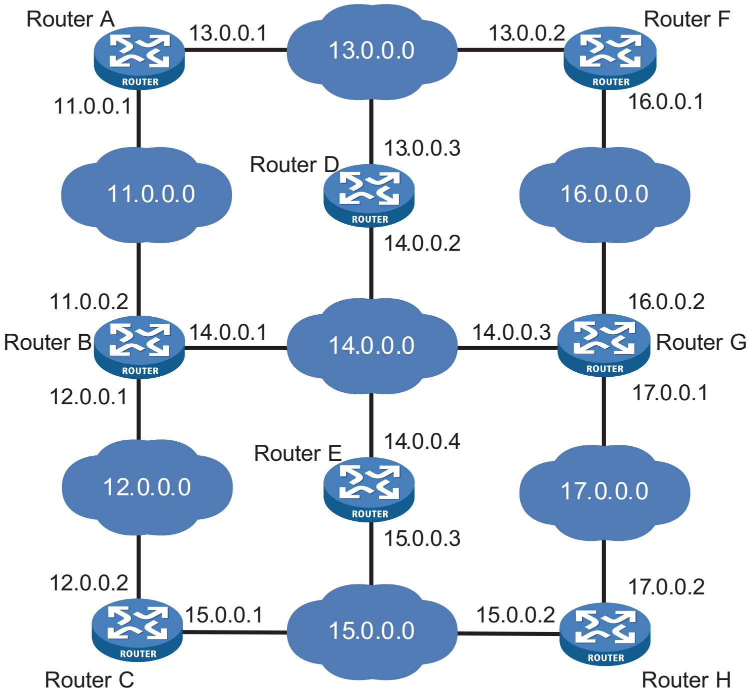
Based on whether the destination is directly connected to a given router, routes can be divided into:

- 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 entry in the routing table will be forwarded through the default route.

In the following figure, the IP address on each cloud represents the address of the network. Router G resides in three networks and therefore has three IP addresses for its three physical interfaces. Its routing table is shown on the right of the network topology.

Figure 56 A sample routing table



Destination Network	Nexthop	Interface
10.0.0.0	10.0.0.1	2
11.0.0.0	11.0.0.1	1
12.0.0.0	11.0.0.2	1
13.0.0.0	13.0.0.4	3
14.0.0.0	13.0.0.2	3
15.0.0.0	13.0.0.2	3
16.0.0.0	10.0.0.2	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 complicated to configure, and that it not only imposes higher requirements on the system, but also eats away 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, typically include RIP, OSPF, and IS-IS.
- Exterior Gateway Protocols (EGPs): Work between autonomous systems. The most popular one is BGP.

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: Include mainly RIP and BGP. BGP is also considered a path-vector protocol.
- Link-state protocols: Include mainly 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: Includes RIP, OSPF, BGP, and IS-IS.
- Multicast routing protocols: Includes PIM-SM and PIM-DM.

This chapter focuses on unicast routing protocols. For information on multicast routing protocols, refer to *"IPv6 Multicast Routing and Forwarding Configuration" on page 515*.

Version of IP protocol

IPv4 routing protocols: RIP, OSPF, BGP and IS-IS.

IPv6 routing protocols: RIPng, OSPFv3, BGP4+, 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 routing 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	255

- *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.

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.

Sharing of Routing Information

As different routing protocols use different 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 *"Routing Protocol Overview" on page 189*.

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 public network routing table or a VPN routing table	**display ip routing-table** [**vpn-instance** *vpn-instance-name*] **statistics**	Available in any view
Clear statistics for the public 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

To do...	Use the command...	Remarks
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

20 ARP CONFIGURATION

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

ARP Overview

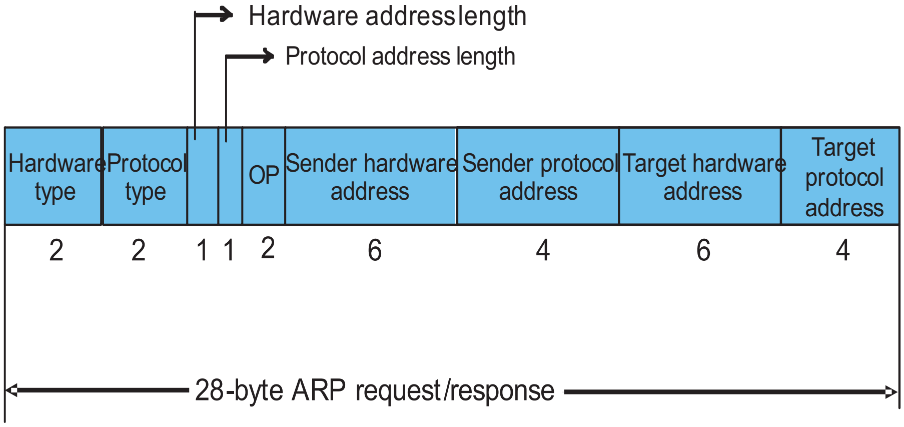
ARP Function

Address resolution protocol (ARP) is used to resolve an IP address into a MAC address.

An IP address is the address of a host at the network layer. To send a network layer packet to a destination host, the device must know the MAC address of the destination host. To this end, the IP address must be resolved into the corresponding MAC address. Each host maintains an IP-to-MAC mapping table that contains IP and MAC addresses of devices that communicated with the host recently.

ARP Message Format

Figure 57 ARP message format



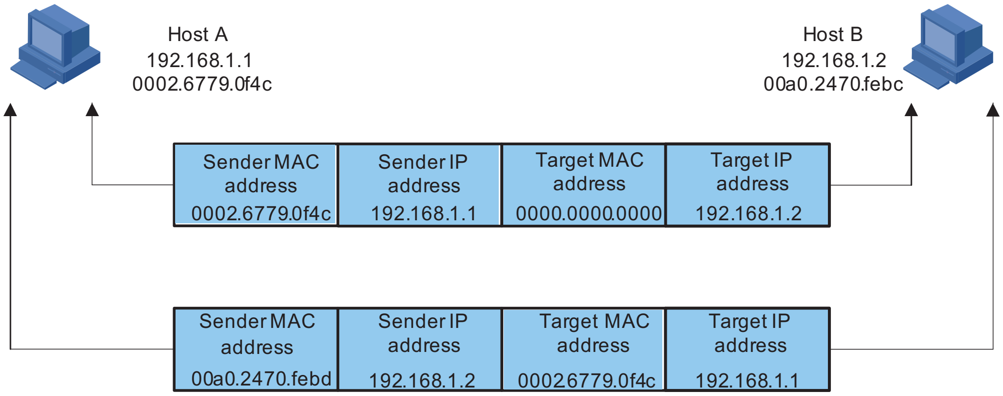
The following explains the fields in Figure 57.

- Hardware type: This field specifies the type of a hardware address. The value "1" represents an Ethernet address.

- Protocol type: This field specifies the type of the protocol address to be mapped. The hexadecimal value "0x0800" represents an IP address.
- 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

Figure 58 ARP address resolution process



Suppose that Host A and Host B are on the same subnet and that Host A sends a message to Host B. The resolution process is as follows:

1. Host A looks in its ARP mapping table to see whether there is an ARP entry for Host B. If Host A finds it, 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 source IP address and source MAC address are respectively the IP address and MAC address of Host A and the destination IP address and MAC address are respectively the IP address of Host B and an all-zero MAC address. Because the ARP request is sent in broadcast mode, all hosts on this subnet can receive the request, but only the requested host (namely, Host B) will process 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 into its ARP mapping 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 into its ARP mapping table for subsequent packet forwarding. Meanwhile, Host A encapsulates the IP packet and sends it out.

When Host A and Host B are not on the same subnet, a gateway helps finish ARP address resolution.

ARP Mapping Table

After obtaining the destination MAC address, the device adds the IP-to-MAC mapping into its own ARP mapping table. This mapping is used for forwarding packets with the same destination in future.

An ARP mapping table contains ARP entries, which fall into two categories: dynamic and static.

1 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.

2 A static ARP entry is manually configured and maintained. It cannot get aged or be overwritten by a dynamic ARP entry. It can be permanent or non-permanent.

- A permanent static ARP entry can be directly used to forward data. When configuring a permanent static ARP entry, you must configure a VLAN and outbound interface for the entry besides the IP address and MAC address.
- A non-permanent static ARP entry cannot be directly used for forwarding data. When configuring a non-permanent static ARP entry, you only need to configure the IP address and MAC address. When forwarding IP packets, the device sends an ARP request. If the source IP and MAC addresses in the received ARP reply are the same as the configured IP and MAC addresses, the device adds the interface receiving the ARP reply into the static ARP entry. Now the entry can be used for forwarding IP packets.

Usually ARP dynamically implements and automatically seeks mappings from IP addresses to MAC addresses, without manual intervention.

Configuring ARP

Configuring a Static ARP Entry

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-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:

- *A static ARP entry is effective when the Ethernet switch 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.*
- *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.*

Configuring the Maximum Number of ARP Entries a VLAN Interface Can Learn

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 vlan-interface** *interface-number*	-
Set the maximum number of dynamic ARP entries that the interface can learn	**arp max-learning-num** *number*	Optional 4096 by default

Setting Aging Time for Dynamic ARP Entries

After dynamic ARP entries expire, the system will delete them from the ARP mapping table. You can adjust the aging time for dynamic ARP entries according to the actual network condition.

Follow these steps to set aging time for dynamic ARP entries:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Set aging time for dynamic ARP entries	**arp timer aging** *aging-time*	Optional 20 minutes by default

Enabling the ARP Entry Check

ARP entry check disables the device from learning multicast MAC addresses.

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 Enabled by default. That is, the device does not learn multicast MAC addresses.

Enabling the Support for ARP Requests from a Natural Network

When learning MAC addresses, if the device finds that the source IP address of an ARP packet and the IP address of the inbound interface are not on the same subnet, the device will further judge whether these two IP addresses are on the same natural network.

Suppose that the IP address of Vlan-interface10 is 10.10.10.5/24 and that this interface receives an ARP packet from 10.11.11.1. Because these two IP addresses are not on the same subnet, Vlan-interface10 cannot process the packet. With this feature enabled, the device will make judgment on natural network basis. Because the IP address of Vlan-interface10 is a Class A address and its default mask length is 8, these two IP addresses are on the same natural network. In this way, Vlan-interface10 can learn the MAC address of the source IP address 10.11.11.1.

Follow these steps to enable the support for ARP requests from a natural network:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the support for ARP requests from a natural network	**naturemask-arp enable**	Required Disabled by default

ARP Configuration Examples

Network requirement

- Disable ARP entry check.
- Set the aging time for dynamic ARP entries to 10 minutes.
- Enable the support for ARP requests from a natural network.
- 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, the MAC address being 00e0-fc01-0000, and the outbound interface being Ethernet1/1/1 of VLAN 10.

Configuration procedure

```
<Sysname> system-view
[Sysname] undo arp check enable
[Sysname] arp timer aging 10
[Sysname] naturemask-arp enable
[Sysname] vlan 10
[Sysname-vlan10] quit
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] port access vlan 10
[Sysname-Ethernet1/1/1] 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 00e0-fc01-0000 10 ethernet1/1/1
```

Configuring Gratuitous ARP

Introduction to Gratuitous ARP

A gratuitous ARP packet is a special ARP packet, in which the source IP address and destination IP address are both the IP address of the sender, the source MAC address is the MAC address of the sender, and the destination MAC address is a broadcast address.

A device can implement 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 can add the information carried in the packet to its own dynamic ARP entry 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	**gratuitous-arp-sending enable**	Optional Disabled by default
Enable the gratuitous ARP packet learning function	**gratuitous-arp-learning enable**	Required Disabled by default

Configuring ARP Source Suppression

Introduction to ARP Source Suppression

If hosts on a network attack the device by sending large amounts of IP packets whose IP addresses cannot be resolved, the following consequences will be resulted in:

- The device sends large amounts of ARP request messages to the destination subnet, which increases the load of the destination subnet.
- The device continuously resolves destination IP addresses, which increase the load of the CPU.

To protect a device against this kind of attack, Switch 8800s provide for the ARP source suppression function. With the function enabled, whenever the number of packets with unresolvable IP addresses that a host sends to the device within five seconds exceeds the specified threshold, the device drops all subsequent packets with the same source IP address in another five coming seconds. This helps in protecting the device against the attack.

Configuring 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 seconds	**arp source-suppression limit** *limit-value*	Optional 10 by default

Configuring ARP Defense against IP Packet Attack

Introduction to ARP Defense against IP Packet Attack

In forwarding an IPv4 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. When large amounts of IP packets for which ARP cannot resolve the IP addresses of the next hops arrive at a device, the software on the device will be called again and again and the CPU of the device will be overburdened. This is called IP packet attack.

To protect a device against IP packet attack, you can configure the ARP defense against IP packet attack function. After receiving an IP packet with the IP address of the next hop unreachable (an IP packet that ARP cannot resolve the MAC address of the next hop), a device with this function creates a black hole route immediately and the forwarding chip simply drops all packets to the address. Note that a black hole route can get aged, in which case a subsequent IP packet with the same next hop triggers the above process. This protects the device against the IP packet attack efficiently, reducing the load of the CPU.

Enabling ARP Defense against IP Packet Attack

The ARP defense against IP packet attack function works for forwarded packets and those originated by the device.

Follow these steps to configure ARP defense against IP packet attack:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable ARP defense against IP packet attack	**arp resolving-route enable**	Optional Enabled by default

Displaying and Maintaining ARP

To do...	Use the command...	Remarks
Display the ARP entries in the ARP mapping table	**display arp** { { **all** | **dynamic** | **static** } [**slot** *slot-id*] | **vlan** *vlan-id* | **interface** *interface-type interface-number* } [[**verbose**] [**|** { **begin** | **exclude** | **include** } *text*] | **count**]	Available in any view
Display the ARP entries for a specified IP address	**display arp** *ip-address* [**slot** *slot-id*] [**verbose**] [**|** { **begin** | **exclude** | **include** } *text*]	Available in any view
Display the ARP entries for a specified VPN instance	**display arp vpn-instance** *vpn-instance-name* [**|** { **begin** | **exclude** | **include** } *text* | **count**]	Available in any view
Display the aging time for dynamic ARP entries	**display arp timer aging**	Available in any view

To do...	Use the command...	Remarks
Display the configuration information of ARP source suppression	**display arp source-suppression**	Available in any view
Clear ARP entries from the ARP mapping table	**reset arp** { **all** | **dynamic** | **static** | **slot** *slot-id* | **interface** *interface-type interface-number* }	Available in user view

21 PROXY ARP CONFIGURATION

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

- "Proxy ARP Overview" on page 201
- "Enabling Proxy ARP" on page 201
- "Displaying and Maintaining Proxy ARP" on page 202

Proxy ARP Overview

For an ARP request of a host on a network to be forwarded to an interface that is on the same network but isolated at Layer 2 or a host on another network, the device connecting the two physical or virtual networks must be able to respond to the request. This is achieved by proxy ARP.

Proxy ARP implements Layer 3 communication between interfaces isolated at Layer 2 or located on different networks.

Proxy ARP involves proxy ARP and local proxy ARP.

In one of the following cases, you need to enable the local proxy ARP:

- Devices connected to different isolated layer 2 ports in the same VLAN need to implement layer 3 communication.
- With the super VLAN function enabled, devices in different sub VLANs need to implement layer 3 communication.
- With the isolate-user-vlan function enabled, devices in different second VLANs need to implement layer 3 communication.

Enabling Proxy ARP

Follow these steps to enable proxy ARP or enable local proxy ARP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface vlan-interface**-*type interface-number*	Required
Enable proxy ARP	**proxy-arp enable**	Required Disabled by default
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** *interface-type* *interface-number*]	Available in any view
Display whether local proxy ARP is enabled	**display local-proxy-arp** [**interface** *interface-type* *interface-number*]	Available in any view

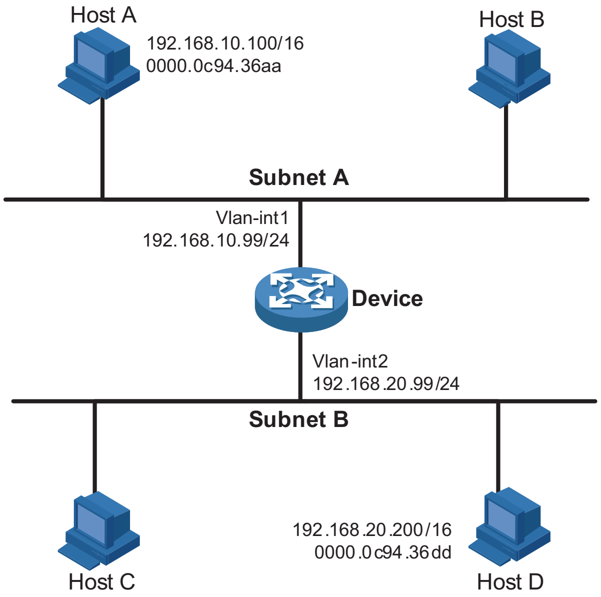
Proxy ARP Configuration Example

Network requirement

Host A belongs to VLAN 1, and Host D belongs to VLAN 2. Configure proxy ARP on the device to enable the communication between the two hosts.

Network diagram

Figure 59 Network diagram for proxy ARP



Configuration procedure

# Configure Proxy ARP on the device to enable the communication between Host A and Host D.

```
<Sysname> system-view
[Sysname] vlan 1
[Sysname-vlan1] vlan 2
[Sysname-vlan2] quit
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] ip address 192.168.10.99 255.255.255.0
[Sysname-Vlan-interface1] proxy-arp enable
[Sysname-Vlan-interface1] quit
[Sysname] interface vlan-interface 2
[Sysname-Vlan-interface2] ip address 192.168.20.99 255.255.255.0
[Sysname-Vlan-interface2] proxy-arp enable
[Sysname-Vlan-interface2] quit
```

For the local proxy ARP configuration example, refer to "Super VLAN Configuration" on page 167.

22 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" on page 205
- "Configuring IP Addresses" on page 207
- "Displaying IP Addressing Configuration" on page 210

IP Addressing Overview

This section covers these topics:

- "IP Address Classes" on page 205
- "Special Case IP Addresses" on page 206
- "Subnetting and Masking" on page 206
- "IP Unnumbered" on page 207

IP Address Classes

IP addressing uses a 32-bit address to identify each host on a network. 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: First several bits of the IP address defining a network, also known as class bits.
- Host-id: Identifies a host on a network.

For administration sake, IP addresses are divided into five classes. Which class an IP address belongs to depends on the first one to four bits of the net-id, as shown in the following figure (in which the blue parts represent the address class).

Figure 60 IP address classes

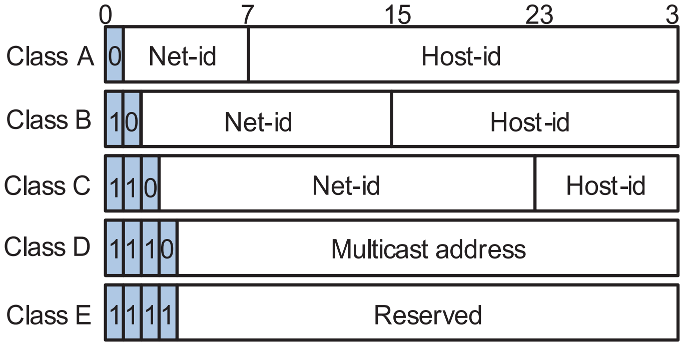


Table 18 describes the address ranges of these five classes. Currently, the first three classes of IP addresses are used in quantity.

Table 18 IP address classes and ranges

Class	Address range	Description
A	0.0.0.0 to 127.255.255.255	This address is used by a host at bootstrap when it does not know its IP address. This address is never a valid destination address. Addresses in the format of 127.X.Y.Z are reserved for the loopback test purpose. 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 address
E	240.0.0.0 to 247.255.255.255	Reserved address

Special Case 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

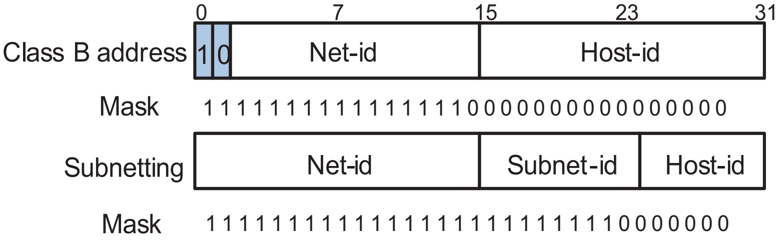
Subnetting was developed to address the risk of IP address exhaustion resulting from fast expansion of the Internet. The idea is to break a network down into smaller networks called subnets by using some bits of the host-id to create a subnet-id. To identify the boundary between the host-id and the combination of net-id and subnet-id, masking is used. (When subnetting is not adopted, a mask identifies the boundary between the host-id and the host-id.)

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.

Subnetting is valid with a single network. All these subnetworks appear as one. As subnetting adds an additional level, subnet-id, to the two-level hierarchy with IP addressing, IP routing now involves three steps: delivery to the site, delivery to the subnet, and delivery to the host.

Figure 61 shows how a Class B network is subnetted.

Figure 61 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 of 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.

IP Unnumbered

Logically, to enable IP on an interface, you must assign this interface a unique IP address. Yet, you can borrow an IP address already configured on one of other interfaces on your device instead. This is called IP unnumbered and the interface borrowing the IP address is called IP unnumbered interface.

You may need to use IP unnumbered to save IP addresses either when available IP addresses are inadequate or when an interface is brought up but for occasional use.

Configuring IP Addresses

Besides directly assigning an IP address to an interface, you may configure the interface to obtain one through DHCP; however, these two methods are mutually exclusive. If you change the way an interface obtains an IP address, from manual assignment to DHCP for example, the IP address obtained through DHCP will overwrite the previous one manually assigned.

This chapter only covers how to assign an IP address manually. For IP address assignment through DHCP, refer to "DHCP Address Allocation" on page 717.

This section includes:

- "Assigning an IP Address to an Interface" on page 208
- "IP Addressing Configuration Example" on page 208

Assigning an IP Address to an Interface

You may assign an interface multiple IP addresses, one primary and multiple secondaries, to connect multiple logical subnets on the same physical subnet.

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 previous one if there is any.*
- *You cannot assign a secondary IP address when the interface is configured to obtain an IP address through DHCP or to borrow one through IP unnumbered.*

IP Addressing Configuration Example

Network requirements

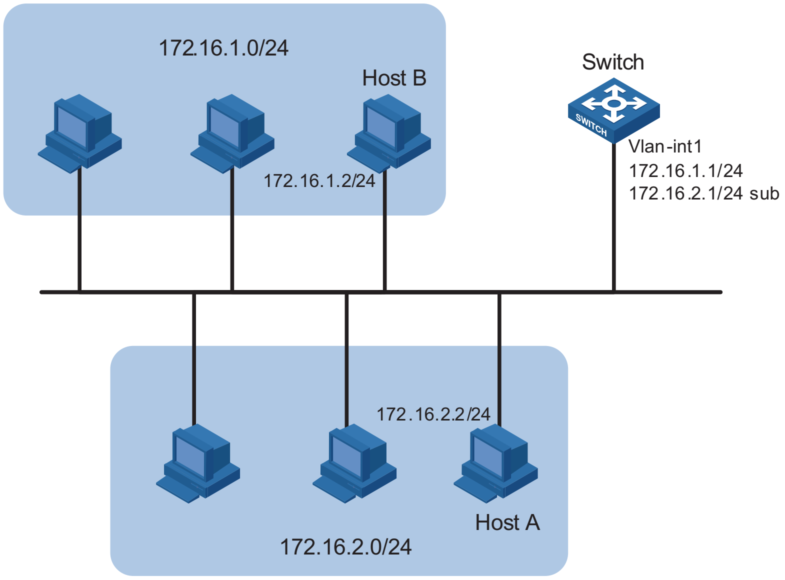
As shown in Figure 62, the Vlan-interface1 on a switch is connected to a LAN comprising two segments: 172.16.1.0/24 and 172.16.2.0/24.

To enable the switch to communicate with the two network segments respectively and the PCs on the LAN can communicate with each other, do the following:

- Assign two IP addresses to Vlan-interface1 on the switch.
- Set the switch as the gateway on all PCs.

Network diagram

Figure 62 Network diagram for IP addressing configuration



Configuration procedure

# Assign a primary IP address and a secondary IP address to Vlan-interface1.

```
<Sysname> system-view
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] ip address 172.16.1.1 255.255.255.0
[Sysname-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 the subnet 172.16.1.0/24, and to 172.16.2.1 on the PCs attached to the subnet 172.16.2.0/24. The configuration method may vary with operating systems, so you need to refer to the operating system configuration manual of the PC.

# Ping a host on the subnet 172.16.1.0/24 from the switch to verify the configuration.

```
<Sysname> 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
```

# Ping a host on the subnet 172.16.2.0/24 from the switch to verify the configuration.

```
<Sysname> 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
```

# Verify that the hosts on the subnets 172.16.1.0/24 and 172.16.2.0/24 can communicate with each other.

Displaying IP Addressing Configuration

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*]	Available in any view

23 IPv6 Basics Configuration

When configuring IPv6 basics, go to these sections for information you are interested in:

- "IPv6 Overview" on page 211
- "Configuring Basic IPv6 Functions" on page 221
- "Configuring IPv6 NDP" on page 222
- "Configuring PMTU Discovery" on page 226
- "Configuring IPv6 TCP Properties" on page 227
- "Configuring IPv6 FIB-Based Forwarding" on page 228
- "Configuring Capacity and Update Period of Token Bucket" on page 228
- "Configuring IPv6 DNS" on page 229
- "Displaying and Maintaining IPv6 Basics Configuration" on page 230
- "IPv6 Configuration Examples" on page 231
- "Troubleshooting IPv6 Basics Configuration" on page 233

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

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:

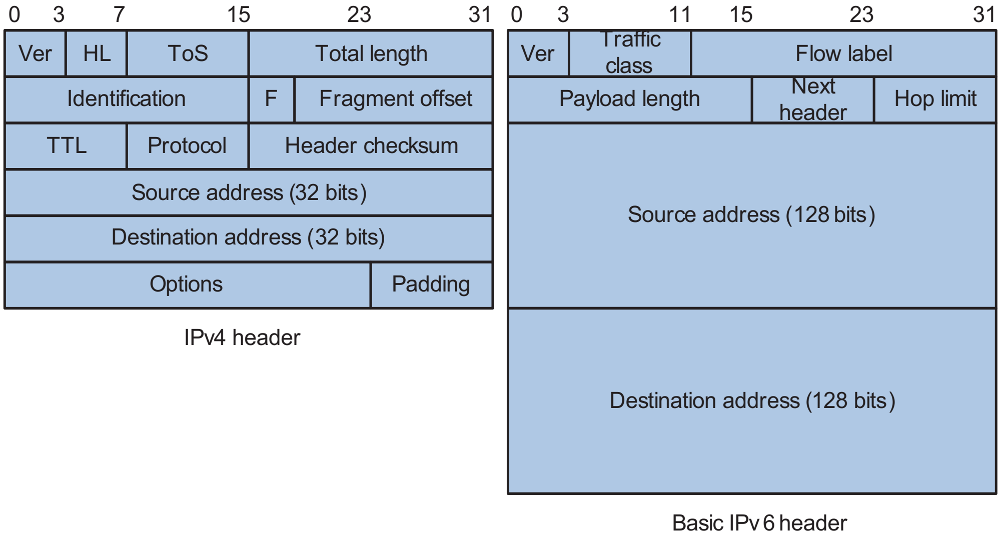
- "IPv6 Features" on page 211
- "Introduction to IPv6 Address" on page 213
- "Introduction to IPv6 Neighbor Discovery Protocol" on page 215
- "IPv6 PMTU Discovery" on page 219
- "Introduction to IPv6 DNS" on page 220
- "Protocols and Standards" on page 220

IPv6 Features

Header format simplification

IPv6 cuts down some IPv4 header fields or move them to IPv6 extension headers to reduce the load of basic IPv6 headers, thus making IPv6 packet handling simple and improving the forwarding efficiency. Although the IPv6 address size is four times that of IPv4 addresses, the size of basic IPv6 headers is 40 bytes, and is only twice that of IPv4 headers (excluding the Options field).

Figure 63 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 completely 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 source occupied by the IPv6 routing table by means of route aggregation.

Automatic address configuration

To simplify the 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, DHCP server).
- Stateless address configuration means that a host automatically configures an IPv6 address and related information on basis of its own link-layer address and the prefix information advertised by the router.

In addition, a host can generate a link-local address on basis of its own link-layer address and the default prefix (FE80::/64) to communicate with other hosts on the 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 improves the interoperability between different IPv6 applications.

QoS support

The Flow Label field in the IPv6 header allows the device to label packets in a flow and provide special handling for these packets.

Enhanced neighbor discovery mechanism

The IPv6 neighbor discovery protocol is a group of Internet control message protocol version 6 (ICMPv6) messages that manages the information exchange between neighbor nodes on the same link. The group of ICMPv6 messages takes the place of address resolution protocol (ARP) message, Internet control message protocol version 4 (ICMPv4) router discovery message, and ICMPv4 redirection message to provide a series of other functions.

Flexible extension headers

IPv6 cancels the Options field in IPv4 packets but introduces multiple extension headers. In this way, IPv6 enhances the flexibility greatly to provide scalability for IP while improving the handling efficiency. The Options field in IPv4 packets contains 40 bytes at most, while the size of IPv6 extension headers is restricted by that of IPv6 packets.

Introduction to IPv6 Address

IPv6 address format

An IPv6 address is represented as a series 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 which are separated by colons, 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 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 the double-colon :: option. For example, the above-mentioned address can be represented in the shortest format as 2001:0:130F::9C0:876A:130B.

CAUTION: *The double-colon :: option can be used only once in an IPv6 address. Otherwise, the device is unable to determine how many zeros 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 IPv6-address is an IPv6 address in any of the notations and prefix-length is a decimal number indicating how many bits from the utmost left of an IPv6 address are 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 nearest one, according to the routing protocols' measure of distance).

There are no broadcast addresses in IPv6. Their function is superseded by multicast addresses.

The type of an IPv6 address is designated by the first several bits called format prefix. Table 19 lists the mappings between address types and format prefixes.

Table 19 Mapping 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 forms of unicast address assignment in IPv6, including global unicast address, link-local address, and site-local address.

- The global unicast address, equivalent to an IPv4 public address, is provided for network service providers. The structure of such a type of address allows efficient route prefix aggregation to restrict the number of global routing entries.
- The link-local address is used for communication between link-local nodes in neighbor discovery and stateless autoconfiguration. Routers must not forward any packets with link-local source or destination addresses to other links.
- IPv6 unicast site-local addresses are similar to private IPv4 addresses. Routers must not forward any packets with site-local source or destination addresses outside of the site (equivalent to 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, but may not use it as a destination IPv6 address.

Multicast address

IPv6 multicast addresses listed in Table 20 are reserved for special purpose.

Table 20 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 addresses of neighbor nodes on the same link and is also used for duplicate address detection (DAD). Each IPv6 unicast or anycast address has one corresponding solicited-node address. The format of a solicited-node multicast address is as follows:

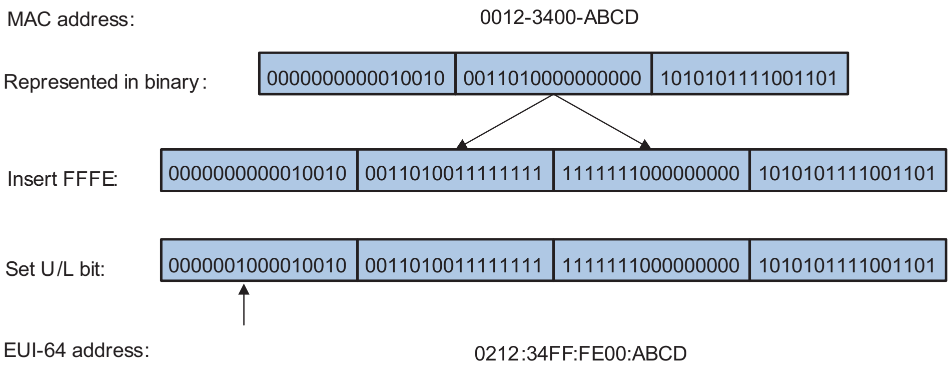
FF02:0:0:0:0:1:FFXX:XXXX

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

Interface identifiers in IPv6 unicast addresses are used to identify interfaces on a link and they are required to be unique on that link. Interface identifiers in IPv6 unicast addresses are currently required to be 64 bits long. An interface identifier in IEEE EUI-64 format is derived from the link-layer address of that interface. Interface identifiers in IPv6 addresses are 64 bits long, while MAC addresses are 48 bits long. Therefore, the hexadecimal number FFFE needs to be inserted in the middle of MAC addresses (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 64 Convert a MAC address into an EUI-64 address



Introduction to IPv6 Neighbor Discovery Protocol

IPv6 neighbor discovery protocol (NDP) uses five types of ICMPv6 messages to implement the following functions:

- "Address resolution" on page 217

- "Neighbor reachability detection" on page 218
- "Duplicate address detection" on page 218
- "Router/prefix discovery and address autoconfiguration" on page 218
- "Redirection" on page 219

Table 21 lists the types and functions of ICMPv6 messages used by the NDP.

Table 21 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 host 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 address prefix 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.
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.

Table 21 Types and functions of ICMPv6 messages

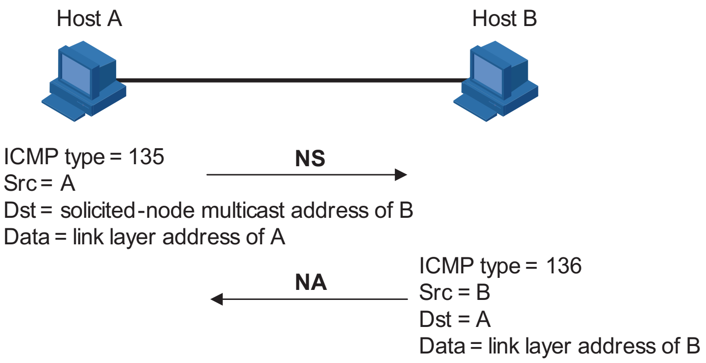
ICMPv6 message	Number	Function
Neighbor advertisement (NA) message	136	Used to respond to an NS message.
		When the link layer changes, the local node initiates an NA message to notify neighbor nodes of the node information change.
Router solicitation (RS) message	133	After started, a host sends an RS 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 address prefix 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 65 shows how node A acquires the link-layer address of node B.

Figure 65 Address resolution


Host A
Host B
ICMP type = 135
Src = A
Dst = solicited-node multicast address of B
Data = link layer address of A
NS
NA
ICMP type = 136
Src = B
Dst = A
Data = link layer address of B

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 an 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 corresponds to the solicited-node multicast address. If yes, node B unicasts an NA message containing its link-layer address.
3. Node A acquires the link-layer address of node B from the NA message. After that, node A and node B can communicate.

Neighbor reachability detection

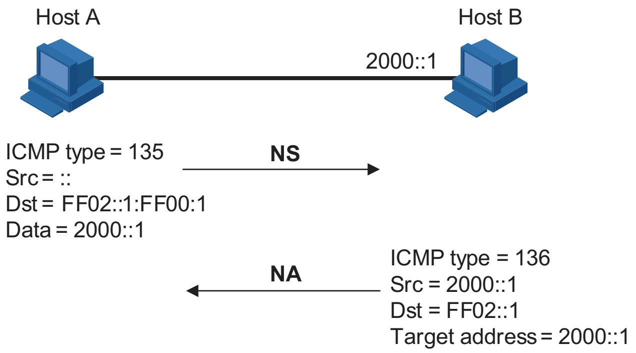
After node A acquires the link-layer address of its neighbor node B, node A can verify whether node B is reachable according to NS and NA messages.

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 other nodes (similar to the gratuitous ARP function of IPv4). DAD is accomplished through NS and NA messages. Figure 65 shows the DAD procedure.

Figure 66 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 host 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 host automatically configures 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 host 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 address prefix and flag bits. (The router also regularly sends an RA message.)
3. The host automatically configures an IPv6 address and other information for its interface according to the address prefix and other configuration parameters in the RA message.

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 will send an IPv6 ICMP redirect message when the following conditions are satisfied:

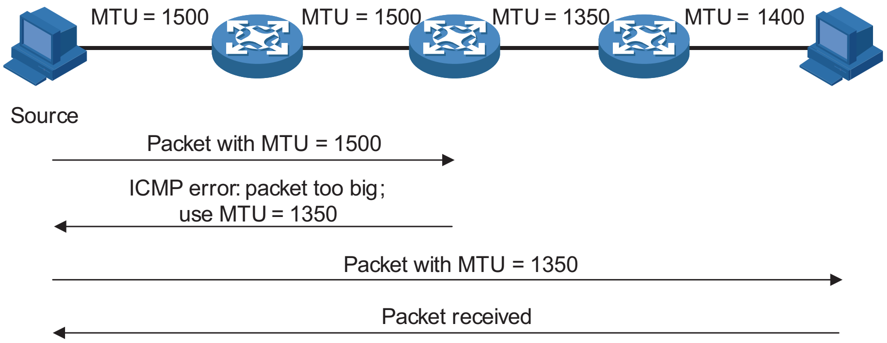
- The receiving interface is the forwarding interface.
- The selected route itself is not created or modified by an IPv6 ICMP redirect message.
- The selected route is not the default route.
- The forwarded IPv6 packet does not contain any routing header.

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 link MTU, the packet will be fragmented at the source end so as to reduce the processing pressure of the forwarding device and utilize network resources rationally.

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 67 shows the working procedure of the PMTU discovery.

Figure 67 Working procedure of the PMTU discovery



The working procedure of the PMTU discovery is as follows:

1. The source host uses its MTU to fragment packets and then sends them to the destination host.
2. If the MTU supported by the forwarding interface is less 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 fragment the packet again and then sends it.
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

In the IPv6 network, a domain name system (DNS) supporting IPv6 converts domain names into IPv6 addresses, instead of IPv4 addresses.

However, just like an IPv4 DNS, an IPv6 DNS also covers 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 an IPv4 DNS. For details, refer to *"DNS Overview" on page 745*.

Usually, the DNS server connecting IPv4 and IPv6 networks not only contain 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

Configuring Basic IPv6 Functions

Enabling the IPv6 Packet Forwarding Function

Before IPv6-related configurations, you must enable the IPv6 packet forwarding function. Otherwise, an interface cannot forward IPv6 packets even if an IPv6 address is configured, resulting in communication failures in the IPv6 network.

Follow these steps to enable the IPv6 packet forwarding function:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the IPv6 packet forwarding function	**ipv6**	Required Disabled by default.

Configuring an IPv6 Unicast Address

IPv6 site-local addresses and global unicast addresses can be configured in either of the following ways:

- EUI-64 format: When the EUI-64 format is adopted to form IPv6 addresses, 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 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::/64) 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 link-local address:

<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><b>system-view</b></td><td>-</td></tr>
<tr><td colspan="2">Enter interface view</td><td><b>interface</b> <i>interface-type interface-number</i></td><td>-</td></tr>
<tr><td rowspan="2">Configure an IPv6 global unicast address or site-local address</td><td>Manually assign an IPv6 address</td><td><b>ipv6 address</b> { <i>ipv6-address prefix-length</i> | <i>ipv6-address</i><b>/</b><i>prefix-length</i> }</td><td rowspan="2">Use either command
By default, no site-local address or global unicast address is configured for an interface.</td></tr>
<tr><td>Adopt the EUI-64 format to form an IPv6 address</td><td><b>ipv6 address</b> <i>ipv6-address</i><b>/</b><i>prefix-length</i> <b>eui-64</b></td></tr>
</table>

To do...		Use the command...	Remarks
Configure an IPv6 link-local address	Automatically generate a link-local address	**ipv6 address auto link-local**	Optional By default, after an IPv6 site-local address or global unicast address is configured for an interface, a link-local address will be generated automatically.
	Manually assign a link-local address for an interface	**ipv6 address** *ipv6-address* **link-local**	

- *After an IPv6 site-local address or global unicast address is configured for an interface, a link-local address will be 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 link-local address takes effect. If the manually assigned link-local address is removed, the automatically generated link-local address takes effect.*
- *The manual assignment takes precedence over the automatic generation. That is, if you first adopt the automatic generation and then the manual assignment, the manually assigned link-local address will overwrite the automatically generated one. If you first adopt the manual assignment and then the 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. You must delete the manually assigned link-local address before adopting the automatic generation.*
- *You must carry out the* ***ipv6 address auto link-local*** *command before the* ***undo ipv6 address auto link-local*** *command. However, if an IPv6 site-local address or 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 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 statically through manual configuration.

The device uniquely identifies a static neighbor entry according to the IPv6 address and the layer 3 interface ID. Currently, there are two configuration methods:

- Configure an IPv6 address and link-layer address for a layer 3 interface.
- Configure an IPv6 address and link-layer address for 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**	-

To do...	Use the command...	Remarks
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 for a VLAN interface.*

- *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 to configure a static neighbor entry, 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 an 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. Too large a neighbor table from which neighbor entries can be dynamically acquired may lead to the forwarding performance degradation of the device. Therefore, 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 on an interface	**ipv6 neighbors max-learning-num** *number*	Optional 1024 by default.

Configuring Parameters Related to an RA Message

You can configure whether the interface sends an RA message, 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 22 lists the configurable parameters in an RA message and their descriptions.

Table 22 Parameters in an RA message and their descriptions

Parameters	Description
Cur hop limit	When sending an IPv6 packet, a host uses the value of this parameter to fill the Cur Hop Limit field in IPv6 headers. Meanwhile, the value of this parameter is equal to the value of the Cur Hop Limit field in response messages of the device.
Prefix information options	After receiving the prefix information advertised by the device, the hosts on the same link can perform stateless autoconfiguration operations.
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. Otherwise, hosts use the stateless autoconfiguration to acquire IPv6 addresses, that is, hosts configure 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 (for example, DHCP server) to acquire information other than IPv6 addresses. 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 of hosts.
Retrans timer	If the device fails to receive a response message within the specified time after sending an NS message, the device will retransmit it.
Reachable time	After the neighbor reachability detection shows that a neighbor is reachable, the device considers the neighbor is reachable within the reachable time. If the device needs to send a packet to a neighbor after the reachable time expires, the device will again confirm whether the neighbor is reachable.

The values of the Retrans Timer field and the Reachable Time field configured for an interface are sent to hosts via RA messages. Furthermore, this interface sends NS messages at intervals of Retrans Timer and considers a neighbor reachable within the time of Reachable Time.

Follow these steps to configure parameters related to an RA message:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Configure the current hop limit	**ipv6 nd hop-limit** *value*	Optional 64 by default.
Enter interface view	**interface** *interface-type interface-number*	-
Disable the RA message suppression	**undo ipv6 nd ra halt**	Optional 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 intervals of a random value 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 options in RA messages	**ipv6 nd ra prefix** { *ipv6-address prefix-length* | *ipv6-address***/***prefix-length* } *valid-lifetime preferred-lifetime* [**no-autoconfig** [**off-link**]*	Optional By default, no prefix information is configured in 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 1,800 seconds by default.
Set the retrans timer	**ipv6 nd ns retrans-timer** *value*	Optional By default, the local interface sends NS messages at intervals of 1,000 milliseconds and the Retrans Timer field in RA messages sent by the local interface is equal to 0.
Set the reachable time	**ipv6 nd nud reachable-time** *value*	Optional By default, the neighbor reachable time on the local interface is 30,000 milliseconds and 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 Number of Attempts to Send an NS Message for DAD

An interface sends a neighbor solicitation (NS) message for DAD 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 attempts to send an NS message reaches the maximum, the acquired address is considered available.

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 the Interface MTU

IPv6 routing devices do not support packet fragmentation. After an IPv6 routing device receives an IPv6 packet, if the packet size is greater than the MTU of the forwarding interface, the device will discard the packet. Meanwhile, the device sends the MTU to the source host through an ICMPv6 packet - Packet Too Big message. The source host fragments the packet according to the MTU and resends it. To reduce the extra flow overhead resulting from packets being discarded, a proper interface MTU should be configured according to the actual networking environment.

Follow these steps to configure the interface MTU:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure the interface MTU	**ipv6 mtu** *mtu-size*	Optional

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 packets 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 PMTU

After the MTU of the path from the source host to the destination host is dynamically determined (refer to "IPv6 PMTU Discovery" on page 219), the source host sends subsequent packets to the destination host on basis of this MTU. After the aging time expires, the dynamically determined PMTU is removed and the source host re-determines an MTU to send packets through the PMTU mechanism.

The aging time is invalid for static PMTU.

Follow these steps to configure the aging time for PMTU:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure aging time for PMTU	**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 other packets are received, the finwait timer is reset from the last received packet and the connection is terminated after the finwait timer expires.
- Size of the connection-oriented socket buffer.

Follow these steps to configure IPv6 TCP properties:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Set the finwait timer of IPv6 TCP packets	**tcp ipv6 timer fin-timeout** *wait-time*	Optional 675 seconds by default
Set the synwait timer of IPv6 TCP packets	**tcp ipv6 timer syn-timeout** *wait-time*	Optional 75 seconds by default
Set the size of the IPv6 TCP buffer	**tcp ipv6 window size**	Optional 8 KB by default

Configuring IPv6 FIB-Based Forwarding

With the caching function of IPv6 FIB enabled, the device searches the FIB cache when forwarding packets, thus reducing the time in searching IP packets and improving the forwarding efficiency.

In the load sharing mode of IPv6 FIB, the device can decide how to select an equal cost multi-path (ECMP) route to forward packets. Currently, two load sharing modes are supported:

- Load sharing based on the HASH algorithm: A certain algorithm based on the source IPv6 address and destination IPv6 address is adopted to select an ECMP route to forward packets.
- Load sharing based on polling: Each ECMP route is used in turn to forward packets.

Follow these steps to configure the IPv6 FIB-based forwarding:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable the caching function of IPv6 FIB		**ipv6 fibcache** { *slot-number* | **all** }	Required Disabled by default
Configure the IPv6 FIB load sharing mode	Configure the load sharing based on the HASH algorithm	**ipv6 fib-loadbalance-type hash-based**	Optional By default, the load sharing based on polling is adopted, that is, each ECMP route is used in turn to forward packets.
	Configure the load sharing based on polling	**undo ipv6 fib-loadbalance-type hash-based**	

Configuring Capacity and Update Period of Token Bucket

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 period of the token bucket, namely, the interval for updating the number of tokens in the token bucket to 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 1. If the number of ICMPv6 error packets successively sent exceeds the capacity of the token bucket, subsequent ICMPv6 error packets cannot be sent out until the number of tokens in the token bucket is updated and new tokens are added to the bucket.

Follow these steps to configure the capacity and update period of the token bucket:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Configure the capacity and update period 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 period is 100 milliseconds. That is, at most 10 IPv6 ICMP error packets can be sent within these 100 milliseconds. The update period "0" indicates that the number of ICMPv6 error packets sent is not restricted.

Configuring IPv6 DNS

Configuring Static IPv6 DNS

You can establish the mapping between host name and IPv6 address through the following configuration. When applying such applications as Telnet, you can directly use 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 a host name and the corresponding IPv6 address:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure a host name and the corresponding IPv6 address	**ipv6 host** *hostname ipv6-address*	Required

Configuring Dynamic IPv6 DNS

If you want to use the dynamic domain name function, you can use the following command to enable the dynamic domain name resolution function. In addition, you should 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 some fields 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 DNS:

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 a value for *interface-type* and *interface-number*.

To do...	Use the command...	Remarks
Configure the DNS suffix.	**dns domain** *domain-name*	Required By default, no DN suffix is configured, that is, the domain name is resolved according to the input information.

*The **dns resolve** and **dns domain** commands are the same as those of IPv4 DNS. For details about the commands, refer to the Switch 8800 Command Reference Guide.*

Displaying and Maintaining IPv6 Basics Configuration

To do...	Use the command...
Display DNS suffix information	**display dns domain** [**dynamic**]
Display IPv6 dynamic domain name cache information	**display dns ipv6 dynamic-host**
Display DNS server information	**display dns server** [**dynamic**]
Display the FIB entries	**display ipv6 fib** [*slot-number*] [*ipv6-address*]
Display the total number of routes in the FIB cache	**display ipv6 fibcache** *slot-number*
Display the mappings between host names and IPv6 addresses in the static DNS database.	**display ipv6 host**
Display the IPv6 information of an interface	**display ipv6 interface** [*interface-type interface-number* | **brief**]
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** } *text*]
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 information related to a specified socket	**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	**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**
Clear FIB entries from the cache	**reset ipv6 fibcache** { *slot-number* | **all** }
Clear IPv6 neighbor information	**reset ipv6 neighbors** { **all** | **dynamic** | **interface** *interface-type interface-number* | **slot** *slot-number* | **static** }
Clear the corresponding PMTU	**reset ipv6 pathmtu** { **all** | **static** | **dynamic**}

To do...	Use the command...
Clear the statistics of IPv6 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**

*The **display dns domain** and **display dns server** commands are the same as those of IPv4 DNS. For details about the commands, refer to the Switch 8800 Command Reference Guide.*

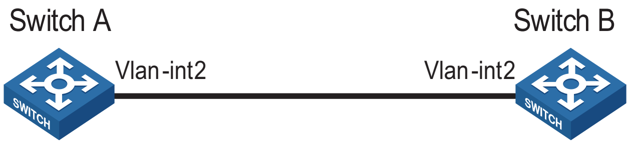
IPv6 Configuration Examples

Network requirements

Switch A and Switch B are directly connected through two Ethernet ports that belong to VLAN 2. Different types of IPv6 addresses are configured for the interface VLAN-interface 2 to verify the connectivity between two switches. The IPv6 prefix in the EUI-64 format is 2001::/64, the global unicast address of Switch A is 3001::1/64, and the global unicast address of Switch B is 3001::2/64.

Network diagram

Figure 68 Network diagram for IPv6 address configuration



Configuration procedure

- Configure Switch A

  # Enable the IPv6 packet forwarding function.

```
<SwitchA> system-view
[SwitchA] ipv6
```

  # Configure the interface VLAN-interface 2 to automatically generate a link-local address.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address auto link-local
```

  # Configure an EUI-64 address for the interface VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ipv6 address 2001::/64 eui-64
```

  # Configure a global unicast address for the interface VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ipv6 address 3001::1/64
```

- Configure Switch B

# Enable the IPv6 packet forwarding function.

```
<SwitchB> system-view
[SwitchB] ipv6
```

# Configure the interface VLAN-interface 2 to automatically generate a link-local address.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address auto link-local
```

# Configure an EUI-64 address for the interface VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ipv6 address 2001::/64 eui-64
```

# Configure a global unicast address for VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ipv6 address 3001::2/64
```

Verification

# Display the IPv6 information of the interface on Switch A.

```
[SwitchA-Vlan-interface2] display ipv6 interface vlan-interface 2
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE49:8048
  Global unicast address(es):
    2001::20F:E2FF:FE49:8048, subnet is 2001::/64
    3001::1, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:1
    FF02::1:FF49:8048
    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
```

# Display the IPv6 information of the interface on Switch B.

```
[SwitchB-Vlan-interface2] display ipv6 interface vlan-interface 2
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1
  Global unicast address(es):
    2001::20F:E2FF:FE00:1, subnet is 2001::/64
    3001::2, subnet is 3001::/64
  Joined group address(es):
    FF02::1:FF00:2
    FF02::1:FF00:1
    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
```

# From Switch A, ping the link-local address, EUI-64 address, and global unicast address of Switch B, respectively. If the configurations are correct, the three types of IPv6 addresses above can be pinged.

CAUTION: *When you ping a link-local address, you should use the "-i" parameter to specify an interface for the link-local address.*

```
[SwitchA-Vlan-interface2] ping ipv6 FE80::20F:E2FF:FE00:1 -i vlan-interface2
  PING FE80::20F:E2FF:FE00:1 : 56  data bytes, press CTRL_C to break
    Reply from FE80::20F:E2FF:FE00:1
    bytes=56 Sequence=1 hop limit=255  time = 80 ms
    Reply from FE80::20F:E2FF:FE00:1
    bytes=56 Sequence=2 hop limit=255  time = 60 ms
    Reply from FE80::20F:E2FF:FE00:1
    bytes=56 Sequence=3 hop limit=255  time = 60 ms
    Reply from FE80::20F:E2FF:FE00:1
    bytes=56 Sequence=4 hop limit=255  time = 70 ms
    Reply from FE80::20F:E2FF:FE00:1
    bytes=56 Sequence=5 hop limit=255  time = 60 ms

  --- FE80::20F:E2FF:FE00:1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 60/66/80 ms
[SwitchA-Vlan-interface2] ping ipv6 2001::20F:E2FF:FE00:1
  PING 2001::20F:E2FF:FE00:1 : 56  data bytes, press CTRL_C to break
    Reply from 2001::20F:E2FF:FE00:1
    bytes=56 Sequence=1 hop limit=255  time = 40 ms
    Reply from 2001::20F:E2FF:FE00:1
    bytes=56 Sequence=2 hop limit=255  time = 70 ms
    Reply from 2001::20F:E2FF:FE00:1
    bytes=56 Sequence=3 hop limit=255  time = 60 ms
    Reply from 2001::20F:E2FF:FE00:1
    bytes=56 Sequence=4 hop limit=255  time = 60 ms
    Reply from 2001::20F:E2FF:FE00:1
    bytes=56 Sequence=5 hop limit=255  time = 60 ms

  --- 2001::20F:E2FF:FE00:1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 40/58/70 ms

[SwitchA-Vlan-interface2] ping ipv6 3001::2
  PING 3001::2 : 56  data bytes, press CTRL_C to break
    Reply from 3001::2
    bytes=56 Sequence=1 hop limit=255  time = 50 ms
    Reply from 3001::2
    bytes=56 Sequence=2 hop limit=255  time = 60 ms
    Reply from 3001::2
    bytes=56 Sequence=3 hop limit=255  time = 60 ms
    Reply from 3001::2
    bytes=56 Sequence=4 hop limit=255  time = 70 ms
    Reply from 3001::2
    bytes=56 Sequence=5 hop limit=255  time = 60 ms

  --- 3001::2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 50/60/70 ms
```

Troubleshooting IPv6 Basics Configuration

Symptom:

The peer IPv6 address cannot be pinged.

Solution:

- Carry out the **display current-configuration** command in any view or the **display this** command in system view to check that the IPv6 packet forwarding function is enabled.
- Carry out the **display ipv6 interface** command in any view to check that the IPv6 address of the interface is correct and that the interface is up.
- Carry out the **debugging ipv6 packet** command in user view to enable the debugging for IPv6 packets and make judgment according to the debugging information.

24 IP PERFORMANCE CONFIGURATION

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

IP Performance Overview

In some network environments, you need to adjust the IP parameters to achieve best network performance. IP performance configuration includes:

- Enabling forwarding of directed broadcasts
- Configuring TCP timers
- Configuring the TCP buffer size
- Configuring TCP MSS for the interface
- Enabling ICMP error packet sending

Enabling Forwarding of Directed Broadcasts to a Directly Connected Network

Directed broadcasts refer to broadcast packets sent to a specific network. In the destination IP address of a directed broadcast, the network ID is a network-specific number and the host ID is all ones.

Enabling the device to receive and forward directed broadcasts to a directly connected network will give hackers an opportunity to attack the network. Therefore, this feature is disabled by default.

When this feature is required in some network applications, you can configure the device to forward directed broadcasts in system view or interface view. If you disable this feature in system view, the switch discards directed broadcasts directly; otherwise, the system determines whether to discard directed broadcasts according to the interface configuration.

Switch 8800s can still receive broadcasts from a designated UDP port even if they are disabled from receiving directed broadcasts.

Enabling Forwarding of Directed Broadcasts to a Directly Connected Network (in System View)

Follow these steps to enable the device to forward directed broadcasts:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the device to forward directed broadcasts	**ip forward-broadcast**	Required By default, the device is disabled from forwarding directed broadcasts.

Enabling Forwarding of Directed Broadcasts to a Directly Connected Network (in Interface View)

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.

- *You can reference an ACL to forward only directed broadcasts permitted by the ACL.*
- *If you execute the* ***ip forward-broadcast acl*** *command on an interface repeatedly, the last execution overwrites the previous one. If the command executed last time does not include the* ***acl*** *acl-number, the ACL configured previously will be removed.*

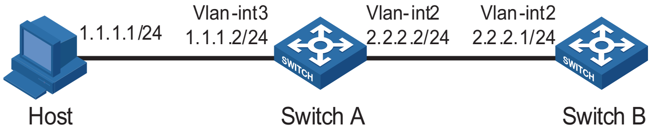
Configuration Example

Network requirements

As shown in Figure 69, the host's interface and Vlan-interface3 of Switch A are on the same network segment (1.1.1.0/24). Vlan-interface2 of Switch A and Vlan-interface2 of Switch B are on another network segment (2.2.2.0/24). The default gateway of the host is Vlan-interface3 (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.

Network diagram

Figure 69 Figure 25-1 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-interface3 and Vlan-interface2.

```
[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-interface2 to forward directed broadcasts.

```
[SwitchA-Vlan-interface2] ip forward-broadcast
```

- l 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-interface2.

```
[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-interface2 of Switch A on the host, the ping packets can be received by Vlan-interface2 of Switch B. However, if you execute the **undo ip forward-broadcast** command, the ping packets cannot be received by Vlan-interface2 of Switch B.

Configuring TCP Attributes

TCP attributes that can be configured include:

synwait timer: When sending a SYN packet, TCP starts the synwait timer. If no response packets are received within the synwait timer timeout, the TCP connection is not successfully created.

finwait timer: When the TCP connection is in FIN_WAIT_2 state, finwait timer will be started. If no FIN packets are received within the timer timeout, the TCP connection will be terminated. If FIN packets are received, the TCP connection state changes to TIME_WAIT. If non-FIN packets are received, the system restarts the timer from receiving the last non-FIN packet. The connection is terminated 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**	-

To do...	Use the command...	Remarks
Configure TCP synwait timer's timeout value	**tcp timer syn-timeout** *time-value*	Optional By default, the timeout value is 75 seconds.
Configure TCP finwait timer's timeout value	**tcp timer fin-timeout** *time-value*	Optional By default, the timeout value is 675 seconds.
Configure the size of TCP receive/send buffer	**tcp window** *window-size*	Optional By default, the buffer is 8 kilobytes.

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 TCP MSS for the Interface

The TCP maximum segment size (MSS) on an interface determines whether TCP packets need to be fragmented when forwarded. If the size of a packet is smaller than the TCP MSS, the packet is unnecessarily to be fragmented; otherwise, it will be fragmented according to the TCP MSS.

Follow these steps to configure TCP MSS for the interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure TCP MSS for the interface	**tcp mss** *value*	Required 1,460 bytes by default.

Configuring ICMP Error Packet Sending

Sending error packets is a major function of Internet control message protocol (ICMP). In case of network abnormalities, ICMP packets are usually sent by the network or transport layer protocols to notify corresponding devices, thus facilitating control and management.

Advantage 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 one default route to the default gateway in its routing table after startup. If certain conditions are satisfied, the default gateway will send ICMP redirect packets to the source host and notify it to reselect a correct next hop router to send the subsequent packets.

Switch 8800s will send ICMP redirect packets to the source host under the following conditions:

- 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 switch
- There is no source route option in the packet

When performing hardware forwarding, Switch 8800s will not forward ICMP redirect packets even if the above conditions are satisfied.

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.

Switch 8800s will send ICMP timeout packets under the following conditions:

- If the switch 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 switch receives the first fragment of an IP datagram whose destination is the device itself, it will start a timer. If the timer times out before all the fragments of the datagram are received, the switch 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.

Switch 8800s will send an ICMP destination unreachable error packet under the following conditions:

- 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 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.

When performing hardware forwarding, Switch 8800s will not forward ICMP destination unreachable packets even if the above conditions are satisfied.

Disadvantage of sending ICMP error packets

Although sending ICMP error packets facilitate control and management, it still has the following disadvantages:

- Sending a lot of ICMP packets will increase network traffic.
- If the switch receives a lot of malicious packets that cause it to send ICMP error packets, the 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 switch from sending ICMP error packets.

Follow these steps to disable sending ICMP error packets:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Disable sending ICMP redirect packets	**undo ip redirects**	Required Enabled by default
Disable sending ICMP timeout packets	**undo ip ttl-expires**	Required Enabled by default.
Disable sending ICMP destination unreachable packets	**undo ip unreachables**	Required Enabled by default

- *The switch stops sending "network unreachable" and "source route failure" ICMP error packets after sending ICMP destination unreachable packets is disabled. However, other destination unreachable packets can be sent normally.*
- *The switch 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

To do...	Use the command...
Display current TCP connection state	**display tcp status**
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*]

To do...	Use the command...
Display FIB forward information	**display fib** [| { **begin** | **include** | **exclude** } *text* | **acl** *acl-number* | **ip-prefix** *ip-prefix-name*]
Display FIB forward information matching the specified destination IP address	**display fib** *ip-address1* [{ *mask1* | *mask-length1* } [*ip-address2* { *mask2* | *mask-length2* } | **longer**] | **longer**]
Display statistics about the FIB items	**display fib statistics**
Clear statistics of IP packets	**reset ip statistics** [**slot** *slot-number*]
Clear statistics of TCP connections	**reset tcp statistics**
Clear statistics of UDP flows	**reset udp statistics**

25 ROUTING POLICY CONFIGURATION

A routing policy is used on a router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed.

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

- "Introduction to Routing Policy" on page 243
- "Routing Policy Configuration Task List" on page 245
- "Defining Filtering Lists" on page 245
- "Configuring a Routing Policy" on page 247
- "Displaying and Maintaining the Routing Policy" on page 251
- "Routing Policy Configuration Examples" on page 252
- "Troubleshooting Routing Policy Configuration" on page 256

The term "router" refers to a router in a generic sense or an Ethernet switch running routing protocols in this document.

Introduction to Routing Policy

Routing Policy

A routing policy is used on the router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed.

When distributing or receiving routing information, a router can apply a policy to filter routing information. For example, a router handles only routing information that matches some criteria of a routing policy; a routing protocol redistributes from another protocol only routes matching some criteria of a routing policy and modifies some attributes of these routes to satisfy its needs according to the routing policy.

To implement a routing policy, you need define a set of match criteria according to attributes in routing information, such as destination address, advertising router's address and so on. The match criteria can be set beforehand and then apply them to a routing policy for route distribution, reception and redistribution.

Filters

Routing protocols can use six filters: ACL, IP prefix list, AS path ACL, community list, extended community list and routing policy.

ACL

ACL involves IPv4 ACL and IPv6 ACL. When defining an ACL, you can specify IP addresses and prefixes to match destinations or next hops of routing information.

For ACL configuration, refer to *"ACL Overview" on page 801*.

IP prefix list

IP prefix list involves IPv4 and IPv6 prefix list.

IP prefix list plays a role similar to ACL, but it is more flexible than ACL and easier to understand. When an IP prefix list is applied to filtering routing information, its matching object is the destination address of routing information. Moreover, you can specify the **gateway** option to indicate that only routing information advertised by certain routers will be received. For **gateway** option information, refer to *"RIP Configuration" on page 269* and *"OSPF Configuration" on page 301.*

An IP prefix list is identified by name. Each IP prefix list can comprise multiple items, and each item, which is identified by an index number, can specify a matching range in network prefix format. The index number indicates the matching sequence of items in the IP prefix list.

During matching, the router compares the packet with the items in the ascending order. If one item is matched, the IP prefix list filter is passed, and the packet will not go to the next item.

AS-path list

AS path list is only applicable to BGP. There is an AS-path field in the BGP packet. An AS path list specifies matching conditions according to the AS-path field.

Community list

Community list only applies to BGP. The BGP packet contains a community attribute field to identify a community. A community list specifies matching conditions based on the community attribute.

Extended community list

Extended community list (extcommunity-list) applies to BGP only. It involves two attributes: Route-Target extcommunity for VPN, Source of Origin extcommunity. An extcommunity-list specifies matching conditions according to the two attributes.

The Source of Origin extcommunity attribute, which is the application in the source routing feature, is not supported currently.

Routing policy

A routing policy is used to match against some attributes in given routing information and modify the attributes of the information if match conditions are satisfied. It can reference the above mentioned filters to define its own match criteria.

A routing policy can comprise multiple nodes, which are in logic OR relationship. Each node is a match unit, and the system compares each node to a packet in the order of node sequence number. Once a node is matched, the routing policy is passed and the packet will not go through the next node.

Each node comprises a list of **if-match** and **apply** clauses. The **if-match** clauses define the match criteria. The matching objects are some attributes of routing information. The different **if-match** clauses on a node is in logical AND relationship. Only when the matching conditions specified by all the **if-match** clauses on the node are satisfied, can routing information pass the node. The **apply** clauses specify the actions performed after the node is passed, concerning the attribute settings for routing information.

Routing Policy Application

A routing policy is applied in two ways:

- When redistributing routes from other routing protocols, a routing protocol accepts only routes passing the routing policy.
- When receiving or advertising routing information, a routing protocol uses the routing policy to filter routing information.

Routing Policy Configuration Task List

To configure a routing policy, perform the tasks described in the following sections:

Task	
"Defining Filtering Lists" on page 245	"Defining an IP-prefix List" on page 245
	"Defining an AS Path ACL" on page 247
	"Defining a Community List" on page 247
	"Defining an Extended Community List" on page 247
"Configuring a Routing Policy" on page 247	"Creating a Routing Policy" on page 248
	"Defining if-match Clauses for the Routing Policy" on page 248
	"Defining apply Clauses for the Routing Policy" on page 250

Defining Filtering Lists

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, each IPv4 prefix list can comprise multiple items. Each item specifies a matching address range in the form of network prefix identified by index number.

During matching, the system compares the route to each item identified by index number in the ascending order. If one item matches, the route passes the IP-prefix list, without needing to match against the next item.

To define an IPv4 prefix list, use the following commands:

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

*If all 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** mode 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 ipv6-prefix abc index 10 deny 10.1.0.0 16
[Sysname] ip ipv6-prefix abc index 20 deny 10.2.0.0 16
[Sysname] ip ipv6-prefix abc index 30 deny 10.3.0.0 16
[Sysname] ip ipv6-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 matching address range in the form of network prefix, which is identified by index number.

During matching, the system compares the route to each item in the ascending order of index number. If one item is matched, the route passes the IP-prefix list, without needing to match the next item.

To define an IPv6 prefix list, use the following commands:

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

*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** mode 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 ip-prefix abc index 10 deny 2000:1:: 48
[Sysname] ip ip-prefix abc index 20 deny 2000:2:: 48
[Sysname] ip ip-prefix abc index 30 deny 2000:3:: 16
[Sysname] ip ip-prefix abc index 40 permit :: 0 less-equal 128
```

Defining an AS Path ACL

You can define multiple items for an AS path ACL that is identified by number. During matching, the relation between items is logical OR, that is, if the route matches one of these items, it passes the AS path ACL.

To define an AS path ACL, use the following commands:

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.

To define a community list, use the following commands:

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
	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.

To define an extended community list, use the following commands:

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** { *as-number***:***nn* | *ip-address***:***nn* } }&<1-16>	Required Not defined by default

Configuring a Routing Policy

A routing policy is used to filter routing information according to some attributes, and modify some attributes of the routing information that matches the routing policy. Match criteria can be configured using filters above mentioned.

A routing policy can comprise multiple nodes, each 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 performed after specified match criteria are satisfied, concerning attribute settings for passed routing information.

Prerequisites

Before configuring this task, you have completed:

- Filtering list configuration
- Routing protocol configuration

You also need to decide on:

- Name of the routing policy, node sequence numbers
- Match criteria
- Attributes to be modified

Creating a Routing Policy

To create a routing policy, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a routing policy and enter its view	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required

- *If a node has the* ***permit*** *keyword specified, routing information meeting the node's conditions will be handled using the* ***apply*** *clauses of this node, without needing to match against the next node. If routing information does not meet the node's conditions, it will go to the next node for a match.*
- *If a node is specified as* ***deny****, the* ***apply*** *clauses of the node will not be executed. When routing information meets all* ***if-match*** *clauses, it cannot pass the node, nor can it go to the next node. If route information cannot meet any* ***if-match*** *clause of the node, it will go to the next node for a match.*
- *When a routing policy is defined with more than one node, at least one node should be configured with the* ***permit*** *keyword. If the routing policy is used to filter routing information, routing information that does not meet any node's conditions cannot pass the routing policy. If all nodes of the routing policy are set using the* ***deny*** *keyword, no routing information can pass it.*

Defining if-match Clauses for the Routing Policy

To define if-match clauses for a route-policy, use the following command:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter routing policy view	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required

To do…		**Use the command…**	**Remarks**
Define match criteria for IPv4 routes	Match IPv4 routes having destinations specified in the ACL	**if-match acl** *acl-number*	Optional Not configured by default
	Match IPv4 routes having destinations specified in the IP prefix list	**if-match ip-prefix** *ip-prefix-name*	
	Match IPv4 routes having next hops or sources 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
Match IPv6 routes having the next hop or source 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 routes having AS path attributes specified in the AS path ACL(s)		**if-match as-path** *as-path-number*&<1-16>	Optional Not configured by default
Match routes having community attributes in the specified 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 routes having extended attributes contained in the extended community list(s)		**if-match extcommunity** *ext-comm-list-number*&<1-16>	Optional Not configured by default
Match routes having specified outbound interface(s)		**if-match interface** { *interface-type interface-number* }&<1-16>	Optional Not configured by default
Match routes having MPLS label		**if-match mpls-label**	Optional Not configured by default
Match routes 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

To do...	Use the command...	Remarks
Match RIP, OSPF, or IS-IS routes having the specified tag value	**if-match tag** *value*	Optional Not configured by default

- *The **if-match** clauses of a route-policy are in logic AND relationship, namely, routing information has to satisfy all **if-match** clauses before being executed with **apply** clauses.*
- *You can specify no or multiple **if-match** clauses for a routing policy. If no **if-match** clause is specified, and the routing policy is in permit mode, all routing information can pass the node; if in deny mode, no routing information can pass.*
- *A routing policy should use a non VPN ACL for filtering.*
- *The differences between defining **if-match** clauses for IPv4 and IPv6 routing policies are commands for matching the destination, next hop and source address.*

Defining apply Clauses for the Routing Policy

To define apply clauses for a route-policy, use the following command:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a routing policy and enter its view	**route-policy** *route-policy-name* { **permit** | **deny** } **node** *node-number*	Required Not created by default
Set AS_Path attribute for BGP routes	**apply as-path** *as-number*&<1-10> [**replace**]	Optional Not set by default
Specify a community list according to which to delete community attributes of BGP routing information	**apply comm-list** *comm-list-number* **delete**	Optional Not configured by default
Set community attribute for BGP routes	**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 routes	**apply cost** [**+** | **-**] *value*	Optional Not set by default
Set a cost type for routes	**apply cost-type** [**external** | **internal** | **type-1** | **type-2**]	Optional Not set by default
Set the extended community attribute for BGP routes	**apply extcommunity** { **rt** { *as-number***:***nn* | *ip-address***:***nn* } }&<1-16> [**additive**]	Optional Not set by default

To do...		Use the command...	Remarks
Set a next hop	for IPv4 routes	**apply ip-address next-hop** *ip-address*	Optional Not set by default The next hop set using the **apply ip-address next-hop** command does not take effect for route redistribution.
	for IPv6 routes	**apply ipv6 next-hop** *ipv6-address*	Optional Not set by default The next hop set using the **apply ip-address next-hop** command does not take effect for route redistribution.
Redistribute routes to a specified ISIS level		**apply isis** { **level-1** | **level-1-2** | **level-2** }	Optional Not configured by default
Set a local preference for BGP routes		**apply local-preference** *preference*	Optional Not set by default
Set MPLS label		**apply mpls-label**	Optional Not set by default
Set an origin attribute for BGP routes		**apply origin** { **igp** | **egp** *as-number* | **incomplete** }	Optional Not set by default
Set a preference for the matched routing protocol		**apply preference** *preference*	Optional Not set by default
Set a preferred value for BGP routes		**apply preferred-value** *preferred-value*	Optional Not set by default
Set a tag value for RIP, OSPF or IS-IS routes		**apply tag** *value*	Optional Not set by default

- *The difference between IPv4 and IPv6 apply clauses is the command of 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 Routing Policy

To do...	Use the command...	Remarks
Display BGP AS path ACL 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 routing policy information	**display route-policy** [*route-policy-name*]	

To do...	Use the command...	Remarks
Clear IPv4 prefix list statistics	**reset ip ip-prefix** [*ip-prefix-name*]	Available in user view
Clear IPv6 prefix statistics	**reset ip ipv6-prefix** [*ipv6-prefix-name*]	

Routing Policy Configuration Examples

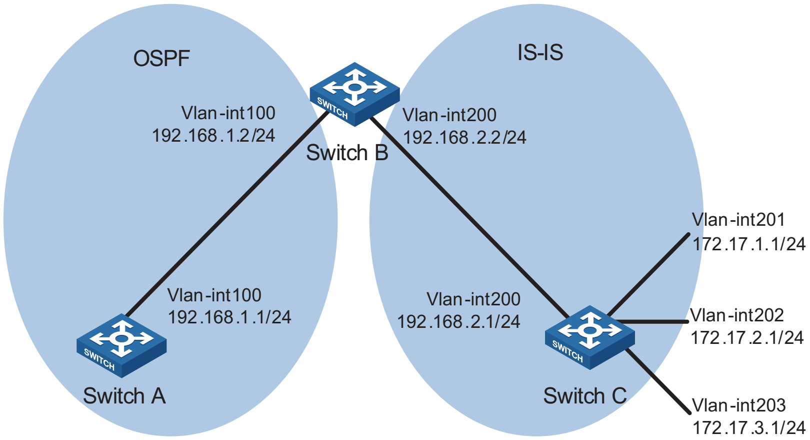
Applying Routing Policy When Redistributing IPv4 Routes

Network Requirements

- Switch B exchanges routing information with Switch A via OSPF, with Switch C via IS-IS.
- On Switch B, configure route redistribution from IS-IS to OSPF and apply a routing policy to set attributes of redistributed routes, setting the cost of route 172.17.1.0/24 to 100, tag of route 172.17.2.0/24 to 20.

Network diagram

Figure 70 Network diagram for routing 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 Switch A: enable OSPF.

```
<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
```

# Configure Switch B: enable OSPF and redistribute routes 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 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 an ACL with the number of 2002, letting pass 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 an IP prefix list named prefix-a, letting pass route 172.17.1.0/24.

```
[SwitchB] ip ip-prefix prefix-a index 10 permit 172.17.1.0 24
```

5 Configure a routing 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 routing policy to route redistribution.

# Configure Switch B: apply the routing 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. You can find the cost of route 172.17.1.0/24 is 100, tag of route 172.17.1.0/24 is 20, and other external routes have no change.

```
[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
```

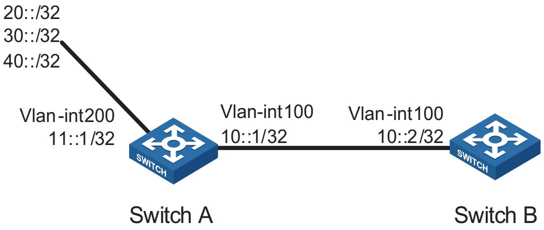
Applying Routing Policy When Redistributing IPv6 Routes

Network requirements

- Enable RIPng and configure three static routes on Switch A.
- Apply a routing policy when redistributing static routes, making routes in 20::0/32 and 40::0/32 pass, routes in 30::0/32 filtered.
- Display RIPng routing table information on Switch B to verify the configuration.

Network diagram

Figure 71 Network diagram for routing 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 routing 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 redistribute static routes.

```
[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
```

Troubleshooting Routing Policy Configuration

IPv4 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while 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.

Processing procedure

1 Use the **display ip ip-prefix** command to display IP prefix list information.

2 Use the **display route-policy** command to display routing policy information.

IPv6 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while 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.

Processing procedure

1 Use the **display ip ipv6-prefix** command to display IP prefix list information.

2 Use the **display route-policy** command to display routing policy information.

26 STATIC ROUTING CONFIGURATION

When configuring a static route, go to the following sections for information you are interested in:

- "Introduction" on page 259
- "Configuring a Static Route" on page 260
- "Displaying and Maintaining Static Routes" on page 262
- "Configuration Example" on page 262

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

Introduction

Static Route

A static route is a special route that is manually configured by the network administrator. If a network's topology is simple, you only need configure static routes for the network to work normally. The proper configuration and usage of static routes can improve a network's performance and ensure bandwidth for important network applications.

The disadvantage of using a static route is that, if a fault or a topological change occurs to the network, the routes will be unavailable and the network breaks. In this case, the network administrator has to modify the static routes manually.

Default Route

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.

You can create the default route with both destination and mask being 0.0.0.0, and some dynamic routing protocols, such as OSPF, RIP and IS-IS, can also generate the default route.

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. The next hop address can not be a local interface IP address; otherwise, the route configuration will not take effect.

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.

When specifying the output interface, note that:

- If the output interface is a NULL0 or loopback interface, there is no need to configure the next hop address.
- You are not recommended to specify a broadcast interface (such as an Ethernet interface, virtual template, or VLAN interface) as the output interface, because a broadcast interface may have multiple next hops. If you have to do so, you must specify the corresponding next hop for the output interface.

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.

You can also enable bidirectional forwarding detection (BFD) to implement fast detection on the next hops of static routes. When a next hop is unreachable, the system can switch to a backup route instantly.

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* } { *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*] **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*]	Required By default, the preference of static routes is 60, tag is 0, and no description information is configured.
Configure the default preference for static routes	**ip route-static default-preference** *default-preference-value*	Optional 60 by default

- *If you specify the next hop of a static route and then configure the next hop as the IP address of a local interface such as a VLAN interface, the static route cannot take effect.*
- *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.*
- *A description can describe the usage, function of some specific static routes for easy and flexible management, classification and configuration of 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.*

Displaying and Maintaining Static Routes

To do...	Use the command...	Remarks
View the current configuration information	**display current-configuration**	Available in any view
View the brief information of the IP routing table	**display ip routing-table**	
View 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

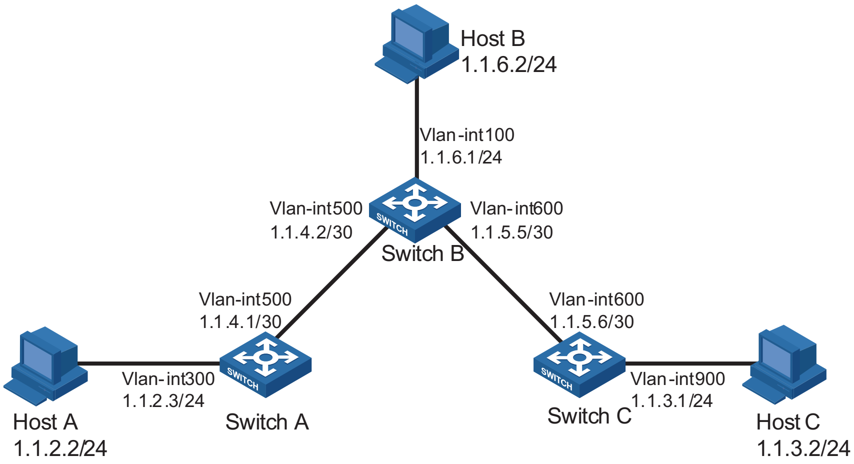
Configuration Example

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.

Network diagram

Figure 72 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.

4 View the configuration result

# 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.

```
[HostB] 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.
```

27 IPv6 Static Routing Configuration

When configuring IPv6 Static Routing, go to these sections for information you are interested in:

- "Introduction to IPv6 Static Routing" on page 265
- "Configuring an IPv6 Static Route" on page 265
- "Displaying and Maintaining IPv6 Static Routes" on page 266
- "IPv6 Static Routing Configuration Example" on page 266

The term "router" in this document refers to either a router in a generic sense or an Ethernet 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

To do...	Use the commands...	Remarks
Enter system view	**System-view**	-
Configure an IPv6 static route with the output interface being a broadcast or NBMA interface	**ipv6 route-static** *ipv6-address prefix-length* [*interface-type interface-number*] *nexthop-address* [**preference** *preference-value*]	Required Not configured by default The default preference of IPv6 static routes is 60.
Configure an IPv6 static route with the output interface being a point-to-point interface	**ipv6 route-static** *ipv6-address prefix-length* { *interface-type interface-number* | *nexthop-address* } [**preference** *preference-value*]	

While configuring a static route, you can configure either the output interface or the next-hop address depending on the situations:

- *If the output interface is a broadcast interface, or an NBMA interface, the next hop address must be specified.*
- *If the output interface is a point-to-point interface, you can specify either the output interface or the next hop address, but not both.*

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

*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.*

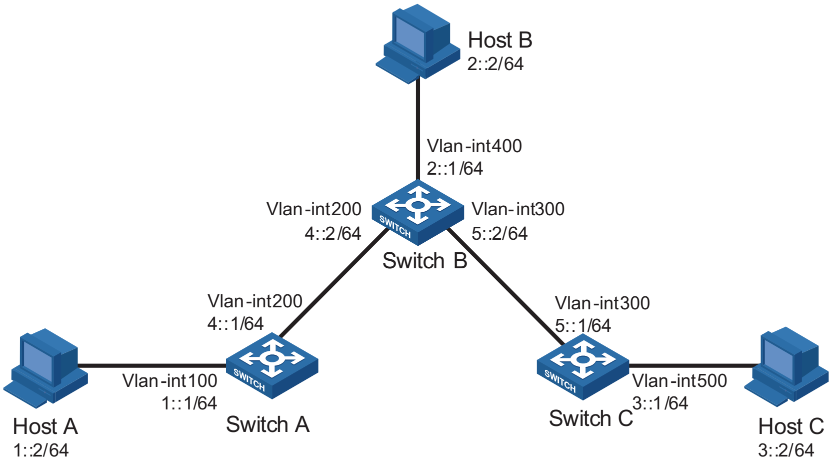
IPv6 Static Routing Configuration Example

Network requirements

With IPv6 static routes configured, all hosts and switches can interact with each other.

Network diagram

Figure 73 Network diagram for static routes (on switches)



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  : ::                                   Protocol     : Static
 NextHop      : FE80::510A:0:8D7:1                   Preference   : 60
 Interface    : Vlan-interface200                    Cost         : 0
```

```
Destination  : ::1                                  Protocol     : Direct
NextHop      : ::1                                  Preference   : 0
Interface    : InLoop0                              Cost         : 0

Destination  : 1::                                  Protocol     : Direct
NextHop      : 1::1                                 Preference   : 0
Interface    : Vlan-interface100                    Cost         : 0

Destination  : 1::1                                 Protocol     : Direct
NextHop      : ::1                                  Preference   : 0
Interface    : InLoop0                              Cost         : 0

Destination  : FE80::                               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
```

28 RIP CONFIGURATION

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

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 structured 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.

RIP Working Mechanism

Basic concept of RIP

RIP is a Distance-Vector-based 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 is known as metric. The hop count from a router to its directly connected network is 0. The hop count from one 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 bigger) 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

Each 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 routing policy to flexibly control routes. For information about routing policy, refer to *"Routing Protocol Overview" on page 189*.

RIP initialization and running procedure

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 all information about their routing tables.
2. The router updates its local routing table, and broadcasts the triggered update messages to its neighbors. All routers on the network do the same to keep the latest routing information.
3. RIP ages out timed out routes by adopting an aging mechanism to keep only valid routes.

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 after the aging time elapses, 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) based 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 the 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 useless 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 the network convergence.

RIP Version

RIP has two versions, RIP-1 and RIP-2.

RIP-1, a Classful Routing Protocol, supports message advertisement via broadcast only. RIP-1 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 RIP-1 does not support discontiguous subnet.

RIP-2 is a Classless Routing Protocol. Compared with RIP-1, RIP-2 has the following advantages.

- Supporting route tags. The route tag is used in routing policies to flexibly control routes.
- Supporting masks, route summarization and classless inter-domain routing (CIDR).
- Supporting designated next hop to select the best next hop on broadcast networks.
- Supporting multicast routing update to reduce resource consumption.
- Supporting Plain text authentication and MD5 authentication to enhance security.

RIP-2 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 RIP-2 broadcast mode can also receive RIP-1 messages.

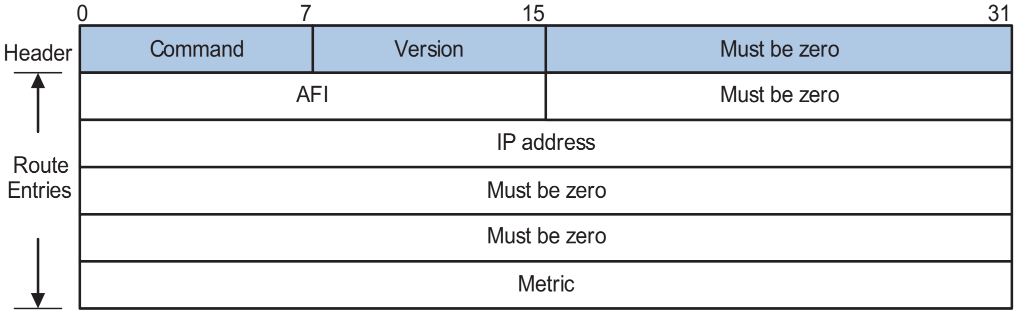
RIP Message Format

RIP-1 message format

A RIP message consists of the Header and up to 25 route entries.

Figure 74 shows the format of RIP-1 message.

Figure 74 RIP-1 Message Format



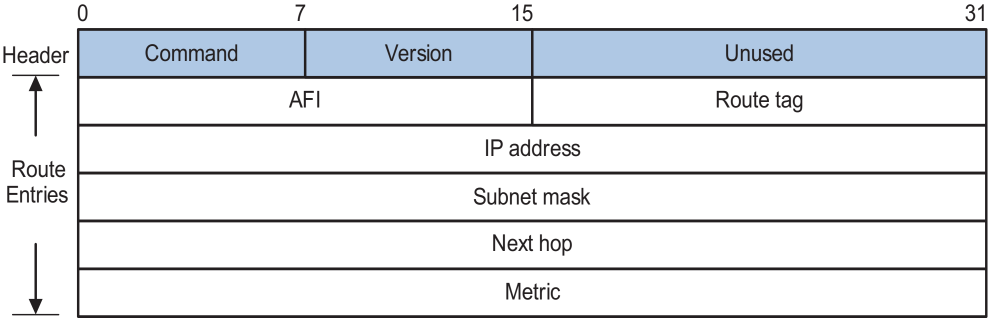
- Command: The type of message. 1 indicates Request, 2 indicates Response.
- Version: The version of RIP, 0x01 for RIP-1.
- AFI: Address Family Identifier, 2 for IP.

- IP Address: Destination IP address of the route; can be a natural network, subnet or a host address.
- Metric: Cost of the route.

RIP-2 message format

The format of RIP-2 message is similar with RIP-1. Figure 75 shows it.

Figure 75 RIP-2 Message Format



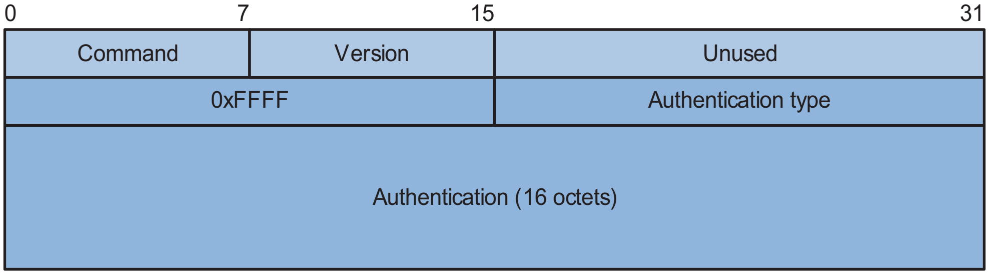
The differences from RIP-1 are stated as following.

- Version: Version of RIP. For RIP-2 the value is 0x02.
- Route Tag: Route Tag.
- IP Address: Destination IP address. It could 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 that the originator of the route.

RIP-2 authentication

RIP-2 sets the AFI field of the first route entry to 0xFFFF to identify authentication information. See Figure 76.

Figure 76 RIP-2 Authentication Message



- Authentication Type: 2 represents plain text authentication, while 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.

RFC 1723 only defines plain text authentication. For information about MD5 authentication, refer to “Configuring RIP-2 Message Authentication” on page 281.

TRIP

Triggered RIP (TRIP), a RIP extension on WAN, is mainly used in dial-up network.

Working mechanism

Routing information is sent in triggered updates rather than periodic broadcasts to reduce the routing management cost the WAN.

- Only when data in the routing table changes or the next hop is unreachable, a routing update message is sent.
- Since the periodic update delivery is canceled, an acknowledgement and retransmission mechanism is required to guarantee successful updates transmission on WAN.

Message types

RIP use three new types of message which are identified by the value of the Command filed.

- Update Request (type value 9): Requests needed routes from the peer.
- Update Response (type value 10): Contains the routes requested by the peer.
- Update Acknowledge (type value 11): Acknowledges received Update Response messages.

TRIP retransmission mechanism

- If receiving no Update Responses after sending an Update Request, a router sends the request again after a specified interval. If still receiving no Update Response after the upper limit for sending requests is reached, the router considers the neighbor unreachable.
- If receiving no Update Acknowledge after sending an Update Response, a router sends the Update Response again after a specified interval. If still receiving no Update Acknowledge after the upper limit for sending Update Responses is reached, the router considers the neighbor unreachable.

RIP Features Supported

The current implementation supports the following RIP features.

- RIP-1 and RIP-2
- RIP Multi-instance. This means that RIP can serve as an internal VPN routing protocol, running between CE and PE on the BGP/MPLS VPN network.
- TRIP

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: RIP-2 MD5 Authentication

RFC 2091: Triggered Extensions to RIP to Support Demand Circuits

Configuring RIP Basic Functions

Configuration Prerequisites

Before configuring RIP features, finish the following tasks.

- Configure the link layer protocol.
- Configure the IP address on each interface, and make sure all adjacent routers are reachable with each other at the network layer.

Configuration Procedure

Enable RIP and specify networks

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 network of an interface	**network** *network-address*	Required Disabled by default

- *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 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** { **all** | *interface-type interface-number* }	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

To do...	Use the command...	Remarks
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 RIP-1 broadcasts and can receive RIP-1 broadcast and unicast packets, RIP-2 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 RIP-1 configured, an interface sends RIP-1 broadcasts, and can receive RIP-1 broadcasts and RIP-1 unicasts.
- With RIP-2 configured, a multicast interface sends RIP-2 multicasts and can receive RIP-2 unicasts, broadcasts and multicasts.
- With RIP-2 configured, a broadcast interface sends RIP-2 broadcasts and can receive RIP-1 unicasts, and broadcasts, RIP-2 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*]	--

Specify a global RIP version	**version** { **1** | **2** }	Optional;
		RIP-1 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 RIP-1 broadcasts, and receive RIP-1 broadcasts, unicasts, RIP-2 broadcasts, multicasts and unicasts.

Return to system view	**quit**	-
Enter interface view	**interface** *interface-type interface-number*	--
Specify a RIP version	**rip version** { **1** | **2** [**broadcast** | **multicast**] }	Optional

Configuring RIP Advanced Functions

In some complex network environments, you need to configure advanced RIP functions.

This section covers the following topics:

- "Configuring an Additional Routing Metric" on page 276

- "Configuring RIP-2 Route Summarization" on page 276
- "Disabling Host Route Reception" on page 277
- "Advertising a Default Route" on page 277
- "Configuring Inbound/Outbound Route Filtering Policies" on page 278
- "Configuring a Priority for RIP" on page 278
- "Configuring RIP Route Redistribution" on page 278

Before configuring RIP routing feature, finish the following tasks:

- Configure an IP address for each interface, and make sure all routers are reachable.
- Configure basic RIP functions

Configuring an Additional Routing Metric

An additional routing metric can be added to the metric of an inbound/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 additional routing metric:

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** *value*	Optional 0 by default
Define an outbound additional routing metric	**rip metricout** *value*	Optional 1 by default

Configuring RIP-2 Route Summarization

The route summarization means that subnet routes in a natural network are summarized with a natural network that is sent to other networks. This function can reduce the size of routing tables.

Configure RIP-2 route automatic summarization

Disable RIP-2 route automatic summarization if you want to advertise all subnet routes.

Follow these steps to configure 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*]	--

To do...	Use the command...	Remarks
Enable RIP-2 automatic route summarization	**summary**	Optional Enabled by default

Advertise a summary route

You can configure RIP-2 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 RIP-2 automatic route summarization	**undo summary**	Required Enabled by default
Exit to system view	**quit**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure to advertise a summary route	**rip summary-address** *ip-address* { *mask* | *mask-length* }	Required

You need disable RIP-2 route automatic summarization before advertising a summary route on an interface.

Disabling Host Route Reception

Sometimes a router may receive many host routes from the same network, which are not helpful for routing and occupy 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

Advertising a Default Route

You can configure RIP to advertise a default route with the specified metric to RIP neighbors.

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*]	--

To do...	Use the command...	Remarks
Enable RIP to advertise a default route	**default-route originate cost** *value*	Required Not enabled by default

The router enabled to advertise a default route does not receive default routes from RIP neighbors.

Configuring Inbound/Outbound Route Filtering Policies

Route filtering is supported by the router. You can filter routes by configuring the inbound and outbound route filtering policies via referencing an ACL and IP prefix list. You can also specify to receive only routes from a specified neighbor.

Follow these steps to configure a routing policy:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter RIP view	**rip** [*process-id*]	--
Define a filtering policy for 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 By default, no inbound filtering is configured by default.
Define a filtering policy for outgoing routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*] | *interface-type interface-number*]	Required No outbound filtering is configured by default.

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 from other routing protocols, you should 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

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*]	--

To do...	Use the command...	Remarks
Configure a default metric for redistributed routes	**default-cost** *value*	Optional The default metric is applied if no metric is specified when redistributing routes.
Redistribute routes from another protocol	**import-route** *protocol* [*process-id*] [**allow-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag*] *	Required

Optimizing the RIP Network

This section covers the following topics:

- "Configuring RIP Timers" on page 279
- "Configuring the Split Horizon and Poison Reverse" on page 279
- "Configuring the Maximum Number of Load Balanced Routes" on page 280
- "Configuring RIP Message Check" on page 280
- "Configuring RIP-2 Message Authentication" on page 281
- "Configuring a RIP Neighbor" on page 281
- "Configuring RIP-to-MIB Binding" on page 282

Finish the following tasks before configuring the RIP network optimization.

- Configure network addresses on interfaces, and make sure neighboring nodes are reachable
- Configure basic RIP 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 By default, 30s for update timer, 180s for timeout timer, 120s for suppress timer, and 120s for garbage-collect timer

Based on the network performance, you should make RIP timers of RIP routers identical to each other to avoid unnecessary traffic or route oscillation.

Configuring the Split Horizon and Poison Reverse

Configure split horizon

The split horizon function disables an interface from sending routes received by the interface itself, so as to prevent routing loops between adjacent routers.

Follow these steps to configure the split horizon function:

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

Disabling the split horizon function on a point-to-point link does not take effect.

Configure the poison reverse

The poison reverse function allows an interface to advertise the routes received by itself, but the metric of these routes is set to 16, making them unreachable.

Follow these steps to configure the poise reserve function:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Enable the poison reverse function	**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

Configuring RIP Message Check

Some fields in the RIP-1 message must be zero. These fields are called zero fields. You can enable the zero field check on received RIP-1 messages. If any such field contains a non-zero value, the RIP-1 message will not be processed. If you are sure that all messages are trusty, you can disable the zero field check to save the CPU processing time.

In addition, you can enable the source IP address validation on received messages. For the message received on an Ethernet interface, 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. For a message received on a serial interface, RIP checks whether the source address of the message is the IP address of the peer interface. If not, RIP discards the message.

Follow these steps to configure RIP message check:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter RIP view	**rip** [*process-id*] [**vpn-instance** *vpn-instance-name*]	--
Enable the zero field check on received RIP-1 messages	**checkzero**	Optional Enabled by default
Enable the source IP address validation on received RIP messages	**validate-source-address**	Optional Enabled by default

- *The zero field check is invalid for RIP-2 messages.*
- *The source IP address validation should be disabled when a non direct RIP neighbor exists.*

Configuring RIP-2 Message Authentication

RIP-2 supports two authentication modes: plain text and MD5.

In plain text authentication, the authentication information is sent with the RIP message, which cannot meet high security needs.

Follow these steps to configure RIP-2 message authentication:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enter interface view	**interface** *interface-type interface-number*	--
Configure RIP-2 authentication mode	**rip authentication-mode** { **md5** { **rfc2082** *key-string key-id* | **rfc2453** *key-string* } | **simple** *password* }	Required

Configuring a RIP Neighbor

Usually, RIP sends messages to broadcast or multicast addresses. On non broadcast or multicast links, you need to manually specify a RIP neighbor. If the specified neighbor is not directly connected, you must disable the source address check on update messages.

Follow these steps to configure 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*]	--
Specify a RIP neighbor	**peer** *ip-address*	Required
Disable source address check on received RIP update messages	**undo validate-source-address**	Required Not disabled by default

*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.*

Configuring RIP-to-MIB Binding

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 the RIP process with the smallest process ID

Displaying and Maintaining RIP Configuration

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** [**statistics** | *ip-address* { *mask* | *mask-length* } | **peer** *ip-address*]	
Clear the statistics of a RIP process	**reset rip** *process-id* **statistics**	Available in user view

RIP Configuration Examples

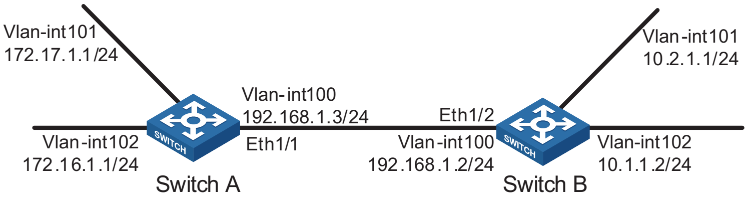
Configuring the RIP Version

Network requirements

As shown in Figure 77, enable RIP-2 on all interfaces on Switch A and Switch B.

Network diagram

Figure 77 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] vlan 100
[SwitchA-vlan100] port ethernet1/1
```

```
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.1 3 24
```

# Configure Switch B.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port ethernet1/2
[SwitchB-vlan100] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.168.1.2 24
```

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 RIP-1 uses natural mask.

3 Configure RIP version

# Configure RIP-2 on Switch A.

```
[SwitchA] rip
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
```

# Configure RIP-2 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 RIP-2 uses classless subnet mask.

Since RIP-1 routing information has a long aging time, it will still exist until aged out after RIP-2 is configured.

Configuring RIP Route Redistribution

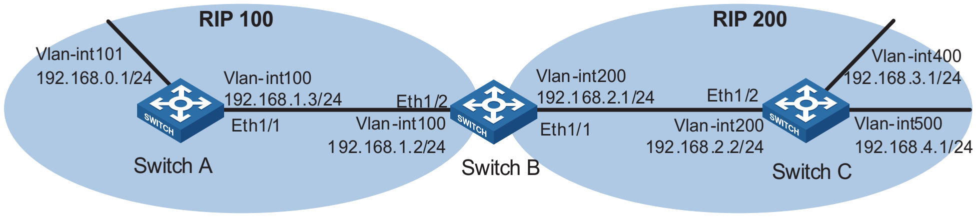
Network requirements

As shown in Figure 78, two RIP processes are running on Switch B, which communicates with Switch A through RIP100 and with Switch C through RIP 200.

Configure route redistribution on Switch B, letting the two RIP processes redistribute routes from each other. Set the cost of redistributed routes from RIP 200 to 3. Configure a filtering policy on Switch B to filter out the route 192.168.4.0/24 from RIP200, making the route not advertised to Switch A.

Network diagram

Figure 78 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 on Switch A.

```
<SwitchA> system-view
[SwitchA] rip 100
[SwitchA-rip-100] network 192.168.0.0
[SwitchA-rip-100] network 192.168.1.0
```

# Enable RIP 100 and RIP 200 on Switch B.

```
<SwitchB> system-view
[SwitchB] rip 100
[SwitchB-rip-100] network 192.168.1.0
[SwitchB-rip-100] quit
[SwitchB] rip 200
[SwitchB-rip-200] network 192.168.2.0
[SwitchB-rip-200] quit
```

# Enable RIP 200 on Switch C.

```
<SwitchC> system-view
[SwitchC] rip 200
[SwitchC-rip-200] network 192.168.2.0
[SwitchC-rip-200] network 192.168.3.0
[SwitchC-rip-200] network 192.168.4.0
```

# Display the routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 10        Routes : 10

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
172.16.1.0/24       Direct 0    0            172.16.1.1      Eth1/3
172.16.1.1/32       Direct 0    0            127.0.0.1       InLoop0
172.17.1.0/24       Direct 0    0            172.17.1.1      Eth1/2
172.17.1.1/32       Direct 0    0            127.0.0.1       InLoop0
192.168.1.0/24      Direct 0    0            192.168.1.3     Vlan100
192.168.1.3/32      Direct 0    0            127.0.0.1       InLoop0
192.168.0.0/24      Direct 0    0            192.168.0.1     Vlan101
192.168.0.1/32      Direct 0    0            127.0.0.1       InLoop0
```

3. Configure route redistribution

   # Configure route redistribution between the two RIP processes on Switch B.

```
[SwitchB] rip 100
[SwitchB-rip-100] default cost 3
[SwitchB-rip-100] import-route rip 200
[SwitchB-rip-100] quit
[SwitchB] rip 200
[SwitchB-rip-200] import-route rip 100
[SwitchB-rip-200] quit
```

# Display the routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12

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
172.16.1.0/24       Direct 0    0            172.16.1.1      Eth1/3
172.16.1.1/32       Direct 0    0            127.0.0.1       InLoop0
172.17.1.0/24       Direct 0    0            172.17.1.1      Eth1/2
172.17.1.1/32       Direct 0    0            127.0.0.1       InLoop0
192.168.1.0/24      Direct 0    0            192.168.1.3     Vlan100
192.168.1.3/32      Direct 0    0            127.0.0.1       InLoop0
192.168.0.0/24      Direct 0    0            192.168.0.1     Vlan101
192.168.0.1/32      Direct 0    0            127.0.0.1       InLoop0
192.168.3.0/24      RIP    100  4            192.168.1.2     Vlan100
192.168.4.0/24      RIP    100  4            192.168.1.2     Vlan100
```

4. Configure an filtering policy to filter redistributed routes

   # Define ACL2000 and reference it to a filtering policy to filter routes redistributed from RIP 200 on Switch B.

```
[SwitchB] acl number 2000
[SwitchB-acl-basic-2000] rule deny source 192.168.4.0 0.0.0.255
[SwitchB-acl-basic-2000] rule permit
[SwitchB-acl-basic-2000] quit
[SwitchB] rip 100
[SwitchB-rip-100] filter-policy 2000 export rip 200
```

# Display the routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
         Destinations : 11       Routes : 11

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
172.16.1.0/24       Direct 0    0            172.16.1.1      Eth1/3
172.16.1.1/32       Direct 0    0            127.0.0.1       InLoop0
172.17.1.0/24       Direct 0    0            172.17.1.1      Eth1/2
172.17.1.1/32       Direct 0    0            127.0.0.1       InLoop0
192.168.1.0/24      Direct 0    0            192.168.1.3     Vlan100
192.168.1.3/32      Direct 0    0            127.0.0.1       InLoop0
192.168.0.0/24      Direct 0    0            192.168.0.1     Vlan101
192.168.0.1/32      Direct 0    0            127.0.0.1       InLoop0
192.168.3.0/24      RIP    100  4            192.168.1.2     Vlan100
```

Troubleshooting RIP Configuration

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 larger 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.

29 IPv6 RIPng Configuration

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

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

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 made the following changes to 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 multicast address.
- Destination Prefix: 128-bit destination address prefix.
- Next hop: 128-bit IPv6 address.
- Source address: RIPng uses the link-local address as the source for sending RIPng route updates.

RIPng Working Mechanism

RIPng is a routing protocol based on the distance vector (D-V) algorithm. RIPng uses UPD 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 after 180 seconds, the routes learned from the neighbor are considered as unreachable. After 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 routing policy to control routing information.

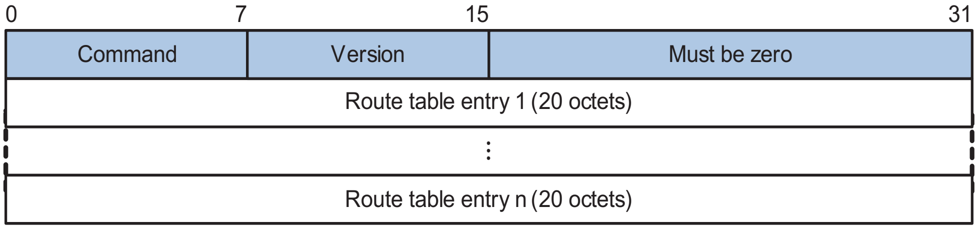
RIPng Packet Format

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 MTU of the sending interface.

Figure 79 shows the packet format of RIPng.

Figure 79 RIPng basic packet format


0
7
15
31
Command
Version
Must be zero
Route table entry 1 (20 octets)
⋮
Route table entry n (20 octets)

- 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.

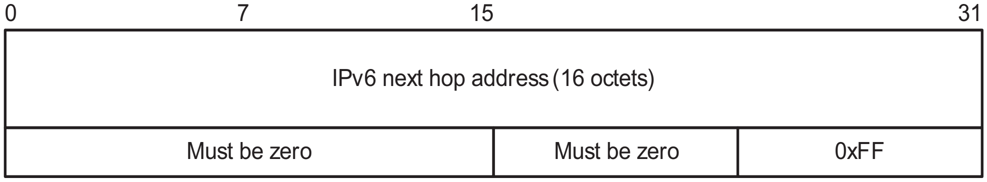
RTE format

There are two types of RTE 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 80 shows the format of the next hop RTE:

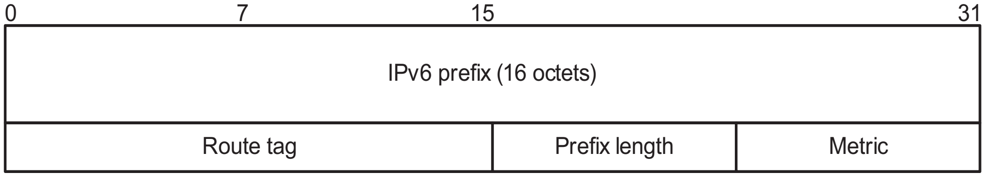
Figure 80 Next hop RTE format



IPv6 next hop address is the IPv6 address of the next hop.

Figure 81 shows the format of the IPv6 prefix RTE.

Figure 81 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, whether the port number is correct. The response packet failed the check will be discarded.

Protocols and Standards

- RFC2080: RIPng for IPv6

- RFC2081: RIPng Protocol Applicability Statement
- RFC2453: RIP Version 2

Configuring RIPng Basic Functions

In this section, you are presented with 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.

Configuration Procedure

Follow these steps to configure the basic RIPng functions:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Enable 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

If RIPng is not enabled on an interface, the interface will not send and receive any RIPng route.

Configuring RIPng Route Control

This section describes how to configure the attributes of RIPng routes, such as configuration of RIPng route preference and cost, and how to control incoming and outgoing routes and how to redistribute routes from other protocols.

This section covers the following topics:

- "Configuring an Additional Route Metric" on page 293
- "Configuring RIPng Route Summarization" on page 293
- "Advertising a Default Route" on page 293
- "Configuring a RIPng Route Filtering Policy" on page 294
- "Configuring the RIPng Priority" on page 294
- "Configuring RIPng Route Redistribution" on page 294

Prerequisites

Before the configuration, accomplish the following tasks first:

- Configure an IPv6 address for each interface.
- Configure RIPng basic functions

- Define an IPv6 ACL before using it for route filtering. Refer to *"IPv6 ACL Configuration" on page 815* for related information.
- Define an IPv6 address prefix list before using it for route filtering. Refer to *"Defining Filtering Lists" on page 245* for related information.

Configuring an Additional Route Metric

An additional route 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 route additional metric	**ripng metricin** *value*	Optional 0 by default
Specify an outbound route 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

After configuration, the device advertises a summary IPv6 prefix rather than a specific route through the interface.

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

With this feature enabled, a default route is advertised via 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 the RIPng Priority

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 By default, 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 If no cost is specified, the default metric applies.

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. This section covers the following topics:

- "Configuring RIPng Timers" on page 295
- "Configuring Split Horizon" on page 295
- "Configuring Poison Reverse" on page 296
- "Enabling Zero Field Check on RIPng Packets" on page 296
- "Configuring the Maximum Number of Equal Cost Routes for Load Balancing" on page 296

Prerequisites

Before tuning and optimizing the RIPng network, complete the following tasks:

- Configure a network layer address for each interface
- Configure the basic RIPng functions

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

When adjusting RIPng timers, you should consider the network performance and perform unified configurations on routers running RIPng to avoid unnecessary network traffic.

Configuring Split Horizon

The split horizon function disables a route learned from an interface from being advertised via the interface to prevent routing loops between neighbors.

Follow these steps to configure the 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

Generally, you are recommended to enable the split horizon to prevent routing loops.

Configuring Poison Reverse

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

- *The poison reverse function enables a route learned from an interface to be advertised through that same interface. However, the metric of the route is set to 16 when advertised outside of the interface the from which the route was learned. Thus making the route unreachable from the interface on which is was learned.*
- *If both the split horizon and poison reverse are configured, only the poison reverse function takes effect.*

Enabling Zero Field Check on RIPng Packets

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 zero field check on RIPng packets	**checkzero**	Optional Enabled by default

Some fields in the RIPng packet must be zero. These fields are called zero fields. With the zero field check on RIPng packets enabled, if such a field contains a non-zero value, the entire RIPng packet will be discarded.

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

Displaying and Maintaining RIPng Configuration

To do...	Use the command...	Remarks
Display configuration information of a RIPng process	**display ripng** [*process-id*]	Available in any view

To do...	Use the command...	Remarks
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

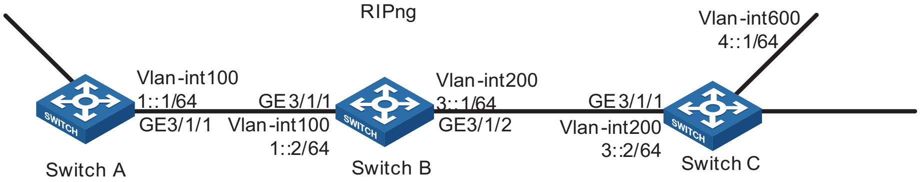
Network requirements

As shown in Figure 82, all switches run RIPng. The prefix length of all IP addresses in the figure is 64, and neighboring switches use link-local IPv6 addresses for interconnection.

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.

Network diagram

Figure 82 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] 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 200
[SwitchA-Vlan-interface200] ripng 1 enable
[SwitchA-Vlan-interface200] quit
```

# Configure Switch B.

```
[SwitchB] ripng 1
[SwitchB-ripng-1] quit
[SwitchB] interface Vlan-interface 101
[SwitchB-Vlan-interface101] ripng 1 enable
[SwitchB-Vlan-interface101] quit
[SwitchB] interface Vlan-interface 100
```

```
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

# Configure Switch C.

```
[SwitchC] ripng 1
[SwitchC-ripng-1] quit
[SwitchC] interface Vlan-interface 101
[SwitchC-Vlan-interface101] ripng 1 enable
[SwitchC-Vlan-interface101] quit
[SwitchC] interface Vlan-interface 300
[SwitchC-Vlan-interface300] ripng 1 enable
[SwitchC-Vlan-interface300] quit
[SwitchC] interface Vlan-interface 400
[SwitchC-Vlan-interface400] ripng 1 enable
[SwitchC-Vlan-interface400] quit
```

# Display the routing table of Switch B.

```
[SwitchB] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::F54C:0:9FDB:1  on Vlan-interface101
 Dest 2::/64,
     via FE80::F54C:0:9FDB:1, cost  1, tag 0, A, 3 Sec
 Dest 3::/64,
     via FE80::F54C:0:9FDB:1, cost  1, tag 0, A, 3 Sec

 Peer FE80::D472:0:3C23:1  on Vlan-interface100
 Dest 1::/64,
     via FE80::D472:0:3C23:1, cost  1, tag 0, A, 4 Sec
```

# Display the routing table of Switch A.

```
[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::476:0:3624:1  on Vlan-interface100
 Dest 2::/64,
     via FE80::476:0:3624:1, cost  2, tag 0, A, 21 Sec
 Dest 3::/64,
     via FE80::476:0:3624:1, cost  2, tag 0, A, 21 Sec
```

3 Configure Switch B to filter incoming 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
```

# 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::F54C:0:9FDB:1  on Vlan-interface101
 Dest 2::/64,
     via FE80::F54C:0:9FDB:1, cost  1, tag 0, A, 14 Sec

 Peer FE80::D472:0:3C23:1  on Vlan-interface100
 Dest 1::/64,
     via FE80::D472:0:3C23:1, cost  1, tag 0, A, 25 Sec

[SwitchA] display ripng 1 route
   Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
 ----------------------------------------------------------------

 Peer FE80::476:0:3624:1  on Vlan-interface100
 Dest 2::/64,
     via FE80::476:0:3624:1, cost  2, tag 0, A, 7 Sec
```

30 OSPF CONFIGURATION

Open Shortest Path First (OSPF) is a link state based 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:

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

Introduction to OSPF

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 the 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 the 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 the routes into intra-area, inter-area, external type-1, and external type-2 routes.

- Authentication: Supports interface-based packet authentication to guarantee the security of packet exchange.
- Multicast: Supports packet multicasting 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 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

To run OSPF, a router must have a Router ID, which is a 32-bit unsigned integer, the unique identifier of the router in the AS.

You may assign a Router ID to an OSPF router manually. If no Router ID is specified, the system automatically selects one for the router as follows:

- If the loopback interfaces are configured, select the highest IP address among them.
- If no loopback interface is configured, select the highest IP address among addresses of active interfaces on the router.

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 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 peer. After exchanging the DD packets, the two routers know which LSAs of the neighbor routers are missing from the local LSDBs. In this case, they send LSR packets, requesting the missing LSAs. The packets contain the digests of the missing LSAs.
- LSU (Link State Update) Packet: Transmits the needed LSAs to the peer router.

- LSAck (Link State Acknowledgment) Packet: Acknowledges received LSU packets. It contains the Headers of LSAs requiring acknowledgement (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 Autonomous System.
- 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 and Route Summarization

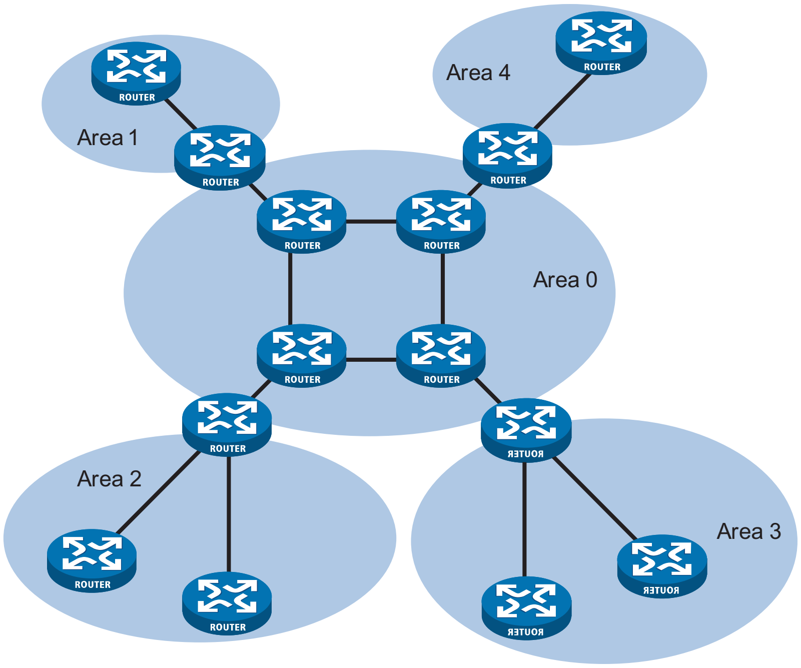
Area partition

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 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 83 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.

Classification of Routers

The OSPF router falls 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.

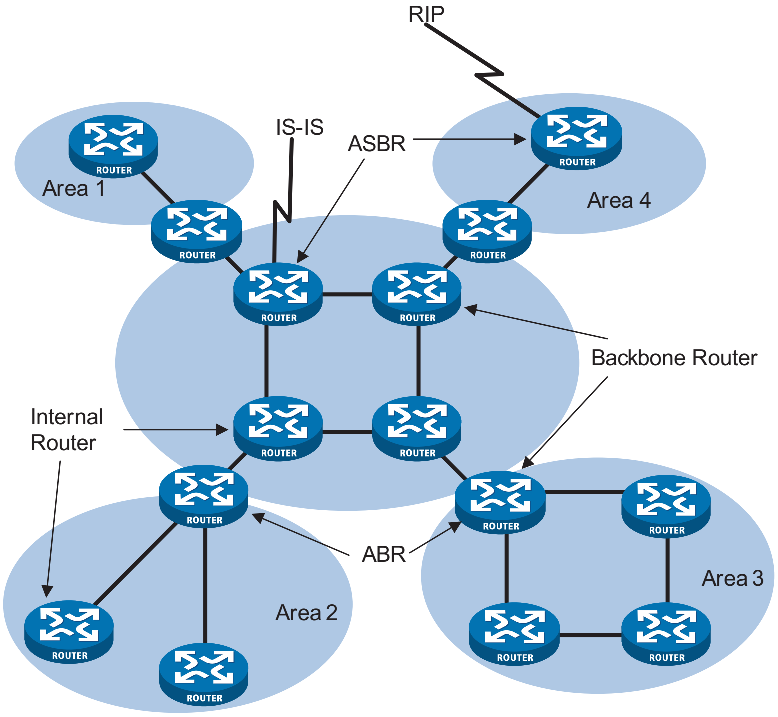
3 Backbone Router

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 84 OSPF router types



Backbone area and virtual links

An AS has a unique area called 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:

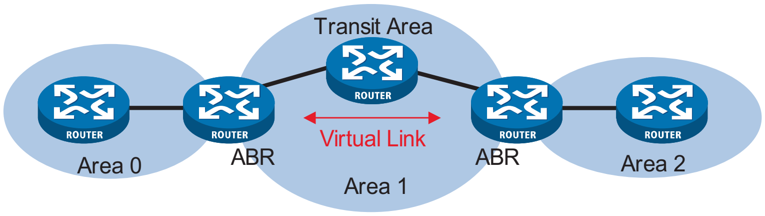
- All non-backbone areas must maintain connectivity to the backbone area.
- The backbone area itself must maintain connectivity.

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".

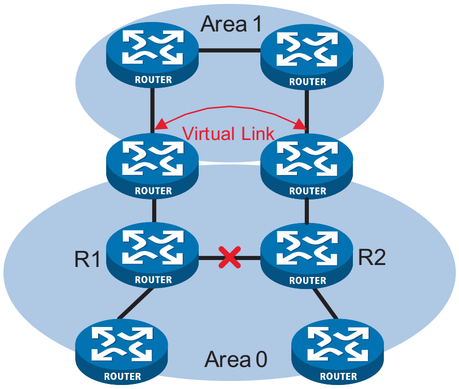
In the following figure, Area 2 has no direct physical link to the backbone area 0. Configuring a virtual link between ABRs can connect Area 2 to the backbone area.

Figure 85 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 86 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, the OSPF routers in between simply convey these OSPF packets as normal IP packets.

(Totally) Stub area

The ABR in a stub area does not distribute Type5 LSAs into the area, so the routing table scale and amount of routing information in this area are reduced significantly.

You can also configure the stub area as a Totally Stub area, where the ABR advertises neither the routes of other areas nor the external routes.

Stub area configuration is optional, and not every area is qualified 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
- The **stub** command must be configured on routers in 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.

NSSA area

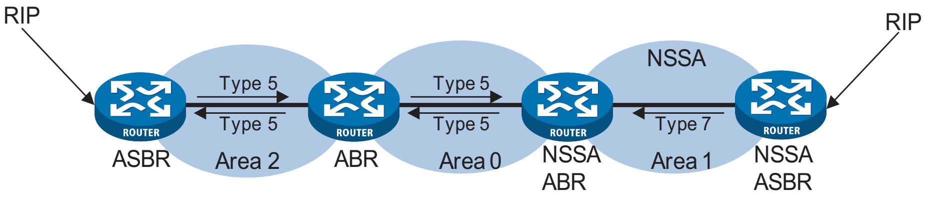
Similar to a stub area, an NSSA area imports no AS external LSA (Type5 LSA) but can import Type7 LSAs that are generated by the ASBR and distributed throughout the NSSA area. When traveling to the NSSA ABR, Type7 LSAs are translated into Type5 LSAs by the ABR for advertisement to other areas.

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 Type7 LSAs and advertises them throughout Area 1. When these LSAs travel to the NSSA ABR, the ABR translates Type7 LSAs to Type5 LSAs for advertisement to Area 0 and Area 2.

On the left of the figure, RIP routes are translated into Type5 LSAs by the ASBR of Area 2 and distributed into the OSPF AS. However, Area 1 is an NSSA area, so these Type5 LSAs cannot travel to Area 1.

Similar to stub areas, virtual links cannot transit NSSA areas.

Figure 87 NSSA area



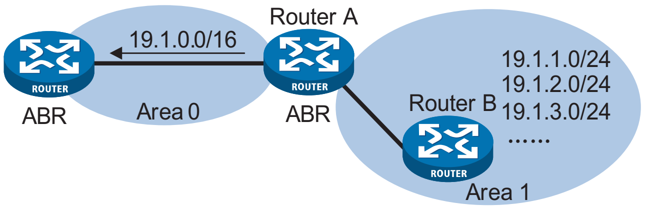
Route summarization

Route summarization: An ABR or ASBR summarizes routes with the same prefix with a single route and distribute it to other areas.

Via route summarization, routing information across areas and the size of routing tables on routers will be reduced, improving calculation speed of routers.

For example, as shown in the following figure, in Area 1 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 Router A, the three routes are summarized with the route 19.1.0.0/16 that is advertised into Area 0.

Figure 88 Route summarization



OSPF has two types of route summarization:

1 ABR route summarization

To distribute routing information to other areas, an ABR generates Type3 LSAs on a per network segment basis for an attached non-backbone area. If contiguous network segments are available in the area, you can summarize them with a single network segment. The ABR in the area distributes only the summary LSA to reduce the scale of LSDBs on routers in other areas.

2 ASBR route summarization

If summarization for redistributed routes is configured on an ASBR, it will summarize redistributed Type5 LSAs that fall into the specified address range. If in an NSSA area, it also summarizes Type7 LSAs that fall into the specified address range.

If this feature is configured on an ABR, the ABR will summarize Type5 LSAs translated from Type7 LSAs.

Route types

OSPF prioritize routes into four levels:

- Intra-area route
- Inter-area route
- Type1 external route
- Type2 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: Type1 and Type2.

A Type1 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 Type1 external route= the cost from the router to the corresponding ASBR+ the cost from the ASBR to the destination of the external route.

A Type2 external route is an EGP route, which has low credibility, so OSPF considers the cost from the ASBR to the destination of the Type2 external route is much bigger than the cost from the ASBR to an OSPF internal router. Therefore, the cost from the internal router to the destination of the Type2 external route= the cost from the ASBR to the destination of the Type2 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, packets are sent to multicast addresses (such as 224.0.0.5 and 224.0.0.6).

- 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.

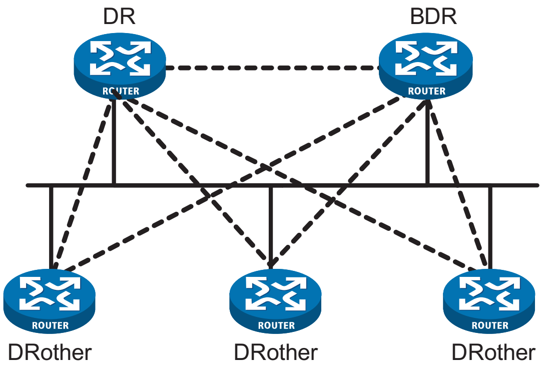
If the DR fails to work, routers on the network have to elect another DR and synchronize information with the new DR. It is time-consuming and prone to routing calculation errors. The Backup Designated Router (BDR) was introduced to reduce the synchronization period.

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 with each other and exchange no routing information, 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 89 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.

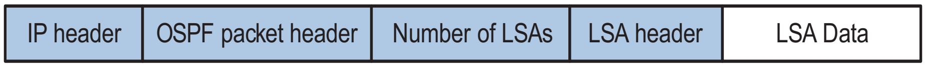
Note that:

- 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, taking a LSU packet as an example, is shown below.

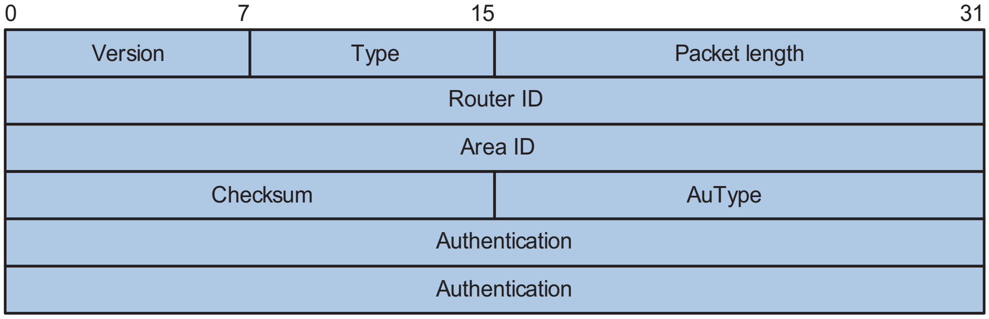
Figure 90 OSPF packet format



OSPF packet header

OSPF packets are classified into five types that have the same packet header, as shown below.

Figure 91 OSPF packet header



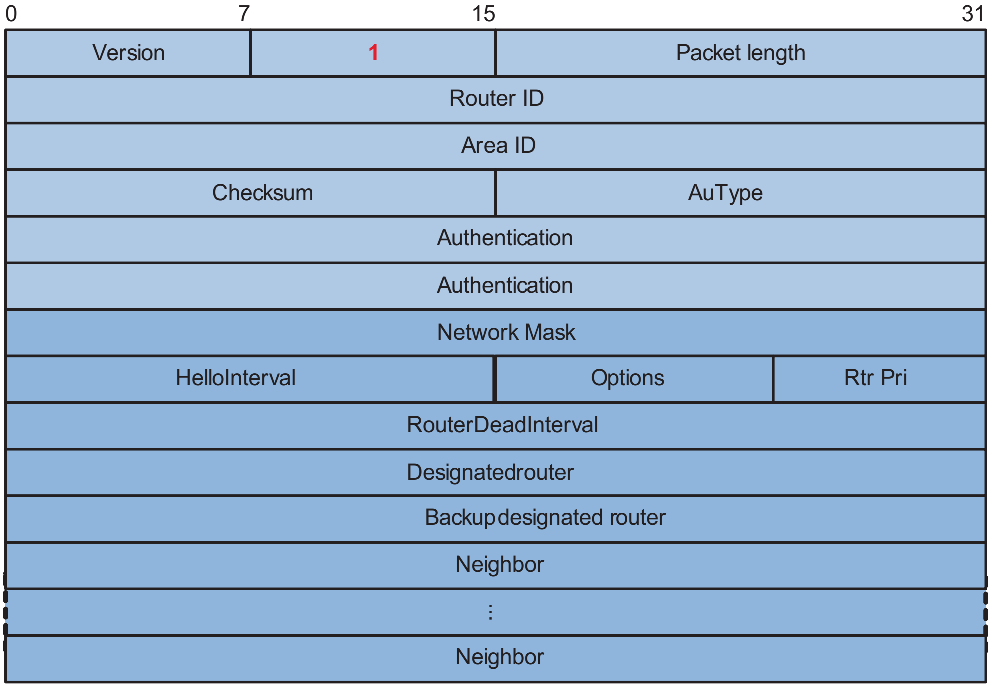
- Version: OSPF version number, which is 2 for OSPFv2.
- Type: OSPF packet type from 1 to 5, corresponding with hello, DD, LSR, LSU and LSAck respectively.
- Packet length: Total length of the OSPF packet in bytes, including the header
- Router ID: ID of the advertising router
- Area ID: ID of the area 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, which is not defined for authentication type 0, password information for authentication type 1, information about Key ID, MD5 authentication data length and sequence number for authentication type 2.

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 DR/BDR, including information about values of timers, DR, BDR and neighbors already known. The format is shown below:

Figure 92 Hello packet format



Major fields:

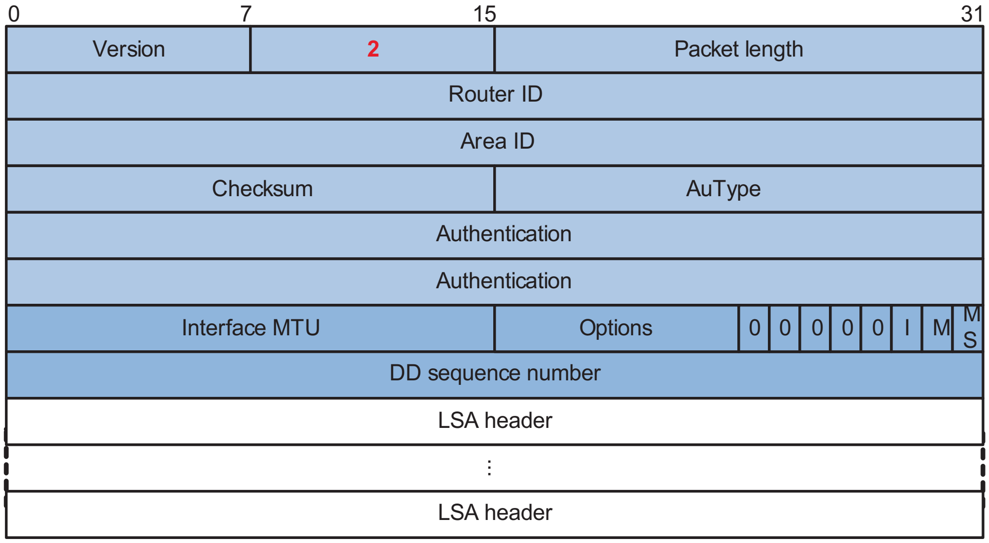
- Network Mask: The network mask associated with the router's sending interface. If two routers have different network masks, they cannot become neighbors.
- HelloInterval: The interval between the router's hello packets. If two routers have different intervals, they cannot become neighbors.
- Rtr Pri: Router priority. A value of 0 means the router cannot become the DR/BDR.
- RouterDeadInterval: The time value before declaring a silent router down. If two routers have different time values of this kind, 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, so reducing traffic between routers. The recipient checks whether the LSA is available using the LSA header.

The DD packet format:

Figure 93 DD packet format



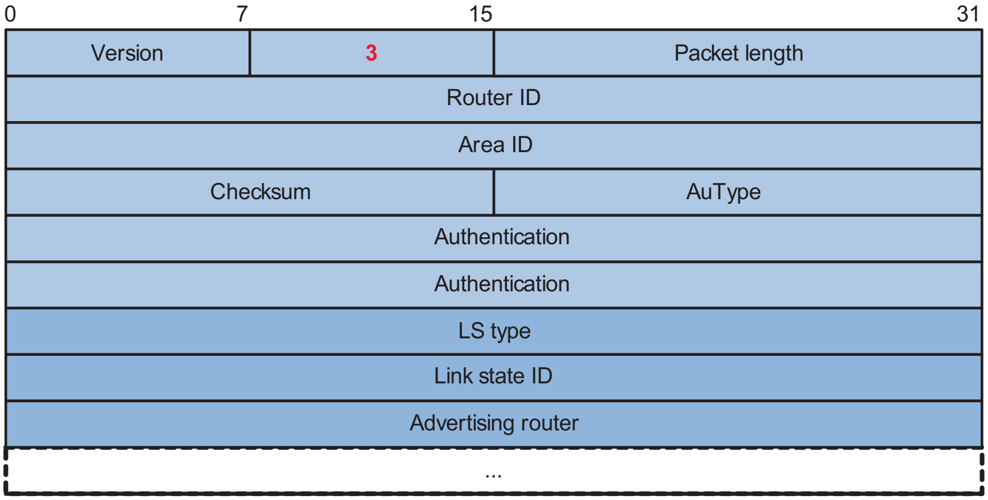
Major fields:

- Interface MTU: The 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 in the sequence 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 in the sequence 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 94 LSR packet format



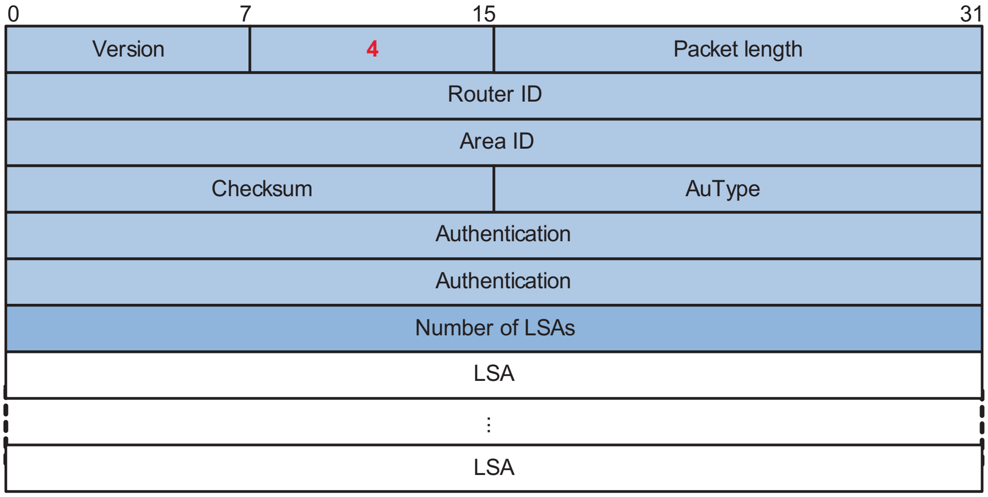
Major fields:

- LS type: The 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: The 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.

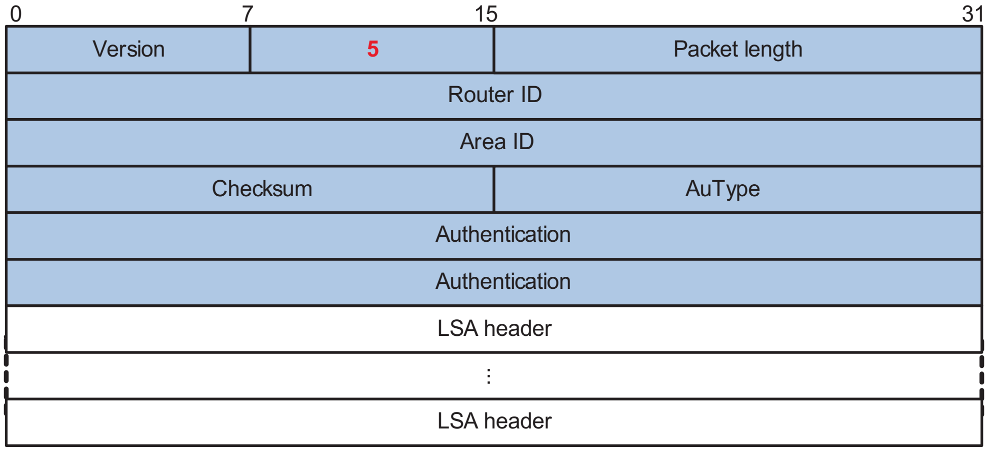
Figure 95 LSU packet format



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.

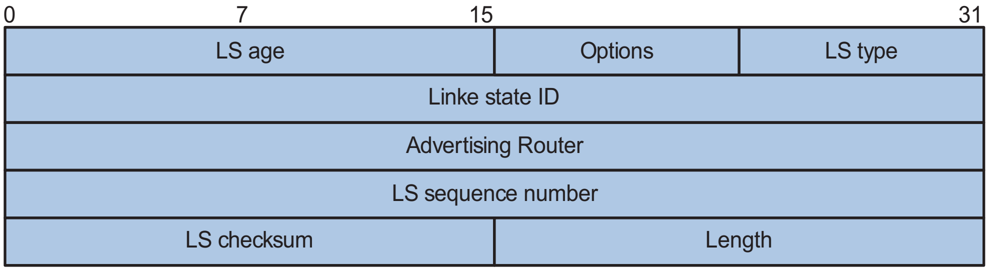
Figure 96 LSAck packet format



LSA header format

All LSAs have the same header, as shown in the following figure.

Figure 97 LSA header format



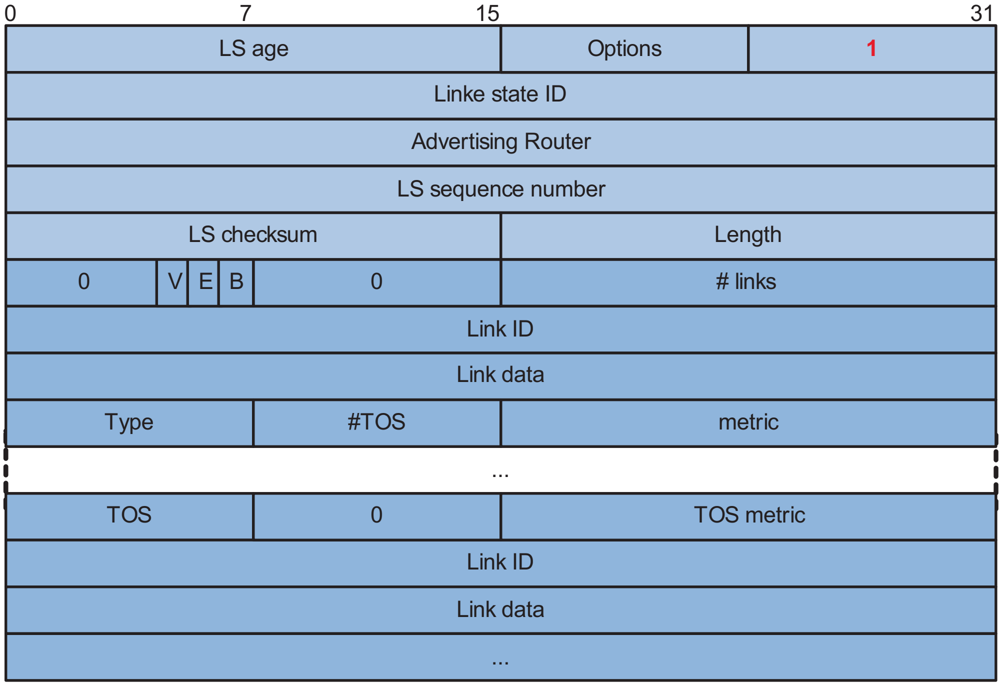
Major fields:

- LS age: The time in seconds elapsed since the LSA was originated. A LSA ages in the LSDB (added 1 per second), but does not in transmission.
- LS type: The 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: The length in bytes of the LSA, including the LSA header

Formats of LSAs

1 Router LSA

Figure 98 Router LSA format



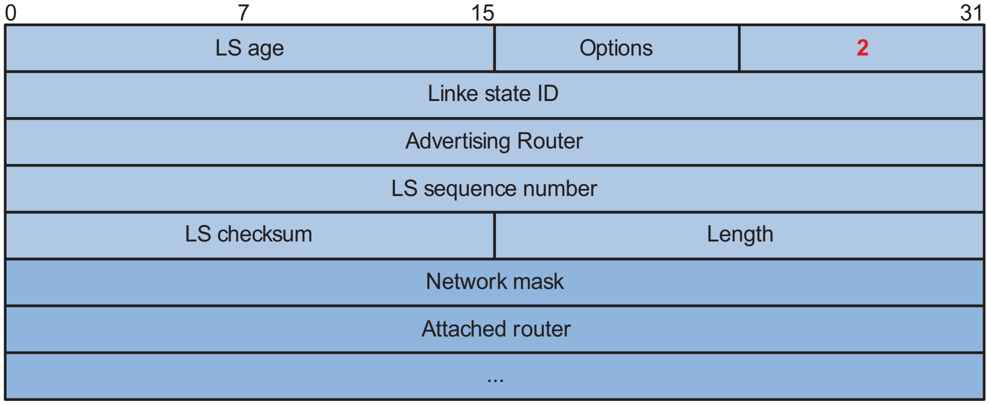
Major fields:

- Link State ID: The 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: The 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: The number of different TOS metrics given for this link.
- metric: The 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 99 Network LSA format



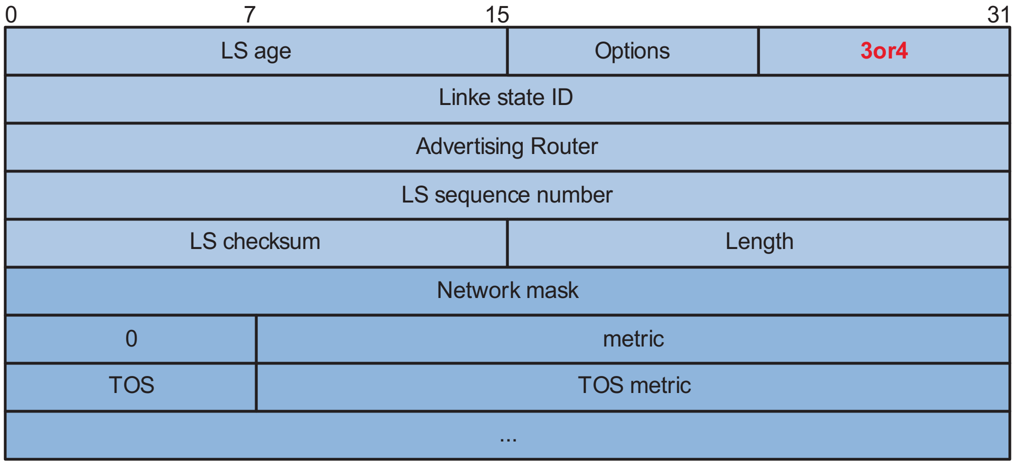
Major fields:

- 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 (Type3 LSAs) and ASBR summary LSAs (type4 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 100 Summary LSA format



Major fields:

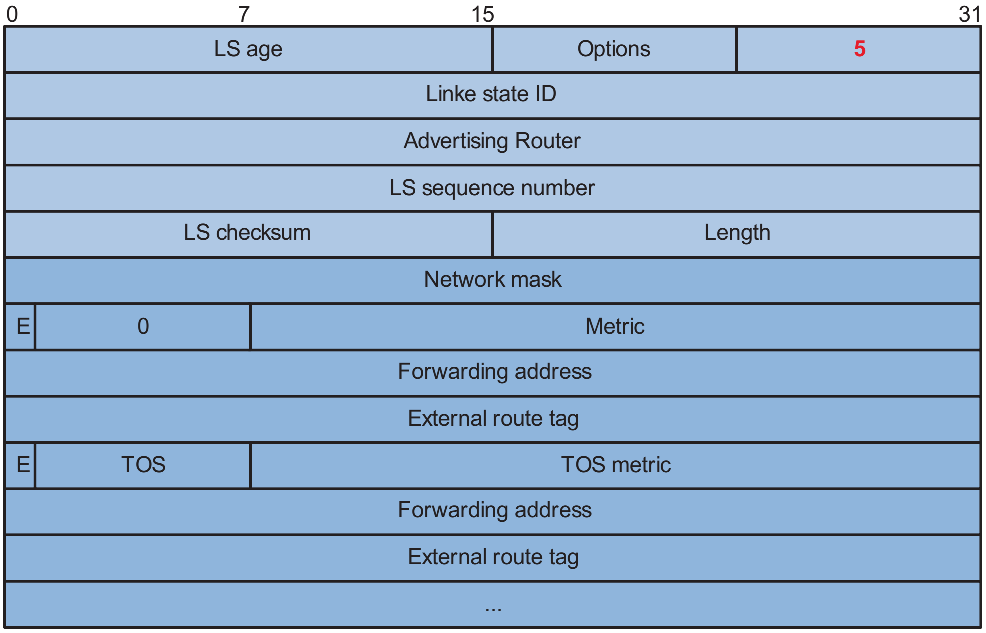
- Link State ID: For a Type3 LSA, it is an IP address outside the area; for a type 4 LSA, it is the router ID of an ASBR outside the area.
- Network Mask: The network mask for the type 3 LSA; set to 0.0.0.0 for the type4 LSA
- metric: The metric to the destination

A Type3 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 101 AS external LSA format



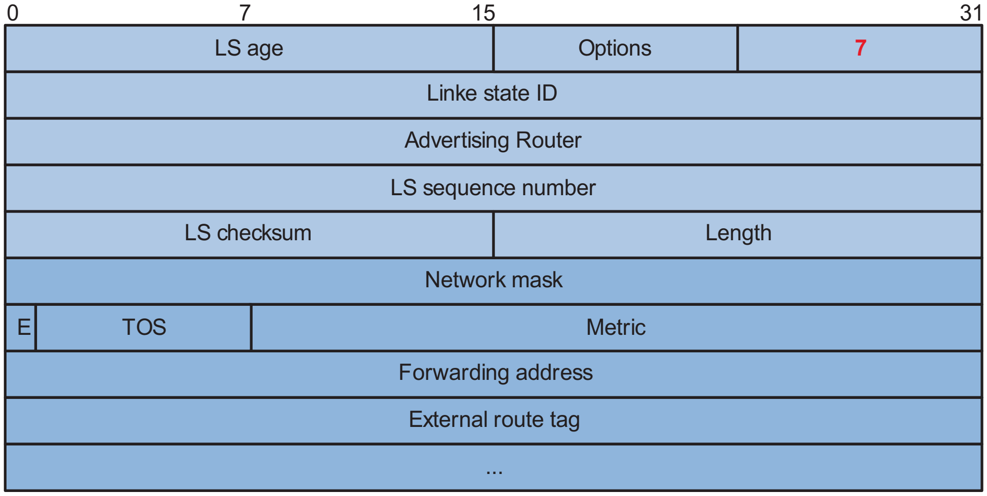
Major fields:

- 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” on page 308 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 102 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.

Hot Standby and GR

Distributed routers support OSPF Hot Standby (HSB). OSPF backups necessary information of the active fabric into the standby fabric. Once the active fabric fails, the standby fabric begins to work to ensure the normal operation of OSPF.

OSPF supports to backup:

- All OSPF data to the standby fabric to make sure OSPF recovers normal operation immediately upon the main fabric failure.
- Only the OSPF configuration information to the standby fabric. Once the main fabric fails, OSPF will perform Graceful Restart (GR), obtaining adjacencies from and synchronizing the Link State Database with neighbors.

The Graceful Restart of the router is mainly used for High Availability (HA) and will not interfere with any other routers.

When a router shuts down, its neighbors will delete it from their neighbor tables and inform other routers, resulting in SPF recalculation. If the router restarts in several seconds, it is unnecessary to perform SPF recalculation, and reestablish adjacencies.

To avoid unnecessary SPF calculation, when a router restarts, it will inform neighboring routers the shutdown is temporary. Then these routers will not delete the router from their neighbor tables, and other routers have no idea about this restart.

After recovering to normal, the router obtains the Link State Database from neighboring routers via the GR related synchronization mechanism.

For OSPF GR configuration, refer to "Configuring OSPF-based Graceful Restart" on page 797.

TE and DS-TETE

OSPF Traffic Engineering (TE) provides for the establishment and maintenance of Label Switch Paths (LSPs) of TE.

When establishing Constraint-based Routed LSPs (CR LSPs), MPLS obtains the TE information of links in the area via OSPF.

OSPF has a new LSA, Opaque LSA, which can be used for carrying TE information.

DiffServ Aware TE (DS-TE) provides for network resource optimization and allocation, flow classification, and indication of network bandwidth consumption of each flow in a link. TE is implemented on the classified type (thin granularity summarization type) rather than the summarized type (thick granularity summarization type) to improve performance and bandwidth utilization.

To support DS-TE application in MPLS, OSPF supports Local Overbooking Multiplier TLV and Bandwidth Constraint (BC) TLV.

IGP Shortcut and Forwarding Adjacency

IGP Shortcut and Forwarding Adjacency enable OSPF to use a LSP as the outbound interface for a destination. Without them, OSPF cannot use the LSP as the outbound interface.

Differences between IGP Shortcut and Forwarding Adjacency:

- If Forwarding Adjacency is enabled only, OSPF can also use an LSP as the outbound interface for a destination
- If LGP Shortcut is enabled only, only the router enabled with it can use LSPs for routing.

VPN

OSPF supports multi-instance, which can run on PEs 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.

Configuring area IDs on PEs can differentiate VPNs. Sites in the same VPN are considered as directly connected. PE routers then exchange OSPF routing information like on a dedicated line, thus network management and OSPF operation efficiency are improved.

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 Type3 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) for a special 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 low 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

OSPF Configuration Task List

To configure OSPF, perform the tasks described in the following sections:

Task		Description
“Configuring OSPF Basic Functions” on page 323		Required
“Configuring OSPF Area Parameters” on page 324		Optional
“Configuring OSPF Network Types” on page 325	“Configuring the OSPF Network Type for an Interface” on page 326	Optional
	“Configuring an NBMA Neighbor” on page 326	Optional
	“Configuring a DR Priority for an OSPF Interface” on page 326	Optional

Task		Description
"Configuring OSPF Route Control" on page 327	"Configuring OSPF Route Summarization" on page 327	Optional
	"Configuring OSPF Inbound Route Filtering" on page 328	Optional
	"Configuring ABR Type3 LSA Filtering" on page 328	Optional
	"Configuring the OSPF Link Cost of an Interface" on page 328	Optional
	"Configuring the Maximum Number of OSPF Routes" on page 329	Optional
	"Configuring the Maximum Number of Equal Cost Routes for Load Balancing" on page 329	Optional
	"Configuring the Priority of OSPF Routes" on page 329	Optional
	"Configuring OSPF Route Redistribution" on page 330	Optional

Task		Description
"Configuring OSPF Network Optimization" on page 330	"Configuring OSPF Packet Timers" on page 331	Optional
	"Configuring the LSA Transmission Delay" on page 332	Optional
	"Configuring the SPF Calculation Interval" on page 332	Optional
	"Configuring the Minimum LSA Repeating Arrival Interval" on page 333	Optional
	"Configuring the LSA Generation Interval" on page 333	Optional
	"Disabling Interfaces from Sending OSPF Packets" on page 333	Optional
	"Configuring Stub Routers" on page 334	Optional
	"Configuring OSPF Authentication" on page 334	Optional
	"Adding Interface MTU into DD Packets" on page 335	Optional
	"Configuring the Maximum Number of External LSAs in LSDB" on page 335	Optional
	"Making External Route Selection Rules Defined in RFC1583 Compatible" on page 335	Optional
	"Configuring OSPF Network Management" on page 336	Optional
	"Enabling the Advertisement and Reception of Opaque LSAs" on page 336	Optional

Configuring OSPF Basic Functions

You need to enable OSPF, specify an interface and area ID first before performing other 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 ensure OSPF stability, you need to decide on router IDs and configure them manually. Any two routers in an AS must have different IDs. In practice, the ID of a router is the IP address of one of its interfaces.

The system supports OSPF multi-process. When a router runs multiple OSPF processes, you need to specify an ID for each process, which takes effect locally and has no interference on packet exchange between routers. Therefore, two routers having different process IDs can exchange packets.

The system supports OSPF multi-instance. You can configure an OSPF process to run in a specified VPN instance to configure an association between the two.

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.

To configure OSPF basic functions, use the following commands:

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*] *	-
Configure a description for the OSPF process	**description** *description*	Optional Not configured by default
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

- *An OSPF process ID is unique, including the process ID for OSPF multi-instance, which cannot be the same as any previously configured ID.*
- *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 Area Parameters

Splitting an OSPF AS into multiple areas reduces the number of LSAs on networks and extends OSPF application. 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 on routers in these areas and the number of LSAs.

A stub area cannot redistribute routes, thus introducing the concept of NSSA, where type 7 LSAs (NSSA External LSAs) are advertised. Type 7 LSAs originate from the ASBR in a NSSA area. When arriving at the ABR in the NSSA area, these LSAs will be translated into type 5 LSAs for advertisement to other areas.

Non-backbone areas exchange routing information via the backbone area. Therefore, the backbone and non-backbone areas, including the backbone itself must maintain connectivity.

If necessary physical links are not available for this connectivity maintenance, 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 network layer.
- OSPF basic functions

Configuration Procedure

To configure OSPF area parameters, use the following commands:

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*	Required
Configure the area as a stub area	**stub** [**no-summary**]	Optional Not configured by default
Configure the area as an NSSA area	**nssa** [**default-route-advertise** | **no-import-route** | **no-summary**] *	Optional Not configured by default
Specify a cost for the default route advertised to the stub or NSSA area	**default-cost** *cost*	Optional Defaults to 1
Create and 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*] *	Optional Configured on both ends of a virtual link Note that **hello** and **dead** parameters must be identical on both ends of the link
Configure and advertise a host route	**host-advertise** *ip-address cost*	Optional Not advertised by default

- *It is required to use the* ***stub*** *command on routers attached to a stub area.*
- *It is required to use the* ***nssa*** *command on routers attached to an NSSA area.*
- *Using the* ***default-cost*** *command only takes effect on the ABR of a stub area or the ABR/ASBR of an NSSA area.*

Configuring OSPF Network Types

OSPF classifies networks into four types upon link layer protocols. Since an NBMA network must be fully meshed, namely, any two routers in the network must have a virtual link in between. In most cases, however, the requirement cannot be satisfied, so you need to change the network type using commands.

For routers having no direct link in between, you can configure related interfaces as the P2MP mode. If a router in the NBMA network has only a single peer, you can also configure associated interfaces as the P2P mode.

In addition, when configuring broadcast and NBMA networks, you can specify for interfaces DR priorities for DR/BDR election. In practice, routers having higher reliability should become the DR/BDR.

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

To configure the OSPF network type for an interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure the network type	**ospf network-type** { **broadcast** | **nbma** | **p2mp** | **p2p** }	Optional Not configured by default

- *Configuring a new network type for an interface overwrites the previous network one (if any).*
- *If the two interfaces on a link are both configured as the broadcast, NBMA or P2MP type, they can not establish neighbor relationship unless they are on the same network segment.*

Configuring an NBMA Neighbor

For NBMA interfaces that cannot broadcast hello packets to find neighbors, you need to specify IP addresses and DR priorities of neighbors manually.

To configure a neighbor and its DR priority, use the following commands:

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 an NBMA neighbor and its DR priority	**peer** *ip-address* [**dr-priority** *dr-priority*]	Required

Configuring a DR Priority for an OSPF Interface

For broadcast or NBMA interfaces, you can configure DR priorities for DR/BDR election. To configure a DR priority for an OSPF interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure a DR priority for the interface	**ospf dr-priority** *priority*	Optional The default DR priority is 1

The DR priority configured with the ***ospf dr-priority*** *command and the one with the* ***peer*** *command have the following differences:*

- The former is for actual DR election.
- The latter is to indicate whether a neighbor has 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, 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 will send a hello packet to the neighbor with priority 0 for adjacency relationship establishment.

Configuring OSPF Route Control

This section is to configure control of OSPF routing information advertisement and reception, and route redistribution from other protocols.

Prerequisites

To configure this task, you have configured:

- IP addresses for interfaces
- OSPF basic functions
- Corresponding filters if routing information filtering is needed.

Configuring OSPF Route Summarization

OSPF route summarization includes:

- Configure route summarization between OSPF areas on an ABR
- Configure route summarization when redistributing routes into OSPF on an ASBR

To configure route summarization between OSPF areas on an ABR, use the following commands:

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*	Required
Configure ABR route summarization	**abr-summary** *ip-address* { *mask* | *mask-length* } [**advertise** | **not-advertise**] [**cost** *cost*]	Required Available on an ABR only Not configured by default

To configure route summarization when redistributing routes into OSPF on an ASBR, use the following commands:

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 Available on an ASBR only Not configured by default

Configuring OSPF Inbound Route Filtering

To configure OSPF to filter received routes, use the following commands:

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 to filter received routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* | **gateway** *ip-prefix-name* } **import**	Required Not configured by default

*Since OSPF is a link state-based internal gateway protocol, routing information is contained in LSAs. However, OSPF cannot filter LSAs. Using the **filter-policy import** command is to filter routes computed by OSPF, and only routes not filtered are added into the routing table.*

Configuring ABR Type3 LSA Filtering

To configure type 3 LSA filtering on an ABR, use the following commands:

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 Type3 LSA filtering	**filter** { *acl-number* | **ip-prefix** *ip-prefix-name* } { **import** | **export** }	Required Not configured by default

Configuring the OSPF Link Cost of an Interface

To configure the link cost for an interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	**-**
Enter interface view	**interface** *interface-type interface-number*	-
Configure the cost value of the interface	**ospf cost** *value*	Optional By default, an interface computes its cost according to the baud rate The cost value defaults to 1 for VLAN interfaces

To configure a bandwidth reference value, use the following commands:

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*]*	**-**

To do...	Use the command...	Remarks
Configure a bandwidth reference value	**bandwidth-reference** *value*	Optional The value defaults to 100 Mbps

If the cost value is not configured for an interface, OSPF computes the interface cost automatically: Interface cost= Bandwidth reference value/Interface bandwidth. If the calculated cost is greater than 65535, the value of 65535 is used.

Configuring the Maximum Number of OSPF Routes

To configure the maximum number of routes, use the following commands:

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

Configuring the Maximum Number of Equal Cost Routes for Load Balancing

If several routes with the same cost to the same destination are available, configuring them as load-balanced routes can improve link utilization.

To configure the maximum number of equal cost routes, use the following commands:

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 equal cost routes for load balancing	**maximum load-balancing** *maximum*	Optional

Configuring the Priority of OSPF Routes

A router may run multiple routing protocols. The router 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.

To configure the priority for OSPF, use the following commands:

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*] *	**-**

To do...	Use the command...	Remarks
Configure the priority of OSPF routes	**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

To configure OSPF route redistribution, use the following commands:

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 other protocols	**import-route** *protocol* [*process-id* | **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
Redistribute a default route	**default-route-advertise** [**always** | **cost** *cost* | **type** *type* | **route-policy** *route-policy-name*]* **default-route-advertise summary cost** *cost*	Optional Not redistributed by default
Configure the default values of 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 Type2.

- *Using the **import-route** command cannot redistribute a default external route. To do so, you need to use the **default-route-advertise** command.*
- *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 will advertise the default route to the CE.*
- *By filtering redistributed routes, OSPF translates only routes, which are not filtered out, into Type5 LSAs or Type7 LSAs for advertisement.*
- *You can configure default values of parameters for redistributed routes, such as the cost, upper limit, tag and type of external routes. The tag is used to indicate information related to protocol, for example, when redistributing BGP routes, OSPF uses the tag to differentiate AS IDs.*

Configuring OSPF Network Optimization

You can optimize your OSPF network in the following ways:

- Change values of 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, 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 retransmit timer: Interval within which if the interface receives no acknowledgement packets after sending a LSA to the neighbor, it will retransmit the LSA.

To configure timers for OSPF packets, use the following commands:

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 dead interval defaults to 40 seconds on P2P, Broadcast interfaces and 120 seconds on P2MP and NBMA interfaces.

To do...	Use the command...	Remarks
Specify the retransmission interval	**ospf timer retransmit** *interval*	Optional The retransmission interval defaults to 5 seconds.

- *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.*

Configuring the LSA Transmission Delay

Since OSPF packets need time for traveling on links, extending LSA age time with some delay time is necessary, especially for low speed links.

To configure the LSA transmission delay time on an interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	**-**
Enter interface view	**interface** *interface-type interface-number*	-
Set the LSA transmission delay time	**ospf trans-delay** *seconds*	Optional Set to 1 second by default

Configuring the SPF Calculation Interval

Link State Database changes lead to SPF calculations. When an OSPF network changes frequently, a large amount of network resources will be occupied, reducing working efficiency of routers. You can adjust the SPF calculation interval for the network to reduce negative influence. To do so, use the following commands:

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*] *	-
Set the SPF calculation interval	**spf-schedule-interval** *maximum-interval* [*minimum-interval* [*incremental-interval*]]	Optional By default, the interval is 5 seconds

With this command 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$2^{n-2}$ (n is the number of calculation times) each time a calculation occurs, up to the maximum-interval.

Configuring the Minimum LSA Repeating Arrival Interval

When an interface receives an LSA that is the same with the previously received LSA within a specified interval, the minimum LSA repeating arrival interval, the interface will discard the LSA. To configure the minimum LSA repeating arrival interval, use the following commands:

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

*The interval set by the **lsa-arrival-interval** command should be smaller or equal to the interval set by the **lsa-generation-interva**l command.*

Configuring the LSA Generation Interval

With this feature configured, you can protect network resources and routers from being over consumed due to frequent network changes.

To configure LSA generation interval, use the following commands:

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 millisecond and the incremental interval is 5000 milliseconds.

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¬$^{2}2^{n-2}$ (n is the number of generation times) each time a generation occurs, up to the maximum-interval.

Disabling Interfaces from Sending OSPF Packets

To disable an interface from sending routing information to other routers, use the following commands:

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

- *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 informs other OSPF routers not to use it to forward data, but they can have a route to the stub router.

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.

To configure a router as a stub router, use the following commands:

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

A stub router has nothing to do with a stub area.

Configuring OSPF Authentication

By supporting packet authentication, OSPF receives packets that pass the authentication only, so failed packets cannot establish neighboring relationship.

To configure OSPF authentication, use the following commands:

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** { **simple** | **md5** }	Required Not configured by default
Exit to OSPF view	**quit**	-
Exit to system view	**quit**	-

To do...	Use the command...	Remarks
Enter interface view	**interface** *interface-type interface-number*	-
Configure the authentication mode (simple authentication) for the interface	**ospf authentication-mode simple** [**plain** | **cipher**] *password*	Optional Not configured by default
Configure the authentication mode (MD5 authentication) for the interface	**ospf authentication-mode** { **md5 | hmac-md5** } *key-id* [**plain** | **cipher**] *password*	

The authentication mode and password for all interfaces attached to the same area must be identical.

Adding 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. To add the interface MTU into DD packets, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Enable OSPF to add 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

To configure the maximum number of external LSAs in the Link State Database, use the following commands:

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 No limitation 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.

To make them compatible, use the following commands:

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

Configuring OSPF Network Management

To Configure OSPF network management, use the following commands:

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 bound with OSPF MIB by default
Enable OSPF trap	**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
Enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*]*	-
Enable messages logging	**enable log** [**config** | **error** | **state**]	Optional Not enabled by default
Enable the logging on neighbor state changes	**log-peer-change**	Optional Enabled by default

Enabling the Advertisement and Reception of Opaque LSAs

With this feature enabled, the OSPF router can receive and advertise the Type 9, Type 10 and Type 11 opaque LSAs.

To enable the advertisement and reception of opaque LSAs, use the following commands:

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

Displaying and Maintaining OSPF Configuration

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 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* }]	
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**	
Remove redistributed routes	**reset ospf** [*process-id*] **redistribution**	

OSPF Configuration Examples

CAUTION: *In these examples, only commands related to OSPF configuration are described.*

Configuring OSPF Basic Functions

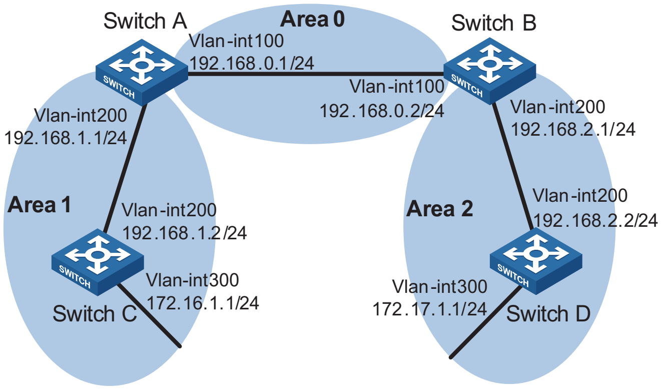
Network requirements

As shown in the following figure, all switches run OSPF. The AS is split into three areas, in which, SwitchA and SwitchB act as ABRs to forward routing information between areas.

After configuration, all switches can learn routes to every network segment in the AS.

Network diagram

Figure 103 Network diagram for OSPF basic configuration



Configuration procedure

1 Configure IP addresses for interfaces (omitted)

2 Configure OSPF basic functions

# Configure SwitchA

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.0.0 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 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

# Configure SwitchB

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.0.0 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 192.168.2.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.2] quit
[SwitchB-ospf-1] quit
```

# Configure SwitchC

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 192.168.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] network 172.16.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

# Configure SwitchD

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 192.168.2.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] network 172.17.1.0 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 SwitchA.

```
[SwitchA] display ospf peer

          OSPF Process 1 with Router ID 192.168.1.1
                  Neighbors

 Area 0.0.0.0 interface 192.168.0.1(Vlan-interface100)'s neighbors
 Router ID: 192.168.2.1      Address: 192.168.0.2      GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 192.168.0.2  BDR: 192.168.0.1  MTU: 0
   Dead timer due in 31  sec
   Neighbor is up for 00:01:09
   Authentication Sequence: [ 0 ]

                  Neighbors

 Area 0.0.0.1 interface 192.168.1.1(Vlan-interface200)'s neighbors
 Router ID: 192.168.1.2      Address: 192.168.1.2      GR State: Normal
   State: Full  Mode: Nbr is Master  Priority: 1
   DR: 192.168.1.2  BDR: 192.168.1.1  MTU: 0
   Dead timer due in 39  sec
   Neighbor is up for 00:01:01
   Authentication Sequence: [ 0 ]
```

# Display OSPF routing information on SwitchA.

```
[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
 172.16.1.0/24      2        Stub    192.168.1.2     192.168.1.2   0.0.0.1
 172.17.1.0/24      3        Inter   192.168.0.2     192.168.2.1   0.0.0.0
 192.168.0.0/24     1        Transit 192.168.0.1     192.168.2.1   0.0.0.0
 192.168.1.0/24     1        Transit 192.168.1.1     192.168.1.2   0.0.0.1
 192.168.2.0/24     2        Inter   192.168.0.2     192.168.2.1   0.0.0.0

 Total Nets: 5
 Intra Area: 3  Inter Area: 2  ASE: 0  NSSA: 0
```

# Display the Link State Database on SwitchA.

```
[SwitchA] display ospf lsdb

          OSPF Process 1 with Router ID 192.168.1.1
                  Link State Database

                          Area: 0.0.0.0
 type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    192.168.1.1     192.168.1.1        536  36    80000004       0
 Router    192.168.2.1     192.168.2.1        528  36    80000004       0
 Network   192.168.0.2     192.168.2.1        528  32    80000002       0
 Sum-Net   192.168.1.0     192.168.1.1        581  28    80000001       1
 Sum-Net   172.17.1.0      192.168.2.1        516  28    80000001       2
 Sum-Net   192.168.2.0     192.168.2.1        582  28    80000001       1
 Sum-Net   172.16.1.0      192.168.1.1        531  28    80000001       2
                          Area: 0.0.0.1
 type      LinkState ID    AdvRouter          Age  Len   Sequence   Metric
 Router    192.168.1.2     192.168.1.2        537  48    80000006       0
 Router    192.168.1.1     192.168.1.1        534  36    80000004       0
 Network   192.168.1.2     192.168.1.2        537  32    80000002       0
 Sum-Net   172.17.1.0      192.168.1.1        515  28    80000001       3
 Sum-Net   192.168.2.0     192.168.1.1        536  28    80000001       2
 Sum-Net   192.168.0.0     192.168.1.1        581  28    80000001       1
```

# Display the OSPF routing table on SwitchD.

```
[SwitchD] display ospf routing

          OSPF Process 1 with Router ID 192.168.2.2
                   Routing Tables

 Routing for Network
 Destination        Cost     type    NextHop         AdvRouter     Area
 172.16.1.0/24      4        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 172.17.1.0/24      1        Stub    172.17.1.1      192.168.2.2   0.0.0.2
 192.168.0.0/24     2        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 192.168.1.0/24     3        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 192.168.2.0/24     1        Transit 192.168.2.2     192.168.2.2   0.0.0.2

 Total Nets: 5
 Intra Area: 2  Inter Area: 3  ASE: 0  NSSA: 0
```

# Ping the IP address 172.16.1.1 to check connectivity.

```
[SwitchD] ping 172.16.1.1
  PING 172.16.1.1: 56  data bytes, press CTRL_C to break
    Reply from 172.16.1.1: bytes=56 Sequence=1 ttl=253 time=62 ms
    Reply from 172.16.1.1: bytes=56 Sequence=2 ttl=253 time=16 ms
    Reply from 172.16.1.1: bytes=56 Sequence=3 ttl=253 time=62 ms
    Reply from 172.16.1.1: bytes=56 Sequence=4 ttl=253 time=94 ms
    Reply from 172.16.1.1: bytes=56 Sequence=5 ttl=253 time=63 ms

  --- 172.16.1.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 16/59/94 ms
```

Configuring an OSPF Stub Area

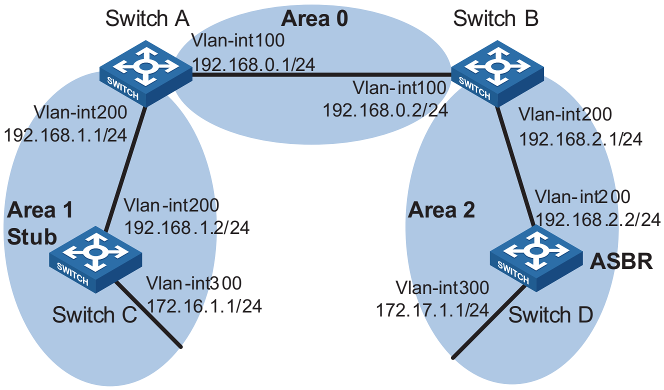
Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. SwitchA and SwitchB act as ABRs to forward routing information between areas. SwitchD acts as the ASBR, redistributing routes (static routes).

It is required to configure Area1 as a Stub area, reducing LSAs to this area without affecting route reachability.

Network diagram

Figure 104 Network diagram for OSPF Stub area configuration



Configuration procedure

1 Configure IP addresses for interfaces (omitted).

2 Configure OSPF basic functions (in the previous example).

3 Configure SwitchD to redistribute static routes.

```
[SwitchD] ip route-static 200.0.0.0 8 null 0
[SwitchD] ospf
[SwitchD-ospf-1] import-route static
[SwitchD-ospf-1] quit
```

# Display ABR/ASBR information on SwitchC

```
[SwitchC] display ospf abr-asbr

          OSPF Process 1 with Router ID 192.168.1.2
                  Routing Table to ABR and ASBR

 type         Destination        Area       Cost  Nexthop         Rttype
 Intra        192.168.1.1        0.0.0.1    1     192.168.1.1     ABR
 Inter        192.168.2.2        0.0.0.1    3     192.168.1.1     ASBR
```

# Display OSPF routing table information on SwitchC.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.2
                   Routing Tables

 Routing for Network
 Destination          Cost     type     NextHop          AdvRouter       Area
 172.16.1.0/24        1        Stub     172.16.1.1       192.168.1.2     0.0.0.1
 172.17.1.0/24        4        Inter    192.168.1.1      192.168.1.1     0.0.0.1
 192.168.0.0/24       2        Inter    192.168.1.1      192.168.1.1     0.0.0.1
 192.168.1.0/24     1        Transit 192.168.1.2     192.168.1.2   0.0.0.1
 192.168.2.0/24     3        Inter   192.168.1.1     192.168.1.1   0.0.0.1

 Routing for ASEs
 Destination        Cost     type    Tag       NextHop         AdvRouter
 200.0.0.0/8         1         Type2    1         192.168.1.1    192.168.2.2

 Total Nets: 6
 Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

In the above output, since SwitchC resides in a normal OSPF area, its routing table contains an external route.

4 Configure Area1 as a Stub area.

# Configure SwitchA.

```
[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 SwitchC.

```
[SwitchC] ospf
[SwitchC-ospf-1] stub-router
[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 the OSPF routing table on SwitchC

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.2
                   Routing Tables

 Routing for Network
 Destination        Cost     type    NextHop         AdvRouter     Area
 0.0.0.0/0          65536    Inter   192.168.1.1     192.168.1.1   0.0.0.1
 172.16.1.0/24      1        Stub    172.16.1.1      192.168.1.2   0.0.0.1
 172.17.1.0/24      65538    Inter   192.168.1.1     192.168.1.1   0.0.0.1
 192.168.0.0/24     65536    Inter   192.168.1.1     192.168.1.1   0.0.0.1
 192.168.1.0/24     65535    Transit 192.168.1.2     192.168.1.2   0.0.0.1
 192.168.2.0/24     65537    Inter   192.168.1.1     192.168.1.1   0.0.0.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 4  ASE: 0  NSSA: 0
```

When SwitchC resides in the Stub area, a default route takes the place of the external route.

# Filter Type3 LSAs for 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 the OSPF routing table on SwitchC.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.2
                   Routing Tables

 Routing for Network
 Destination         Cost     type    NextHop         AdvRouter      Area
 0.0.0.0/0           65536    Inter   192.168.1.1     192.168.1.1    0.0.0.1
 172.16.1.0/24       1        Stub    172.16.1.1      192.168.1.2    0.0.0.1
 192.168.1.0/24      65535    Transit 192.168.1.2     192.168.1.2    0.0.0.1

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

After this configuration, routing table entries on the Stub router are further reduced, containing only one default external route.

Configuring an OSPF NSSA Area

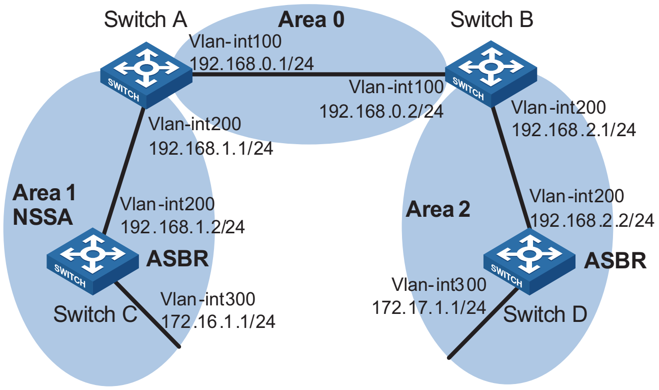
Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. SwitchA and SwitchB act as ABRs to forward routing information between areas. SwitchD acts as the ASBR, redistributing routes (static routes)

It is required to configure Area1 as an NSSA area, RouterC as the ASBR to redistribute static routes into the AS.

Network diagram

Figure 105 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" on page 323).

3 Configure SwitchD to import external static routes (refer to "Authentication" on page 319 previous example)

4 Configure Area1 as an NSSA area.

# Configure SwitchA.

```
[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 SwitchC.

```
[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
```

*It is recommended to configure the **nssa** command with the keyword **default-route-advertise no-summary** on SwitchA (an ABR) to reduce the routing table size on NSSA routers. On other NSSA routers, using the **nssa** command is ok.*

# Display OSPF routing table information on SwitchC.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.2
                   Routing Tables

 Routing for Network
 Destination        Cost     type    NextHop         AdvRouter     Area
 0.0.0.0/0          2        Inter   192.168.1.1     192.168.1.1   0.0.0.1
 172.16.1.0/24      1        Stub    172.16.1.1      192.168.1.2   0.0.0.1
 192.168.1.0/24     1        Transit 192.168.1.2     192.168.1.2   0.0.0.1

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

5 Configure SwitchC to redistribute static routes.

```
[SwitchC] ip route-static 100.0.0.0 8 null 0
[SwitchC] ospf
[SwitchC-ospf-1] import-route static
[SwitchC-ospf-1] quit
```

# Display OSPF routing table information on SwitchD.

```
[SwitchD-ospf-1] display ospf routing

          OSPF Process 1 with Router ID 192.168.2.2
                   Routing Tables

 Routing for Network
 Destination        Cost     type    NextHop         AdvRouter     Area
 172.16.1.0/24      4        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 172.17.1.0/24      1        Stub    172.17.1.1      192.168.2.2   0.0.0.2
 192.168.0.0/24     2        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 192.168.1.0/24     3        Inter   192.168.2.1     192.168.2.1   0.0.0.2
 192.168.2.0/24     1        Transit 192.168.2.2     192.168.2.2   0.0.0.2

 Routing for ASEs
 Destination        Cost     type    Tag       NextHop         AdvRouter
 100.0.0.0/8        1        Type2   1         192.168.2.1     192.168.1.1

 Total Nets: 6
 Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```

You can see on SwitchD an external route imported from the NSSA area.

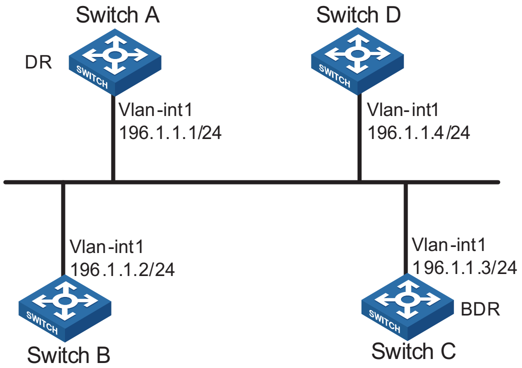
Configuring OSPF DR Election

Network requirements

- In the following figure, OSPF Switches A, B, C and D reside on the same network segment.
- It is required to configure SwitchA as the DR, SwitchC as the BDR.

Network diagram

Figure 106 Network diagram for OSPF DR election configuration



Configuration procedure

1 Configure IP addresses for interfaces (omitted)

2 Configure OSPF basic functions

# Configure SwitchA

```
<SwitchA> system-view
[Switch A] router id 1.1.1.1
[Switch A] ospf
[Switch A-ospf-1] area 0
[Switch A-ospf-1-area-0.0.0.0] network 196.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

# Configure SwitchB

```
<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 196.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

# Configure SwitchC

```
<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 196.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

# Configure SwitchD

```
<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 196.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit
```

# Display OSPF neighbor information on SwitchA.

```
[SwitchA] display ospf peer

          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 DR priorities on interfaces

# Configure SwitchA

```
[SwitchA] interface vlan-interface 1
[RouterA-Vlan-interface1] ospf dr-priority 100
[RouterA-Vlan-interface1] quit
```

# Configure SwitchB

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ospf dr-priority 0
[SwitchB-Vlan-interface1] quit
```

# Configure SwitchC

```
[SwitchC] interface vlan-interface 1
[SwitchC-Vlan-interface1] ospf dr-priority 2
[SwitchC-Vlan-interface] quit
```

# Display neighbor information on SwitchD.

```
[SwitchD] display ospf peer

          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.

In the above output, you can find the priority configuration does not take effect immediately.

4 Restart OSPF process (omitted)

# Display neighbor information on SwitchD.

```
[SwitchD] display ospf peer

          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 ]
```

SwitchA becomes the DR, and SwitchC is the BDR.

If the neighbor state is full, it means SwitchD has established 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
```

The interface state DROther means the interface is not the DR/BDR.

Configuring OSPF Virtual Links

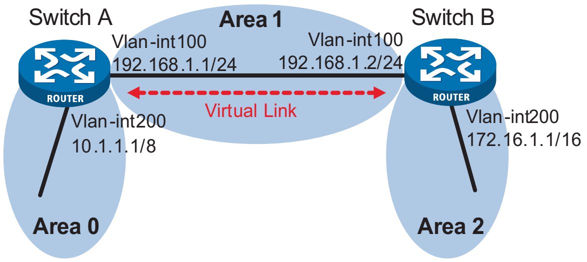
Network requirements

In the following figure, Area2 has no direct connection to Area0, and Area1 acts as the Transit Area to connect Area2 to Area0 via a configured virtual link between SwitchB and SwitchC.

After configuration, SwitchA can learn routes to Area2.

Network diagram

Figure 107 Network diagram for OSPF virtual link configuration



Configuration procedure

1 Configure IP addresses for interfaces (omitted)

2 Configure OSPF basic functions

# Configure SwitchA

```
<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 10.0.0.0 0.255.255.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
```

# Configure SwitchB

```
<SwitchB> system-view
[SwitchB] ospf 1 router-id 2.2.2.2
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.1] quit
[SwitchB-ospf-1] area 2
[SwitchB-ospf-1-area-0.0.0.2] network 172.16.0.0 0.0.255.255
[SwitchB-ospf-1-area-0.0.0.2] quit
```

# Display the OSPF routing table on SwitchA

```
[SwitchA] display ospf routing
          OSPF Process 1 with Router ID 1.1.1.1
                   Routing Tables

 Routing for Network
 Destination         Cost      Type     NextHop          AdvRouter     Area
 10.0.0.0/8          1         Stub     10.1.1.1         1.1.1.1       0.0.0.0
 192.168.1.0/24      1562      Stub     192.168.1.1      1.1.1.1       0.0.0.1

 Total Nets: 2
 Intra Area: 2  Inter Area: 0  ASE: 0  NSSA: 0
```

Since Area2 has no direct connection to Area0, the routing table of RouterA has no route to Area2.

3 Configure a virtual link

# Configure SwitchA

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] vlink-peer 2.2.2.2
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

# Configure SwitchB

```
[SwitchB] ospf 1
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] vlink-peer 1.1.1.1
[SwitchB-ospf-1-area-0.0.0.1] quit
```

# Display the OSPF routing table on SwitchA.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 1.1.1.1
                   Routing Tables

 Routing for Network
 Destination        Cost  Type       NextHop         AdvRouter     Area
 172.16.1.1/16      1563  Inter      192.168.1.2     2.2.2.2       0.0.0.0
 10.0.0.0/8         1     Stub       10.1.1.1        1.1.1.1       0.0.0.0
 192.168.1.0/24     1562  Stub       192.168.1.1     1.1.1.1       0.0.0.1

 Total Nets: 3
 Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

Switch A has learned the route 172.16.1.1/16 to Area2.

Troubleshooting OSPF Configuration

No OSPF Neighbor Relationship Established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower 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), network type. If the network type is broadcast or NBMA, at least one interface must have a DR priority higher than 0.

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.

Processing steps

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 to which 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.

31 IPv6 OSPFv3 Configuration

The term "router" refers to a router in a generic sense or an Ethernet switch running routing protocols in this document.

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

Introduction to OSPFv3

OSPFv3 Overview

OSPFv3 is OSPF (Open Shortest Path First) version 3 for short, supporting IPv6 and compliant with RFC2740 (OSPF for IPv6).

Identical parts 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)
- Mechanisms for finding neighbors and establishing adjacencies
- Mechanisms for LSA flooding and aging

Differences between OSPFv3 and OSPFv2:

- OSPFv3 now 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.

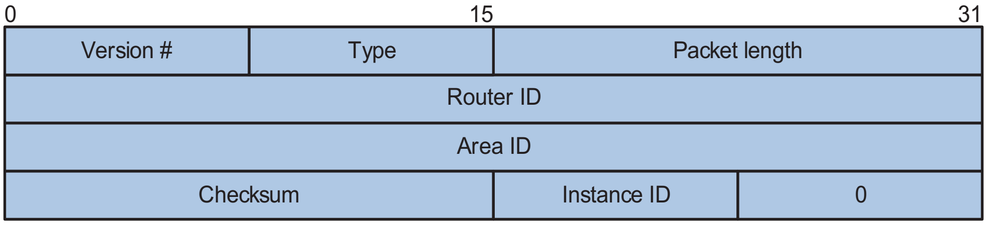
OSPFv3 Packets

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 108 gives the OSPFv3 packet header.

Figure 108 OSPFv3 packet header



Major fields:

- Version #: Version of OSPF, which is 3 for OSPFv3.
- Type: Type of OSPF packet, from 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, which must be 0.

OSPFv3 LSA Types

OSPFv3 sends routing information in LSAs, which as defined in RFC2740 have the following types:

- Router-LSAs: 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-LSAs: 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-LSAs: 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-LSAs: 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-LSAs: 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-LSAs: 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-LSAs: 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.

Timers of OSPFv3

Timers in OSPFv3 include:

- OSPFv3 packet timer
- LSA delay timer
- SPF 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 after a period, it will declare the peer is down. The period is called dead interval.

After sending an LSA to its adjacency, a router waits for an acknowledgment from the adjacency. If no response is received after retransmission interval elapses, the router will send again the LSA. The retransmission interval must be longer than the round-trip time of the LSA in between.

LSA delay time

Each LSA has an age in the local LSDB (incremented by 1 per second), but an LSA is not aged 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 LSDB changes, SPF recalculation happens. If recalculations become so frequent, a large amount of resources will be occupied, reducing operation efficiency of routers. You can adjust SPF calculation interval and delay time to protect networks from being overloaded due to frequent changes.

OSPFv3 Features Supported

- Basic features defined in RFC2740
- OSPFv3 stub area
- OSPFv3 multi-process, which enable a router to run multiple OSPFv3 processes

Related RFCs

- RFC2740: OSPF for IPv6
- RFC2328: OSPF Version 2

IPv6 OSPFv3 Configuration Task List

To configure OSPFv3, perform the tasks described in the following sections:

Task		Description
"Configuring OSPFv3 Basic Functions" on page 356		Required
"Configuring OSPFv3 Area Parameters" on page 357	"Configuring an OSPFv3 Stub Area" on page 357	Optional
	"Configuring OSPFv3 Virtual Links" on page 357	Optional

Task		Description
"Configuring OSPFv3 Routing Information Management" on page 358	"Configuring OSPFv3 Route Summarization" on page 358	Optional
	"Configuring OSPFv3 Inbound Route Filtering" on page 358	Optional
	"Configuring Link Costs for OSPFv3 Interfaces" on page 359	Optional
	"Configuring the Maximum Number of OSPFv3 Load-balanced Routes" on page 359	Optional
	"Configuring a Priority for OSPFv3" on page 359	Optional
	"Configuring OSPFv3 Route Redistribution" on page 359	Optional
"Tuning and Optimizing an OSPFv3 Network" on page 360	"Configuring OSPFv3 Timers" on page 360	Optional
	"Configuring the DR Priority for an Interface" on page 361	Optional
	"Ignoring MTU Check for DD Packets" on page 361	Optional
	"Disable Interfaces from Sending OSPFv3 Packets" on page 362	Optional
	"Enable the Logging on Neighbor State Changes" on page 362	Optional

Configuring OSPFv3 Basic Functions

Prerequisites

- Make neighboring nodes accessible with each other at network layer.
- Enable IPv6 packet forwarding

Configuring OSPFv3 Basic Functions

To configure OSPFv3 basic functions, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	**-**
Enable OSPFv3 and enter its view	**ospfv3** [*process-id*]	Required
Specify a router ID	**router-id** *router-id*	Required
Enter interface view	**interface** *interface-type interface-number*	-
Enable OSPFv3 on the interface	**ospfv3** *process-id* **area** *area-id* [**instance** *instance-id*]	Required Not enabled by default

- *Configure an OSPFv3 process ID when enabling OSPFv3. The process ID takes effect locally, without affecting packet exchange between routers.*

- *When configuring a router ID, make sure each router has a unique ID. If a router runs multiple OSPFv3 processes, you need to specify a router ID for each process.*

Configuring OSPFv3 Area Parameters

The stub area and virtual link support of OSPFv3 has the same principle and application environments with OSPFv2.

Splitting an OSPFv3 AS into multiple areas reduces the number of LSAs on networks and extends OSPFv3 application. 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 on routers in these areas 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 maintain connectivity. In practice, necessary physical links may not be available for connectivity. You can configure virtual links to address it.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring an OSPFv3 Stub Area

To configure an OSPFv3 stub area, use the following commands:

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
Configure the default route cost of sending a packet to the stub area	**default-cost** *value*	Optional Defaults to 1

- *Configurations on routers attached to the same area should be compatible to avoid information exchange failures, information block and routing loops.*
- *You cannot delete 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 automatically.*
- *All routers attached to a stub area must be configured with the **stub** command. The keyword **no-summary** is only available on the ABR.*
- *If you use the **stub** command with the keyword **no-summary** on an ABR, the ABR distributes a default summary LSA into the area rather than generating an AS-external-LSA or Inter-Area-Prefix-LSA. The stub area of this kind is also known as totally stub area.*

Configuring OSPFv3 Virtual Links

You can configure virtual links to maintain connectivity between non-backbone areas and the backbone, or in the backbone itself.

To configure a virtual link, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter OSPFv3 view	**ospfv3** [*process-id*]	-
Enter OSPFv3 area view	**area** *area-id*	-
Create and configure a virtual link	**vlink-peer** *router-id* [**hello** *seconds* | **retransmit** *seconds* | **trans-delay** *seconds* | **dead** *seconds* | **instance** *instance-id*] *	Required

*Both ends of a virtual link are ABRs that are configured with the **vlink-peer** command.*

Configuring OSPFv3 Routing Information Management

This section is to configure management of OSPF routing information advertisement and reception, and route redistribution from other protocols.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring OSPFv3 Route Summarization

To configure route summarization between areas, use the following command on an ABR:

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

*The **abr-summary** command is available on ABRs only. If contiguous network segments are available in an area, you can use the 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.*

Configuring OSPFv3 Inbound Route Filtering

You can configure OSPFv3 to filter routes that are computed from received LSAs according to some rules.

To configure inbound route filtering, use the following commands:

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

*Use of the **filter-policy import** command can only filter routes computed by OSPFv3. Only routes not filtered can be added into the local routing table.*

Configuring Link Costs for OSPFv3 Interfaces

You can configure OSPFv3 link costs for interfaces to adjust routing calculation.

To configure the link cost for an OSPFv3 interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure the cost for the interface	**ospfv3 cost** *value* [**instance** *instance-id*]	Optional By default, OSPFv3 computes an interface's cost according to the bandwidth on it. The cost value defaults to 1 for VLAN interfaces of switches.

Configuring the Maximum Number of OSPFv3 Load-balanced Routes

If multiple routes to a destination are available, using load balancing to send IPv6 packets on these routes in turn can improve link utility. To configure the maximum number of load-balanced routes, use the following commands:

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

Configuring a Priority for OSPFv3

A routing device 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.

To configure a priority for OSPFv3, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter OSPFv3 view	**ospfv3** [*process-id*]	-
Configure a priority for OSPFv3	**preference** [**ase**] [**route-policy** *route-policy-name*] *preference*	Optional By default, the priority of OSPFv3 interval routes is 10, and priority of OSPFv3 external routes is 150.

Configuring OSPFv3 Route Redistribution

To configure OSPFv3 route redistribution, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter OSPFv3 view	**ospfv3** [*process-id*]	-

To do...	Use the command...	Remarks
Specify a default cost for redistributed routes	**default cost** *value*	Optional Defaults to 1
Redistribute routes from another protocol or from another OSPFv3 process	**import-route** { **isisv6** *process-id* | **ospfv3** *process-id* | **ripng** *process-id* | **bgp4+** [**allow-ibgp**] | **direct** | **static** } [**cost** *value* | **type** { **1** | **2** } | **tag** *value* | **route-policy** *route-policy-name*] *	Optional
Configure the filtering of 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

- *Using the **import-route** command on a router makes the router become an ASBR.*
- *Since OSPFv3 is a link state based routing protocol, it cannot directly filter LSAs to be advertised. Therefore, you need to configure filtering redistributed routes before advertising routes that are not filtered in LSAs into the routing domain.*
- *Use of the **filter-policy export** command takes effect only on the local router. However, if the **import-route** command is not configured, executing the **filter-policy export** command does not take effect.*

Tuning and Optimizing an OSPFv3 Network

This section describes configurations of OSPFv3 timers, interface DR priority, MTU check ignorance for DD packets, 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 consumed 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

To configure OSPFv3 timers, use the following commands:

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 hello interval	**ospfv3 timer hello** *seconds* [**instance** *instance-id*]	Optional Defaults to 10 seconds on P2P, broadcast interfaces
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 timer	**spf timers** *delay-interval hold-interval*	Optional By default, *delay-interval* is 5 seconds, and *hold-interval* is 10 seconds

- *The dead interval set on neighboring interfaces cannot be so small. Otherwise, a neighbor is so easy to be considered as down.*
- *The LSA retransmission interval cannot be so small to avoid unnecessary retransmissions.*

Configuring the DR Priority for an Interface

To configure the DR priority for an interface, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	-
Configure the DR priority	**ospfv3 dr-priority** *priority* [**instance** *instance-id*]	Optional Defaults to 1

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 MTU in DD packets in order to improve efficiency.

To ignore MTU check for DD packets, use the following commands:

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 After this command is configured, the interface will not check the MTU field of incoming DD packets.

Disable Interfaces from Sending OSPFv3 Packets

To disable interfaces from sending OSPFv3 packets, use the following commands:

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

- *Multiple processes can disable the same interface from sending OSPFv3 packets. Using 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, 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 on Neighbor State Changes

To enable the logging on neighbor state changes, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter OSPFv3 view	**ospfv3** [*process-id*]	-
Enable the logging on neighbor state changes	**log-peer-change**	Required Enabled 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 lsdb** [[**external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router**] [*link-state-id*] [**originate-router** *ip-address*] | **statistic** | **total**] **display ospfv3** *process-id* **lsdb** [[**external** | **inter-prefix** | **inter-router** | **intra-prefix** | **link** | **network** | **router**] [*link-state-id*] [**originate-router** *ip-address*] | **total**]	
Display OSPFv3 neighbor information	**display ospfv3** [*process-id*] [**area** *area-id*] **peer** [**statistic** | [*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** [**statistics**]	
Display OSPFv3 link state retransmission list information	**display ospfv3** [*process-id*] **retrans-list** [**statistics**]	
Display OSPFv3 statistics	**display ospfv3 statistic**	

OSPFv3 Configuration Examples

Configuring OSPFv3 Areas

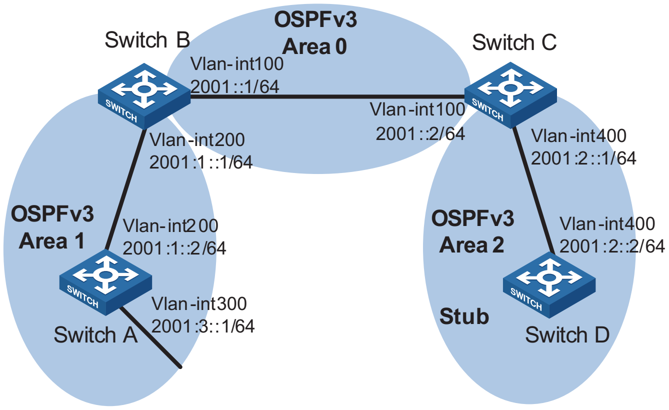
Network requirements

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, reducing LSAs into the area without affecting route reachability.

Network diagram

Figure 109 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,    *  - Selected 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, whose cost is the cost of the directly connected 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,    *  - Selected 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 Switch C, the ABR, to make Area 2 as a totally stub area.

```
[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,    *  - Selected 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

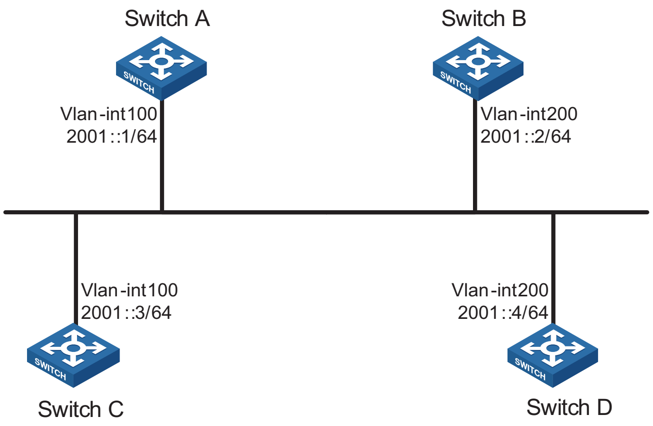
In the following figure:

- 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.
- RouterD has the default priority 1.

Tunnels must be created between switches to configure OSPFv3 and form OSPFv3 neighbor relationships. For related information, refer to “BPDU Tunneling Configuration” on page 149.

Network diagram

Figure 110 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 between Switch D and other switches 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-interface100 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 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 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
```

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, 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 assoiated 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.

32 DUAL STACK CONFIGURATION

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

- "Dual Stack Overview" on page 373
- "Configuring Dual Stack" on page 373

The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

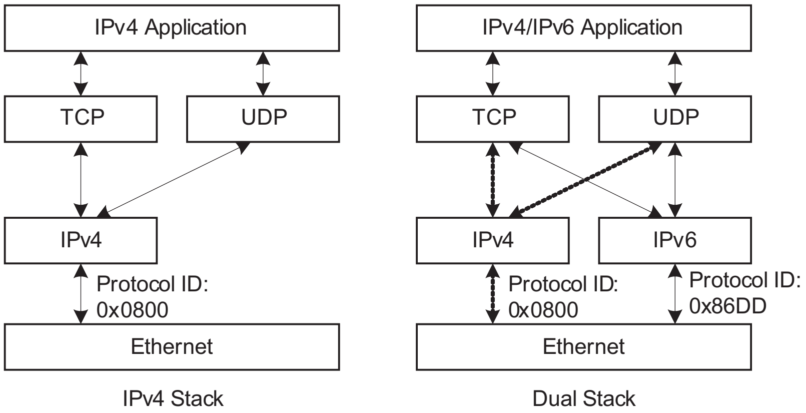
Dual Stack Overview

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 111 illustrates the IPv4/IPv6 dual stack in relation to the IPv4 stack.

Figure 111 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*	-
Configure an IPv4 address for the interface			**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Required By default, no IP address is configured. The support for the **sub** keyword varies with devices.
Configure an IPv6 address on the interface	Configure IPv6 global unicast address or local address	Manually specify an IPv6 address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address/prefix-length* }	Use either command By default, no 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 IPv6 link-local address	Automatically create an IPv6 link-local address	**ipv6 address auto link-local**	Optional By default, after you configured an IPv6 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**	

CAUTION: *For more information about IPv6 address, refer to "Configuring Basic IPv6 Functions" on page 221.*

33 TUNNELING CONFIGURATION

Introduction to Tunneling

The expansion of Internet results in scarce IPv4 addresses. Although the techniques such as temporary IPv4 address allocation and network address translation (NAT) relieve the problem of IPv4 address shortage to some extent, they not only increase the overhead in address resolution and processing, but also lead to high-level 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, the network using the IPv6 protocol stack is expected to communicate with the Internet using IPv4. Therefore, an IPv6-IPv4 interworking technique must be developed to ensure the smooth transition from IPv4 to IPv6. In addition, the interworking technique should provide efficient, seamless information transfer. The Internet Engineering Task Force (IETF) set 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 techniques 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 techniques: dual stack (RFC 2893), tunneling (RFC 2893), and NAT-PT (RFC 2766).

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. 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.

For related configuration about the dual protocol stack, refer to "Dual Stack Overview" on page 373.

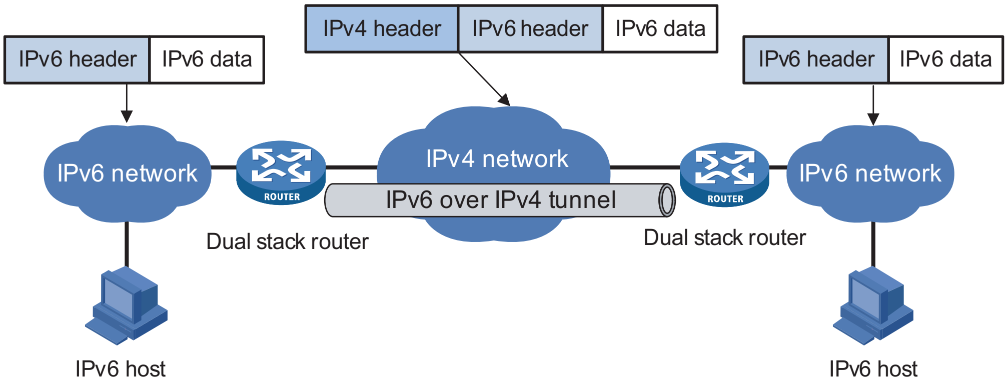
IPv6 over IPv4 Tunnel

Principle

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 112.

***CAUTION:** The devices at both ends of an IPv6 over IPv4 tunnel must support IPv4/IPv6 dual stack.*

Figure 112 Principle of 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 an IPv4 header in the IPv6 packet 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 device at the destination end of the tunnel forwards the packet according to the destination address in the decapsulated IPv6 packet. If the destination address is the device itself, the device at the destination end 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 eventual destination of the IPv6 packet.

According to the way the IPv4 address of the tunnel destination is acquired, tunnels are divided into configured tunnel and automatic tunnel.

- The tunnel destination IPv4 address cannot be acquired from the destination address of the IPv6 packet and it needs to be configured manually. Such a tunnel is called configured tunnel.
- If the tunnel destination is just the eventual destination of the IPv6 packet, an IPv4 address can be embedded into an IPv6 address so that the IPv4 address of the tunnel destination can automatically be acquired from the destination address of the IPv6 packet. Such a tunnel is called automatic tunnel.

Type

According to the way an IPv6 packet is encapsulated, IPv6 over IPv4 tunnels are divided into the following types:

- IPv6 manually configured tunnel
- Automatic IPv4-compatible IPv6 tunnel
- 6to4 tunnel
- ISATAP tunnel
- IPv6-over-IPv4 GRE tunnel (GRE tunnel for short)

Among the above tunnels, the IPv6 manually configured tunnel and GRE tunnel are configured tunnels, while the automatic IPv4 compatible IPv6 tunnel, 6to4 tunnel, and intra-site automatic tunnel address protocol (ISATAP) tunnel are automatic tunnels.

1 IPv6 manually configured tunnel

A manually configured tunnel is a point-to-point link. One link is a separate tunnel. The IPv6 manually configured tunnel is mainly used for stable connections requiring regular secure communication between two border routers or between a border router and a host, or for connections to remote IPv6 networks.

2 Automatic IPv4-compatible IPv6 tunnel

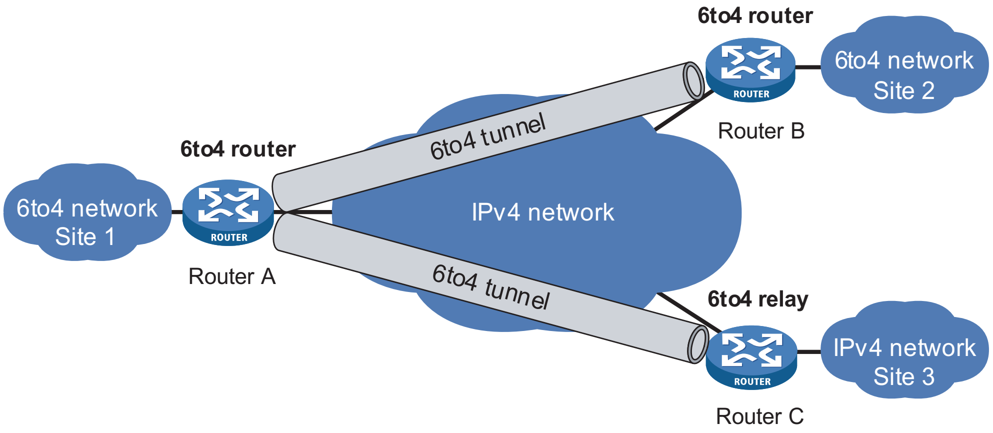
An automatic IPv4-compatible IPv6 tunnel is a point-to-multipoint link. IPv4-compatible IPv6 addresses are adopted at both ends of such a tunnel. The address format is 0:0:0:0:0:0:a.b.c.d/96, where a.b.c.d represents an embedded IPv4 address. The tunnel destination is automatically determined by the embedded IPv4 address, which makes it easy to create a tunnel for IPv6 over IPv4. However, an automatic IPv4-compatible IPv6 tunnel must use IPv4-compatible IPv6 addresses and it is still dependent on IPv4 addresses. Therefore, automatic IPv4-compatible IPv6 tunnels have limitations.

3 6to4 tunnel

A 6to4 tunnel is a point-to-multipoint tunnel and is used to connect multiple isolated IPv6 domains over an IPv4 network to remote IPv6 networks. The embedded IPv4 address in an IPv6 address is used to automatically acquire the destination of the tunnel. The automatic 6to4 tunnel adopts 6to4 addresses. The address format is 2002:abcd:efgh:subnet number::interface ID/64, where abcd:efgh represents the 32-bit source IPv4 address of the 6to4 tunnel, in hexadecimal notation. For example, 1.1.1.1 can be represented by 0101:0101. The tunnel destination is automatically determined by the embedded IPv4 address, which makes it easy to create a 6to4 tunnel.

Since 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 by 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. A 6to4 tunnel interconnects IPv6 networks and overcomes the limitations of an automatic IPv4-compatible IPv6 tunnel.

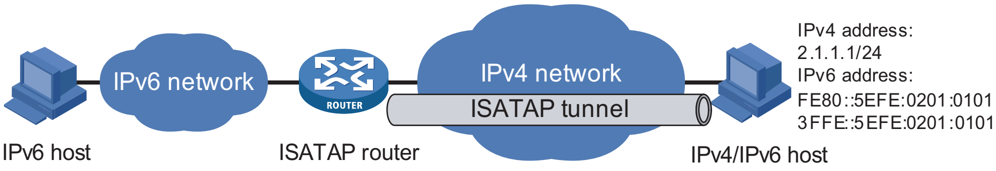
Figure 113 6to4 tunnel



4 ISATAP tunnel

With the application of the IPv6 technique, there will be more and more IPv6 hosts in the existing IPv4 network. The ISATAP tunneling technique 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 addresses: ISATAP addresses. The ISATAP address format is prefix (64bit):0:5EFE:ip-address. The ip-address is in the form of a.b.c.d or abcd:efgh, where abcd:efgh represents a 32-bit source IPv4 address. 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 host and IPv6 router.

Figure 114 ISATAP tunnel



5 GRE tunnel

IPv6 packets can be carried over GRE tunnels to pass through the IPv4 network by using standard GRE protocol to encapsulate them. Like the IPv6 manually configured tunnel, a GRE tunnel is a point-to-point link, too. Each link is a separate tunnel. The GRE tunnel is mainly used for stable connections requiring regular secure communication between two border routers or between a host and a border router. For related configurations, refer to *"GRE Configuration" on page 405*.

Expedite termination

With expedite termination disabled, a tunnel packet arriving at the destination node is first forwarded to the CPU for processing, then the outer IPv4 header is removed, and finally the decapsulated original packet is forwarded. With expedite termination enabled, the tunnel packet is unnecessarily sent to the CPU for processing, but is directly processed as IPv6 packets.

- If the source IP address of the tunnel packet matches the expedite termination subnet, the packet is sent to the IPv6 switch fabric to forward or sent to the CPU for processing.
- If the tunnel packet needs to be forwarded, the IPv6 switch fabric decapsulates the tunnel packet to obtain the original IPv6 packet and then forwards it directly.

The expedite termination function solves the problem that the rate of tunnel packets is restricted by the loopback port in the tunnel service.

With expedite termination enabled, IPv6 packets to be encapsulated still need to be sent to the service loopback port for processing.

Tunnel hybrid insertion

In practice, many cards only support IPv4. However, a tunnel can only be established over IPv6 cards. After tunnel packets arrive on the destination node, it is very likely that an IPv4 card received the packets. The tunnel hybrid insertion function enables IPv4 cards to support the tunnel termination. Through the function, tunnel packets can be terminated without obstruction on the destination node. This function is implemented by configuring an ACL on incoming interfaces of IPv4 cards to redirect tunnel packets to IPv6 cards.

CAUTION: *In the case of tunnel hybrid insertion, the outbound interface of tunnel packets must support IPv6 if expedite termination is enabled. Otherwise, tunnel packets cannot be encapsulated or decapsulated.*

IPv4 over IPv4 Tunnel

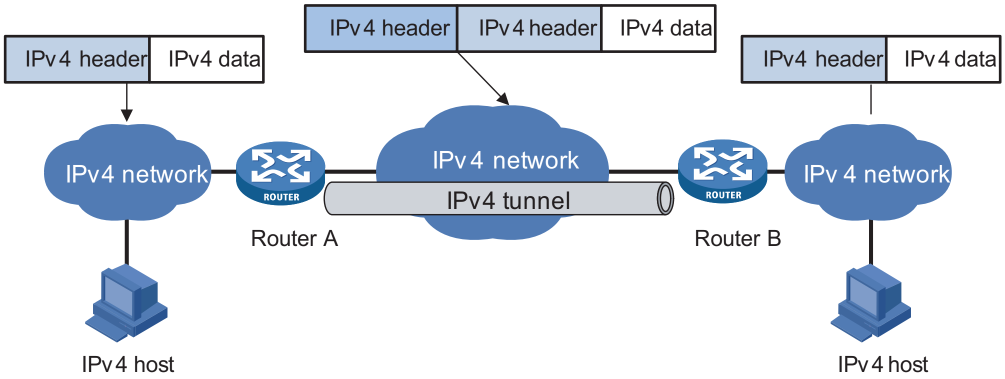
Introduction to IPv4 over IPv4 tunneling protocol

IPv4 over IPv4 tunneling protocol (RFC 1853) is developed for IP data packet encapsulation so that data can be transferred from one IPv4 network to another IPv4 network.

Encapsulation and decapsulation

Packets to be transferred through a tunnel undergo an encapsulation process and decapsulation process. Figure 115 shows these two processes.

Figure 115 Principle of IPv4 over IPv4 tunnel



Tunneling Configuration Task List

Complete these tasks to configure the tunneling feature:

Task		Remarks
Configuring IPv6 over IPv4 GRE tunnel	"Configuring IPv6 Manually Configured Tunnel" on page 380	Optional
	"Configuring Automatic IPv4-Compatible IPv6 Tunnel" on page 384	Optional
	"Configuring 6to4 Tunnel" on page 387	Optional
	"Configuring ISATAP Tunnel" on page 391	Optional
"Configuring IPv4 over IPv4 Tunnel" on page 395		Optional
"Configuring Tunnel Hybrid Insertion" on page 399		Optional

When NAT is also enabled on the VLAN interface serving as the tunnel source interface, if possible, you must enable expedite termination on the tunnel interface to ensure the availability of these two services.

Configuring IPv6 Manually Configured Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface, Ethernet interface, and loopback interface on the device so that they can communicate. These interfaces serve as the source interface of a tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure an IPv6 manually configured tunnel:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the IPv6 packet forwarding function	**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.

To do...		Use the command...	Remarks
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-address***/***prefix-length* }	Required Use any 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 a link-local IPv6 address	**ipv6 address auto link-local**	Optional 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**	
Configure the tunnel to be an IPv6 manually configured tunnel		**tunnel-protocol ipv6-ipv4**	Required By default, the tunnel is a GRE tunnel. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or source interface for the tunnel		**source** { *ip-address* | *ipv6-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.
Configure a link aggregation group ID to be referenced by the tunnel interface		**aggregation-group** *aggregation-group-id*	Required
Enable the expedite termination function		**expediting enable**	Optional By default, the expedite termination function is disabled.
Configure the MTU of IPv6 packets sent over a tunnel interface		**ipv6 mtu** *mtu-size*	Optional

For the configuration of tunnel interface MTU, refer to the ***ipv6 mtu*** *command in the Switch 8800 Command Reference Guide.*

CAUTION:

- *After a tunnel interface is deleted, all the above features configured on the tunnel interface will be deleted.*
- *If the tunnel interface addresses 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 can configure static or dynamic routes. IP addresses must be configured at both ends of the tunnel. For detailed configuration, refer to "IP Routing and Routing Table" on page 187.*

- *When configuring a static route, you need to configure a route to the destination address (the destination IPv6 address of the packet, instead of the tunnel destination IPv4 address) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. Such configurations must be performed at both ends of the tunnel.*
- *Before configuring dynamic routes, you must enable the dynamic routing protocol on the tunnel interfaces at both ends. For configurations, refer to "Static Routing and Dynamic Routing" on page 189.*
- *The interfaces of an IPv6 manually configured tunnel support dynamic routing protocols such as OSPFv3, RIPng, and BGP4+.*
- *When configuring a dynamic routing protocol other than BGP4+ on tunnel interfaces, you need to enable expedite termination on the tunnel interfaces.*
- *The destination address of the route configured on the tunnel interface and the address of the tunnel interface must not be in the same network segment.*
- *Two or more tunnel interfaces using the same encapsulation protocol must have different source and destination addresses.*

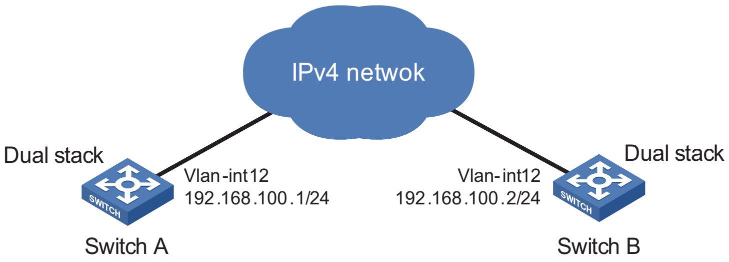
Configuration Example

Network requirements

Two IPv6 networks are connected through an IPv6 manually configured tunnel between Switch A and Switch B. As shown in Figure 116, the interface VLAN-interface 12 on Switch A can communicate with the interface VLAN-interface 12 on Switch B and an IPv4 packet route is available between.

Network diagram

Figure 116 Network diagram for an IPv6 manually configured tunnel (on switches)



Configuration procedure

The following example shows how to configure an IPv6 manually configured tunnel between Switch A and Switch B. Before configuration, you must specify IP addresses for the source and destination of the tunnel.

1 Configure Switch A

# Configure an IPv4 address for the interface VLAN-interface 12.

```
<SwitchA> system-view
[SwitchA] vlan 12
[SwitchA-vlan12] port GigabitEthernet 3/1/1
[SwitchA-vlan12] quit
[SwitchA] interface vlan-interface 12
[SwitchA-vlan-interface12] ip address 192.168.100.1 255.255.255.0
[SwitchA-vlan-interface12] quit
```

# Enable the IPv6 forwarding function.

```
[SwitchA] ipv6
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 3/1/2
[SwitchA-GigabitEthernet3/1/2] stp disable
[SwitchA-GigabitEthernet3/1/2] port link-aggregation group 1
[SwitchA-GigabitEthernet3/1/2] quit
```

# Configure an IPv6 manually configured tunnel.

```
[SwitchA] interface tunnel 0/0/1
[SwitchA-Tunnel0/0/1] ipv6 address 3001::1 64
[SwitchA-Tunnel0/0/1] source vlan-interface 12
[SwitchA-Tunnel0/0/1] destination 192.168.100.2
[SwitchA-Tunnel0/0/1] tunnel-protocol ipv6-ipv4
```

# Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[SwitchA-Tunnel0/0/1] aggregation-group 1
[SwitchA-Tunnel0/0/1] expediting enable
[SwitchA-Tunnel0/0/1] quit
```

# Configure a static route from the interface Tunnel 0/0/1 of Switch A to Switch B.

```
[SwitchA] ipv6 route-static 2::0 64 tunnel 0/0/1
```

2 Configure Switch B.

# Configure an IPv4 address for the interface VLAN-interface 12.

```
<SwitchB> system-view
[SwitchB] vlan 12
[SwitchB-vlan12] port GigabitEthernet 3/1/1
[SwitchB-vlan12] quit
[SwitchB] interface Vlan-interface 12
[SwitchB-Vlan-interface12] ip address 192.168.100.2 255.255.255.0
[SwitchB-Vlan-interface12] quit
```

# Enable the IPv6 forwarding function.

```
[SwitchB] ipv6
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchB] link-aggregation group 2 mode manual
[SwitchB] link-aggregation group 2 service-type tunnel
[SwitchB] interface GigabitEthernet 3/1/2
[SwitchB-GigabitEthernet3/1/2] stp disable
[SwitchB-GigabitEthernet3/1/2] port link-aggregation group 2
[SwitchB-GigabitEthernet3/1/2] quit
```

#Configure an IPv6 manually configured tunnel.

```
[SwitchB] interface tunnel0/0/1
[SwitchB-Tunnel0/0/1] ipv6 address 3001::2 64
[SwitchB-Tunnel0/0/1] source vlan-interface 12
[SwitchB-Tunnel0/0/1] destination 192.168.100.1
[SwitchB-Tunnel0/0/1] tunnel-protocol ipv6-ipv4
```

# Reference link aggregation group 2 and enable expedite termination in tunnel interface view

```
[SwitchB] interface tunnel 0/0/1
[SwitchB-Tunnel0/0/1] aggregation-group 2
[SwitchB-Tunnel0/0/1] expediting enable
[SwitchB-Tunnel0/0/1] quit
```

# Configure a static from the interface Tunnel0/0/1 of Switch B to Switch A.

```
[SwitchB] ipv6 route-static 1::0 64 tunnel 0/0/1
```

Configuration verification

After the above configurations, you can successfully ping the IPv6 address of the peer tunnel interface from one switch.

Configuring Automatic IPv4-Compatible IPv6 Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and Loopback interface on the device so that they can communicate. These interfaces serve as the source interface of the virtual tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure an automatic IPv4-compatible IPv6 tunnel:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable the IPv6 packet forwarding function		**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 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 or site-local address is configured for the interface.
		ipv6 address *ipv6-address* **link-local**	

To do...	Use the command...	Remarks
Configure an automatic IPv4-compatible IPv6 tunnel	**tunnel-protocol ipv6-ipv4 auto-tunnel**	Required By default, the tunnel is a GRE tunnel. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address for the tunnel	**source** { *ip-address* | *ipv6-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Configure a link aggregation group ID to be referenced by the tunnel interface	**aggregation-group** *aggregation-group-id*	Required
Enable the expedite termination function	**expediting enable**	Optional By default, the expedite termination function is disabled.
Configure an address and mask for the expedite termination subnet	**expediting subnet** *ip-address mask*	Optional By default, no expedite termination subnet is configured for a tunnel.
Configure a tunnel interface MTU	**mtu** *mtu-size*	Optional

For the configuration of the tunnel interface MTU, refer to the ***ipv6 mtu*** *command in the Switch 8800 Command Reference Guide.*

CAUTION:

- *For automatic IPv4-compatible IPv6 tunnels, 6to4 tunnels, or ISATAP tunnels, their tunnel interfaces must have different source addresses.*
- *No destination address needs to be configured for an automatic IPv4-compatible IPv6 tunnel.*
- *If the tunnel interface addresses 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. You can configure static or dynamic routes. A forwarding route needs to be configured at both ends of the tunnel. For detailed configuration, refer to "IP Routing and Routing Table" on page 187.*
- *Automatic IPv4-compatible IPv6 tunnels support only BGP4+.*
- *When you configure a static route, you need to configure a route to the destination address (the destination IP address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. A static route must be configured at both ends of the tunnel.*

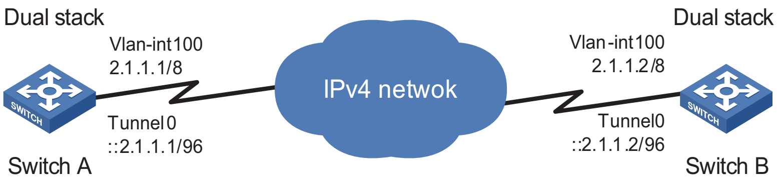
Configuration Example

Network requirements

Between Switch A and Switch B is an IPv4 network. It is required that an IPv6 connection be established through an automatic IPv4-compatible IPv6 tunnel between the two dual-stack switches.

Network diagram

Figure 117 Network diagram for an automatic IPv4-compatible IPv6 tunnel



Configuration procedure

The following example shows how to configure an automatic IPv4-compatible IPv6 tunnel between Switch A and Switch B. No address needs to be specified for the tunnel destination because the tunnel destination address can automatically be obtained from the IPv4 address embedded in the IPv4-compatible IPv6 address.

1 Configure Switch A.

# Enable the IPv6 forwarding function.

```
<SwitchA> system-view
[SwitchA] ipv6
```

# Configure an IPv4 address for the interface VLAN-interface 12.

```
[SwitchA] vlan 12
[SwitchA-vlan12] port GigabitEthernet3/1/1
[SwitchA-vlan12] quit
[SwitchA] interface Vlan-interface 12
[SwitchA-Vlan-interface 12] ip address 2.1.1.1 255.0.0.0
[SwitchA-Vlan-interface 12] quit
```

# Configure an automatic IPv4-compatible IPv6 tunnel.

```
[SwitchA] interface tunnel 0/0/1
[SwitchA-Tunnel0/0/1] ipv6 address ::2.1.1.1/96
[SwitchA-Tunnel0/0/1] source Vlan-interface 12
[SwitchA-Tunnel0/0/1] tunnel-protocol ipv6-ipv4 auto-tunnel
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 3/1/2
[SwitchA-GigabitEthernet3/1/2] stp disable
[SwitchA-GigabitEthernet3/1/2] port link-aggregation group 1
[SwitchA-GigabitEthernet3/1/2] quit
```

# Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[SwitchA] interface tunnel 0/0/1
[SwitchA-Tunnel0/0/1] aggregation-group 1
[SwitchA-Tunnel0/0/1] expediting enable
[SwitchA-Tunnel0/0/1] expediting subnet 2.1.1.0 255.0.0.0
[SwitchA-Tunnel0/0/1] quit
```

2 Configure Switch B

# Enable the IPv6 forwarding function.

```
<SwitchB> system-view
[SwitchB] ipv6
```

# Configure an IPv4 address for the interface VLAN-interface 12.

```
[SwitchB] vlan 12
[SwitchB-vlan12] port GigabitEthernet 3/1/1
[SwitchB] interface Vlan-interface 12
[SwitchB-GigabitEthernet3/1/1] ip address 2.1.1.2 255.0.0.0
[SwitchB-GigabitEthernet3/1/1] quit
```

# Configure an automatic IPv4-compatible IPv6 tunnel.

```
[SwitchB] interface tunnel 0/0/1
[SwitchB-Tunnel0/0/1] ipv6 address ::2.1.1.2/96
[SwitchB-Tunnel0/0/1] source Vlan-interface 12
[SwitchB-Tunnel0/0/1] tunnel-protocol ipv6-ipv4 auto-tunnel
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 3/1/2
[SwitchB-GigabitEthernet3/1/2] stp disable
[SwitchB-GigabitEthernet3/1/2] port link-aggregation group 1
[SwitchB-GigabitEthernet3/1/2] quit
```

# Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[SwitchB] interface tunnel 0/0/1
[SwitchB]-Tunnel0/0/1] aggregation-group 1
[SwitchB-Tunnel0/0/1] expediting enable
[SwitchB-Tunnel0/0/1] expediting subnet 2.1.1.0 255.0.0.0
[SwitchB-Tunnel0/0/1] quit
```

Configuration verification

After the above configurations, you can successfully ping the IPv4-compatible IPv6 address of the peer tunnel interface from one switch.

Configuring 6to4 Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and Loopback interface on the device so that they can communicate. These interfaces serve as the source interface of the virtual tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure a 6to4 tunnel:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the IPv6 packet forwarding function	**ipv6**	Required By default, the IPv6 packet forwarding function is disabled.

To do...		Use the command...	Remarks
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 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 a GRE tunnel. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address for the tunnel		**source** { *ip-address* | *ipv6-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Configure a link aggregation group ID to be referenced by the tunnel interface		**aggregation-group** *aggregation-group-id*	Required
Enable the expedite termination function		**expediting enable**	Optional By default, the expedite termination function is disabled.
Configure an address and mask for the expedite termination subnet		**expediting subnet** *ip-address mask*	Optional By default, no expedite termination subnet is configured for a tunnel.
Configure the tunnel interface MTU		**mtu** *mtu-size*	Optional

For the configuration of the tunnel interface MTU, refer to the ***ipv6 mtu*** *command in the Switch 8800 Command Reference Guide.*

CAUTION:

- *For automatic IPv4-compatible IPv6 tunnels, 6to4 tunnels, or ISATAP tunnels, their tunnel interfaces must have different source addresses.*
- *Two or more tunnel interfaces using the same encapsulation protocol must have different source and destination addresses.*

- *No destination address needs to be configured for an automatic tunnel because the destination address can automatically be obtained from the IPv4 address embedded in the IPv4-compatible IPv6 address.*
- *If the tunnel interface addresses 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. You can configure static or dynamic routes. A forwarding route needs to be configured at both ends of the tunnel. For the detailed configuration, refer to "IP Routing and Routing Table" on page 187.*
- *6to4 tunnels support only BGP4+.*
- *When you configure a static route, you need to configure a route to the destination address (the destination IP address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. A static route must be configured at both ends of the tunnel.*

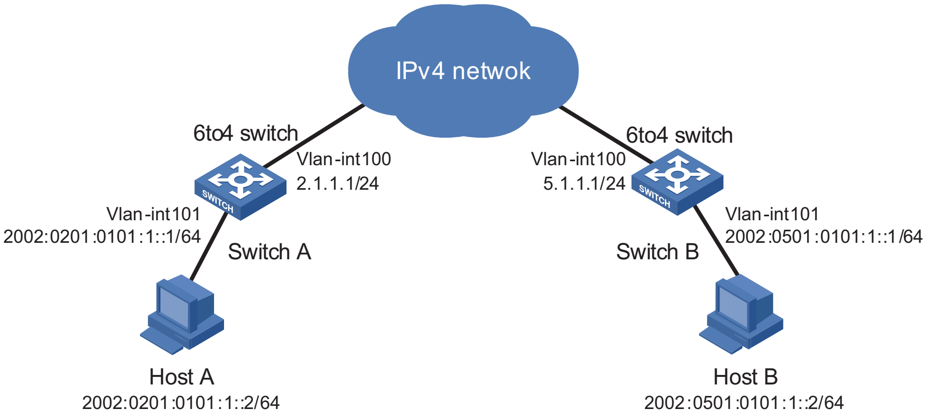
Configuration Example

Network requirements

Isolated IPv6 domains are interconnected through a 6to4 tunnel established in the IPv4 network.

Network diagram

Figure 118 Network diagram for a 6to4 tunnel



Configuration procedure

The following example shows how to configure a 6to4 tunnel between border switches on isolated IPv6 networks. After the IPv4 address 2.1.1.1 is converted into an IPv6 address, the address prefix is 2002:0201:0101::/64. The configured static route directs all traffic destined for the IPv6 address with the prefix 2002::/16 to the tunnel interface of the 6to4 tunnel.

1 Configure Switch A

# Enable the IPv6 forwarding function.

```
<SwitchA> system-view
[SwitchA] ipv6
```

# Configure an IPv4 address for the interface VLAN-interface 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/1/1
```

```
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 2.1.1.1 24
[SwitchA-Vlan-interface100] quit
```

# Configure a route from the interface VLAN-interface 100 to the interface VLAN-interface 100 of Switch B. (Here the next-hop address of the static route is represented by [nexthop]. In practice, you should configure the real next-hop address according to the network.)

```
[SwitchA] ip route-static 5.1.1.1 24 [nexthop]
```

# Configure an IPv6 address for the interface VLAN-interface 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port GigabitEthernet 1/1/2
[SwitchA-vlan101] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 2002:0201:0101:1::1/64
[SwitchA-Vlan-interface101] quit
```

# Configure a 6to4 tunnel.

```
[SwitchA] interface tunnel 0/0/1
[SwitchA-Tunnel0/0/1] ipv6 address 2002:201:101::1 64
[SwitchA-Tunnel0/0/1] source vlan-interface 100
[SwitchA-Tunnel0/0/1] tunnel-protocol ipv6-ipv4 6to4
[SwitchA-Tunnel0/0/1] quit
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 1/1/3
[SwitchA-GigabitEthernet1/1/3] stp disable
[SwitchA-GigabitEthernet1/1/3] port link-aggregation group 1
[SwitchA-GigabitEthernet1/1/3] quit
```

Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[SwitchA] interface tunnel 0/0/1
[SwitchA-Tunnel0/0/1] aggregation-group 1
[SwitchA-Tunnel0/0/1] expediting enable
[SwitchA-Tunnel0/0/1] expediting subnet 5.1.1.0 255.0.0.0
[SwitchA-Tunnel0/0/1] 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 0/0/1
```

2 Configure Switch B

# Enable the IPv6 forwarding function.

```
<SwitchB> system-view
[SwitchB] ipv6
```

# Configure an IPv4 address for the interface VLAN-interface 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/1/1
[SwitchB-vlan100] quit
[SwitchB] interface vlan-interface 100
```

```
[SwitchB-Vlan-interface100] ip address 5.1.1.1 24
[SwitchB-Vlan-interface100] quit
```

# Configure a route from the interface VLAN-interface 100 to the interface VLAN-interface 100 of Switch A. (Here the next-hop address of the static route is represented by [nexthop]. In practice, you should configure the real next-hop address according to the network.)

```
[SwitchB] ip route-static 2.1.1.1 24 [nexthop]
```

# Configure an IPv6 address for the interface VLAN-interface 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port GigabitEthernet 1/1/2
[SwitchB-vlan101] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2002:0501:0101:1::1/64
[SwitchB-Vlan-interface101] quit
```

# Configure a 6to4 tunnel.

```
[SwitchB] interface tunnel0/0/1
[SwitchB-Tunnel0/0/1] ipv6 address 2002:0501:0101::1 64
[SwitchB-Tunnel0/0/1] source vlan-interface 100
[SwitchB-Tunnel0/0/1] tunnel-protocol ipv6-ipv4 6to4
[SwitchB-Tunnel0/0/1] quit
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 1/1/3
[SwitchB-GigabitEthernet1/1/3] stp disable
[SwitchB-GigabitEthernet1/1/3] port link-aggregation group 1
[SwitchB-GigabitEthernet1/1/3] quit
```

# Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[SwitchB] interface tunnel 0/0/1
[SwitchB-Tunnel0/0/1] aggregation-group 1
[SwitchB-Tunnel0/0/1] expediting enable
[SwitchB-Tunnel0/0/1] expediting subnet 2.1.1.0 255.0.0.0
[SwitchB-Tunnel0/0/1] 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 tunnel0
```

Configuration verification

After the above configuration, you can successfully ping Host B from Host A or ping Host A from Host B.

Configuring ISATAP Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and Loopback interface on the device so that they can communicate. These interfaces serve as the source interface of the virtual tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure an ISATAP tunnel:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable the IPv6 packet forwarding function		**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 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 link-local address is configured.
		ipv6 address *ipv6-address* **link-local**	
Set the tunnel to an ISATAP tunnel		**tunnel-protocol ipv6-ipv4 isatap**	Required By default, the tunnel is a GRE tunnel. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or source interface for the tunnel		**source** { *ip-address* | *ipv6-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel.
Configure a link aggregation group ID to be referenced by the tunnel interface		**aggregation-group** *aggregation-group-id*	Required
Enable the expedite termination function		**expediting enable**	Optional By default, the expedite termination function is disabled.
Configure an address and mask for the expedite termination subnet		**expediting subnet** *ip-address mask*	Optional By default, no expedite termination subnet is configured for a tunnel.
Configure the tunnel interface MTU		**mtu** *mtu-size*	Optional

For the configuration of the tunnel interface MTU, refer to the ***ipv6 mtu*** *command in the Switch 8800 Command Reference Guide.*

CAUTION:

- *For automatic IPv4-compatible IPv6 tunnels, 6to4 tunnels, or ISATAP tunnels, their tunnel interfaces must have different source addresses.*
- *If the tunnel interface addresses 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. You can configure static or dynamic routes. A forwarding route needs to be configured at both ends of the tunnel. For the detailed configuration, refer to "IP Routing and Routing Table" on page 187.*
- *When you configure a static route, you need to configure a route to the destination address (the destination IP address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. A static route must be configured at both ends of the tunnel.*
- *Protocol packets can be processed properly only after expedite termination is enabled on the tunnel interface.*

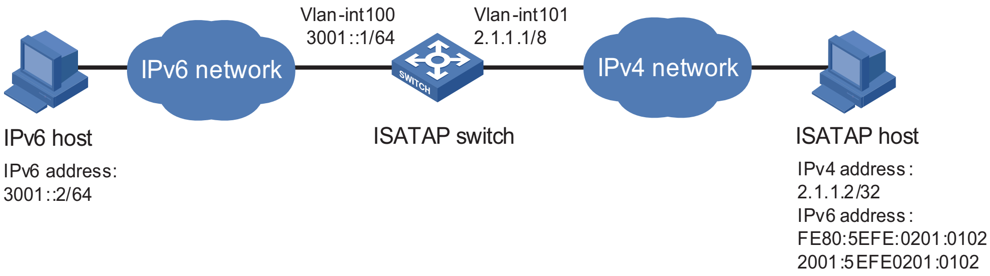
Configuration Example

Network requirements

The destination address of a tunnel is an ISATAP address. It is required that IPv6 hosts in the IPv4 network can access the IPv6 network via an ISATAP tunnel.

Network diagram

Figure 119 Network diagram for an ISATAP tunnel



Configuration procedure

The following example shows how to configure an ISATAP tunnel between the switch and the ISATAP host, which allows a separate ISATAP host to access the IPv6 network.

1 Configure the switch

# Enable the IPv6 forwarding function.

```
<Switch> system-view
[Switch] ipv6
```

# Configure addresses for interfaces.

```
[Switch] vlan 100
[Switch-vlan100] port GigabitEthernet 1/1/1
[Switch-vlan100] quit
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] ipv6 address 3001::1/64
[Switch-Vlan-interface100] quit
[Switch] vlan 101
```

```
[Switch-vlan101] port GigabitEthernet 1/1/2
[Switch-vlan101] quit
[Switch] interface vlan-interface 101
[Switch-Vlan-interface101] ip address 2.1.1.1 255.0.0.0
[Switch-Vlan-interface101] quit
```

Configure a link aggregation group and set the service type to **tunnel**.

```
[Switch] link-aggregation group 1 mode manual
[Switch] link-aggregation group 1 service-type tunnel
[Switch] interface GigabitEthernet 1/1/3
[Switch-GigabitEthernet1/1/3] stp disable
[Switch-GigabitEthernet1/1/3] port link-aggregation group 1
[Switch-GigabitEthernet1/1/3] quit
```

# Reference link aggregation group 1 and enable expedite termination in tunnel interface view.

```
[Switch] interface tunnel 2/0/1
[Switch-Tunnel2/0/1] aggregation-group 1
[Switch-Tunnel2/0/1] expediting enable
[Switch-Tunnel2/0/1] quit
```

# Configure an ISATAP tunnel.

```
[Switch] interface tunnel 2/0/1
[Switch-Tunnel2/0/1] ipv6 address 2001::5efe:0201:0101 64
[Switch-Tunnel2/0/1] source vlan-interface 101
[Switch-Tunnel2/0/1] tunnel-protocol ipv6-ipv4 isatap
[Switch-Tunnel2/0/1] expediting enable
[Switch-Tunnel2/0/1] expediting subnet 2.1.1.0  255.255.255.0
```

# 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-Tunnel2/0/1] undo ipv6 nd ra halt
```

2 Configure 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 an IPv4 address for the ISATAP router to complete the configuration on the host. The ISATAP interface information is as follows:

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
{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
```

# A link-local address (fe80::5efe:2.1.1.2) in the ISATAP format is automatically generated for the ISATAP interface. Configure an IPv4 address for 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
{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
```

# 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 switch 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.

# Configure a static route to the IPv6 host.

```
C:\>ipv6 rtu 3000::/64 2/2001::5efe:2.1.1.1
```

Configuration verification

After the above configurations, the ISATAP host can access hosts in the IPV6 network.

Configuring IPv4 over IPv4 Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and Loopback interface on the device so that they can communicate. These interfaces serve as the source interface of the virtual tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure an IPv4 over IPv4 tunnel:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a tunnel interface and enter tunnel interface view	**interface tunnel** *number*	Required By default, there is no tunnel interface on the device.

To do...	Use the command...	Remarks
Configure an IPv4 address for the tunnel interface	**ip address** *ip-address* { *mask* | *mask-length* } [**sub**]	Required By default, no IPv4 address is configured for the tunnel interface.
Set the tunnel to an IPv4 over IPv4 tunnel	**tunnel-protocol ipv4-ipv4**	Optional By default, the tunnel is a GRE tunnel. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or source interface for the tunnel	**source** { *ip-address* | *ipv6-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.
Configure a link aggregation group ID to be referenced by the tunnel interface	**aggregation-group** *aggregation-group-id*	Required
Configure the tunnel interface MTU	**mtu** *mtu-size*	Optional

CAUTION:

- *If the tunnel interface addresses 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. You can configure static or dynamic routes. A forwarding route needs to be configured at both ends of the tunnel. For the detailed configuration, refer to "IP Routing and Routing Table" on page 187.*
- *Two or more tunnel interfaces using the same encapsulation protocol must have different source and destination addresses.*
- *If the tunnel interface is the source interface, the tunnel source address is the primary IP address of the source interface.*
- *IPv4 over IPv4 tunnels do not support expedite termination.*

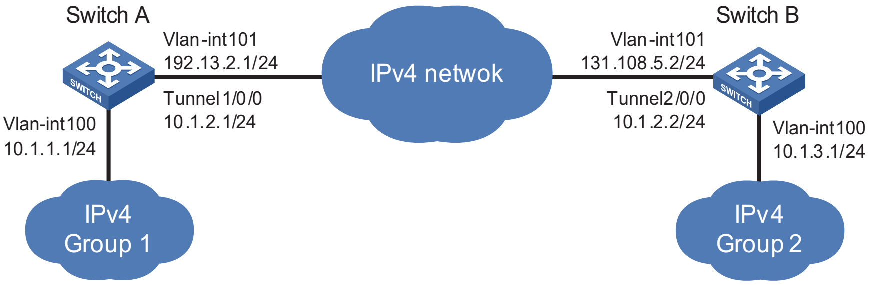
Configuration Example

Network requirements

The two subnets Group 1 and Group 2 running IPv4 are interconnected via an IPv4 over IPv4 tunnel between Switch A and Switch B.

Network diagram

Figure 120 Network diagram for an IPv4 over IPv4 tunnel



Configuration procedure

1 Configure Switch A

# Configure an IPv4 address for the interface VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/1/1
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 10.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

# Configure an IPv4 address for the interface VLAN-interface 101 (the physical interface of the tunnel).

```
[SwitchA] vlan 101
[SwitchA-vlan101] port GigabitEthernet 1/1/2
[SwitchA-vlan101] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ip address 192.13.2.1 255.255.255.0
[SwitchA-Vlan-interface101] quit
```

# Create the interface Tunnel 1/0/0.

```
[SwitchA] interface tunnel 1/0/0
```

# Configure an IPv4 address for the interface Tunnel 1/0/0.

```
[SwitchA-Tunnel1/0/0] ip address 10.1.2.1 255.255.255.0
```

# Configure the tunnel encapsulation mode.

```
[SwitchA-Tunnel1/0/0] tunnel-protocol ipv4-ipv4
```

# Configure a source address for the interface Tunnel 1/0/0.

```
[SwitchA-Tunnel1/0/0] source 192.13.2.1
```

# Configure a source address for the interface Tunnel 1/0/0 (IP address of the interface VLAN-interface 101 of Switch B).

```
[SwitchA-Tunnel1/0/0] destination 131.108.5.2
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 1/1/3
[SwitchA-GigabitEthernet1/1/3] stp disable
[SwitchA-GigabitEthernet1/1/3] port link-aggregation group 1
[SwitchA-GigabitEthernet1/1/3] quit
```

# Reference link aggregation group 1 in tunnel interface view.

```
[SwitchA] interface tunnel 1/0/0
[SwitchA-Tunnel1/0/0] aggregation-group 1
[SwitchA-Tunnel1/0/0] quit
```

# Configure a static route from Switch A through the interface Tunnel 1/0/0 to Group 2.

```
[SwitchA] ip route-static 10.1.3.0 255.255.255.0 tunnel 1
```

2 Configure Switch B.

# Configure an IPv4 address for the interface VLAN-interface 100.

```
<SwitchB> system-view
[SwitchB] vlan 100
[SwitchB-vlan100] port ethernet 1/1/1
[SwitchB-vlan100] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 10.1.3.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

# Configure an IPv4 address for the interface VLAN-interface 101 (the physical interface of the tunnel).

```
[SwitchB] vlan 101
[SwitchB-vlan101] port ethernet 1/1/2
[SwitchB-vlan101] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ip address 131.108.5.2 255.255.255.0
[SwitchB-Vlan-interface101] quit
```

# Create the interface Tunnel 2/0/0.

```
[SwitchB] interface tunnel 2/0/0
```

# Configure an IPv4 address for the interface Tunnel 2/0/0.

```
[SwitchB-Tunnel2/0/0] ip address 10.1.2.2 255.255.255.0
```

# Configure the tunnel encapsulation mode.

```
[SwitchB-Tunnel2/0/0] tunnel-protocol ipv4-ipv4
```

# Configure the source address for the interface Tunnel 2/0/0.

```
[SwitchB-Tunnel2/0/0] source 131.108.5.2
```

# Configure the destination address for the interface Tunnel 2/0/0 (IP address of the interface VLAN-interface 100 of Switch A).

```
[SwitchB-Tunnel2/0/0] destination 192.13.2.1
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 1/1/3
[SwitchB-GigabitEthernet1/1/3] stp disable
[SwitchB-GigabitEthernet1/1/3] port link-aggregation group 1
[SwitchB-GigabitEthernet1/1/3] quit
```

# Reference link aggregation group 1 in tunnel interface view.

```
[SwitchB] interface tunnel 2/0/0
[SwitchB-Tunnel2/0/0] aggregation-group 1
[SwitchB-Tunnel2/0/0] quit
```

Configure a static route from Switch B through the interface Tunnel 2/0/0 to Group 1.

```
[SwitchB] ip route-static 10.1.1.0 255.255.255.0 tunnel 2/0/0
```

Configuration verification

After the above configuration, you can successfully ping the address of the access interface of the peer IPv4 group from one switch.

Configuring Tunnel Hybrid Insertion

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and Loopback interface on the device so that they can communicate. These interfaces serve as the source interface of the virtual tunnel interface to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure tunnel hybrid insertion:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enable the IPv6 packet forwarding function		**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 site-local address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }	Use any command
		ipv6 address *ipv6-address*/*prefix-length* **eui-64**	
	Configure a link-local address	**ipv6 address auto link-local**	
		ipv6 address *ipv6-address* **link-local**	
Configure the source address or source interface of the tunnel interface		**source** { *ip-address* | *ipv6-address* | *interface-type interface-number* }	Required By default, no source address or interface is configured for the tunnel interface.
Create an ACL and enter ACL view		**acl number** *acl-number* [**match-order** { **config** | **auto** }]	Required
Define a ACL rule		**rule** [*rule-id*] { **permit** | **deny** } *protocol* [*rule-string*]	Required
Define a class and enter class view		**traffic classifier** *tcl-name* [**operator** { **and** | **or** }]	Required
Define the packet matching rule		**if-match** [**not**] *match-criteria*	Required
Define a traffic behavior and enter traffic behavior view		**traffic behavior** *behavior-name*	Required

To do...	Use the command...	Remarks
Configure the traffic redirection action for the traffic behavior	**redirect** { **cpu | interface** *interface-type interface-number* | **link-aggregation group** *aggregation-group-id* | **next-hop** { *ipv4-add* [*ipv4-add*] **|** *ipv6-add* [*interface-type interface-number*] [*ipv6-add* [*interface-type interface-number*]] } }	Required
Configure a service loopback group ID to be referenced by the tunnel interface	**aggregation-group** *aggregation-group-id*	Required
Enable the expedite termination function	**expediting enable**	Optional By default, the expedite termination function is disabled.

CAUTION:

- *If the tunnel interface addresses 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. You can configure static or dynamic routes. A forwarding route needs to be configured at both ends of the tunnel. For the detailed configuration, refer to "IP Routing and Routing Table" on page 187.*
- *Two or more tunnel interfaces using the same encapsulation protocol must have different source and destination addresses.*
- *If the tunnel interface is the source interface, the tunnel source address is the primary IP address of the source interface.*

Configuration Example

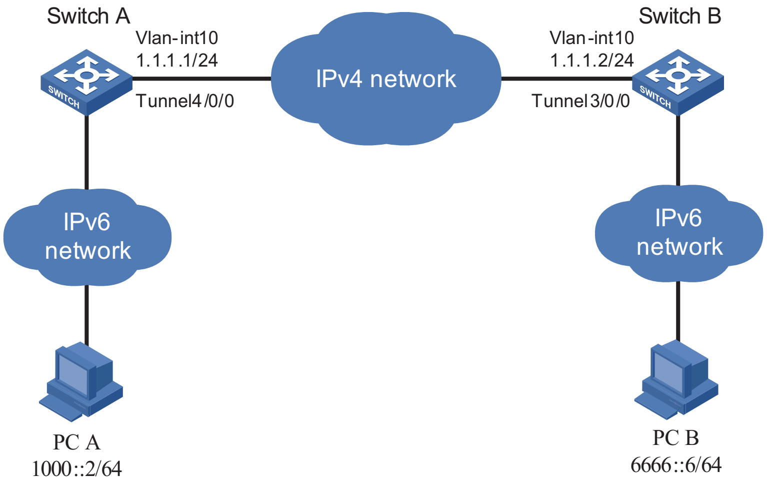
Network requirement

- Switch A and Switch B are configured with IPv6 cards and IPv4 cards. In this example, the tunnel type used for networking is an IPv6 manually configured tunnel, on which the RIPng routing protocol is enabled.
- IPv6 packets (destination IPv6 address is 6666::6) enter the tunnel from the IPv6-supporting interface on Switch A. After encapsulation, the packets are turned into IPv6 over IPv4 tunnel packets.
- IPv6 over IPv4 tunnel packets are sent to the IPv4 network through the port that supports IPv4 only.
- After passing through the IPv4 network, tunnel packets enter the destination dual-stack node, Switch B, through the port that supports IPv4 only.
- On Switch B, the ACL is used to redirect tunnel packets from IPv4 cards to the link aggregation group, whose service type is **tunnel**. This link aggregation group is established over IPv6 cards.
- After expedite termination on IPv6 cards, tunnel packets are forwarded from IPv4 cards to the IPv6 network.

- On PC A, the next hop gateway address of the route to PC B (6666::6/64) is set to 1000::1, and on PC B, the next hop gateway address of the route to PC A (1000::2/64) s is set to 6666::9.

Network diagram

Figure 121 Network diagram for tunnel hybrid insertion



Configuration procedure

1 Configure Switch A.

<SwitchA> system-view# Enable the IPv6 forwarding function.

```
[SwitchA] ipv6
```

# Configure an IPv4 address for the interface VLAN-interface 12.

```
[SwitchA] vlan 12
[SwitchA-vlan12] port GigabitEthernet4/1/3
[SwitchA] interface Vlan-interface 12
[SwitchA-Vlan-interface12] ipv6 address 1000::1 64
[SwitchA-Vlan-interface12] quit
```

# Configure a link aggregation group and set the service type to **tunnel**.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA-GigabitEthernet4/1/1] stp disable
[SwitchA-GigabitEthernet4/1/1] port link-aggregation group 1
[SwitchA-GigabitEthernet4/1/1] quit
```

# Configure the tunnel source interface - VLAN-interface 10 on the IPv4 card.

```
[SwitchA] vlan 10
[SwitchA-vlan10] port GigabitEthernet 3/1/1
[SwitchA] interface Vlan-interface 10
[SwitchA-Vlan-interface10] ip address 1.1.1.1 24
[SwitchA-Vlan-interface10] quit
```

# Configure an IPv6 manually configured tunnel on the interface Tunnel 4/0/0.

```
[SwitchA] interface Tunnel 4/0/0
[SwitchA-Tunnel4/0/0] tunnel-protocol ipv6-ipv4
[SwitchA-Tunnel4/0/0] ipv6 address 3333::1 64
[SwitchA-Tunnel4/0/0] source Vlan-interface 10
```

```
[SwitchA-Tunnel4/0/0] destination 1.1.1.2
[SwitchA-Tunnel4/0/0] aggregation-group 1
```

# Enable expedite termination on the interface Tunnel 4/0/0.

```
[SwitchA-Tunnel4/0/0] expediting enable
[SwitchA-Tunnel4/0/0] quit
```

# Enable RIPng on the interface Tunnel 4/0/0.

```
[SwitchA] ripng
[SwitchA-ripng-1] import-route direct
[SwitchA-ripng-1] quit
[SwitchA] interface Tunnel 4/0/0
[SwitchA-Tunnel4/0/0] ripng 1 enable
```

# Configure an ACL and redirect the tunnel packets that come from the IPv4 cards and should be terminated to IPv6 cards. The protocol number of IPv6 over IPv4 tunnel packets is 41.

```
[SwitchA] acl number 3000
[SwitchA-acl-adv-3000] rule permit 41
[SwitchA-acl-adv-3000] quit
[SwitchA] traffic classifier 1
[SwitchA-classifier-1] if-match acl 3000
[SwitchA-classifier-1] quit
[SwitchA] traffic behavior 1
[SwitchA-behavior-1] redirect link-aggregation group 1
[SwitchA] qos policy 1
[SwitchA-qospolicy-1] classifier 1 behavior 1
[SwitchA-qospolicy-1] quit
[SwitchA] qos vlan-policy 1 vlan 10 inbound
```

2 Configure Switch B.

# Enable IPv6 globally.

```
[SwitchB] ipv6
```

# Configure the tunnel destination address on the interface Tunnel 3/0/0.

```
[SwitchB] vlan 10
[SwitchB-vlan10] port GigabitEthernet 2/1/1
[SwitchB-vlan10] quit
[SwitchB] interface Vlan-interface 10
[SwitchB-Vlan-interface10] ip address 1.1.1.2 24
[SwitchB-Vlan-interface10] quit
```

# Configure a link aggregation group and set the service type to **tunnel** on the IPv6 card.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 3/1/2
[SwitchB-GigabitEthernet3/1/2] stp disable
[SwitchB-GigabitEthernet3/1/2] port link-aggregation group 1
[SwitchB-GigabitEthernet3/1/2] quit
```

# Create the tunnel interfaces.

```
[SwitchB] interface Tunnel 3/0/0
[SwitchB-Tunnel3/0/0] tunnel-protocol ipv6-ipv4
[SwitchB-Tunnel3/0/0] ipv6 address 3333::2 64
[SwitchB-Tunnel3/0/0] source Vlan-interface 10
```

```
[SwitchB-Tunnel3/0/0] destination 1.1.1.1
[SwitchB-Tunnel3/0/0] aggregation-group 1
```

# Enable expedite termination on the interface Tunnel 3/0/0.

```
[SwitchB-Tunnel3/0/0] expediting enable
```

# Enable RIPng on the interface Tunnel 3/0/0.

```
[SwitchB-Tunnel3/0/0] quit
[SwitchB] ripng
[SwitchB-ripng-1] import-route direct
[SwitchB-ripng-1] quit
[SwitchB] interface Tunnel 3/0/0
[SwitchB-Tunnel3/0/0] ripng 1 enable
```

# Configure an ACL and redirect the tunnel packets that come from the IPv4 cards and should be terminated to IPv6 cards. The protocol number of IPv6 over IPv4 tunnel packets is 41.

```
[SwitchB] acl number 3000
[SwitchB-acl-adv-3000] rule permit 41
[SwitchB-acl-adv-3000] quit
[SwitchB] traffic classifier 1
[SwitchB-classifier-1] if-match acl 3000
[SwitchB-classifier-1] quit
[SwitchB] traffic behavior 1
[SwitchB-behavior-1] redirect link-aggregation group 1
[SwitchB-behavior-1] quit
[SwitchB] qos policy 1
[SwitchB-qospolicy-1] classifier 1 behavior 1
[SwitchB-qospolicy-1] quit
[SwitchB] qos vlan-policy 1 vlan 10 inbound
```

# Configure the outbound interface for terminated IPv6 packets.

```
[SwitchB] vlan 12
[SwitchB-vlan12] port GigabitEthernet 3/1/1
[SwitchB-vlan12] quit
[SwitchB] interface Vlan-interface 12
[SwitchB-Vlan-interface12] ipv6 address 6666::9 64
```

Configuration verification

After the above configurations, you can successfully ping the IPv6 address 6666::6 of PC B from PC A or the IPv6 address 1000::2 of PC A from PC B.

Displaying and Maintaining Tunneling Configuration

To do...	Use the command...	Remarks
Display information related to 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*	

For details about BGP, refer to "BGP Overview" on page 419.

Troubleshooting Tunneling Configuration

Symptom: After the configuration of related parameters such as tunnel source address, tunnel destination address, and tunnel type, 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 or down. If the physical interface is down, use the **debugging tunnel event** command in user view to view the cause. For related commands and description of debugging information, refer to the *Switch 8800 Command Reference Guide*.

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.

34 GRE CONFIGURATION

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

- *Routers mentioned and router icons illustrated in the contents below represent the general routers and Ethernet switches running routing protocols. To simplify the description, this explanation will not be provided otherwise.*
- *Currently, the products do not support configuring IS-IS, IPv6-IS-IS or multicasting on tunnels.*

GRE Overview

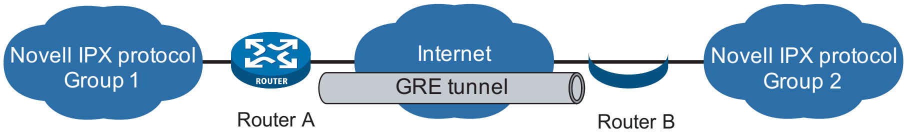
Introduction to GRE

Generic routing encapsulation (GRE) is a protocol designed for performing encapsulation of one network layer protocol (for example, IP or IPX) over another network layer protocol (for example, IP). GRE uses the tunneling technology and serves as a Layer 3 tunneling protocol of virtual private network (VPN).

A tunnel is a virtual point-to-point connection for transferring encapsulated packets. Packets are encapsulated at one end of the tunnel and decapsulated at the other end.

A packet transferred through a tunnel undergoes an encapsulation process and a decapsulation process. Figure 122 depicts the network used to illustrate these two processes.

Figure 122 IPX networks interconnected through the GRE tunnel



Encapsulation process

1. After receiving an IPX packet through the interface connected to IPX network Group1, Router A submits it to the IPX module for processing.
2. The IPX module checks the destination address field in the IPX header to determine how to route the packet.

3 If the packet must be tunneled to reach its destination, Router A sends it to the tunnel interface.

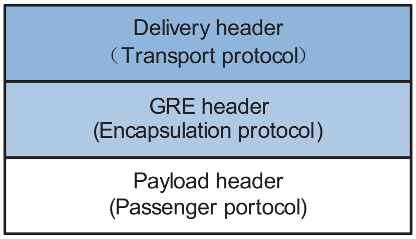
4 Upon receipt of the packet, the tunnel interface encapsulates it in a GRE packet and submits to the IP module.

5 The IP module encapsulates the packet in an IP packet, and then forwards the IP packet out through the corresponding network interface based on its destination address and the routing table.

Format of an encapsulated packet

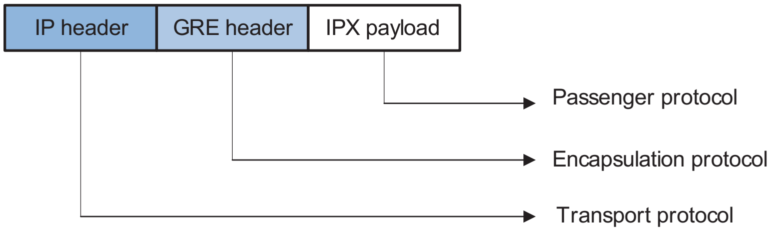
Figure 123 shows the format of an encapsulated packet.

Figure 123 Format of an encapsulated packet



As an example, Figure 124 shows the format of an IPX packet encapsulated for transmission over an IP tunnel.

Figure 124 Format of an IPX packet encapsulated for transmission over an IP tunnel



These are the involved terms:

- Payload: Packet that needs to be encapsulated and routed.
- Passenger protocol: Protocol that the payload packet uses, IPX in the example.
- Encapsulation or carrier protocol: Protocol used to encapsulate the payload packet, that is, GRE.
- Delivery or transport protocol: Protocol used to encapsulate the GRE packet and to forward the resulting packet to the other end of the tunnel, IP in this example.

Decapsulation process

Decapsulation is the reverse process of encapsulation:

1 Upon receiving an IP packet from the tunnel interface, Router B checks the destination address.

2 If the destination is itself, Router B strips off the IP header of the packet and submits the resulting packet to the GRE module.

3 The GRE module checks the key, checksum and sequence number, and then strips off the GRE header and submits the payload to the IPX module.

4 The IPX module performs the subsequent forwarding processing for the packet.

Encapsulation and decapsulation processes on both ends of the GRE tunnel and the resulting increase in data volumes will degrade the forwarding efficiency for the GRE-enabled device to some extent.

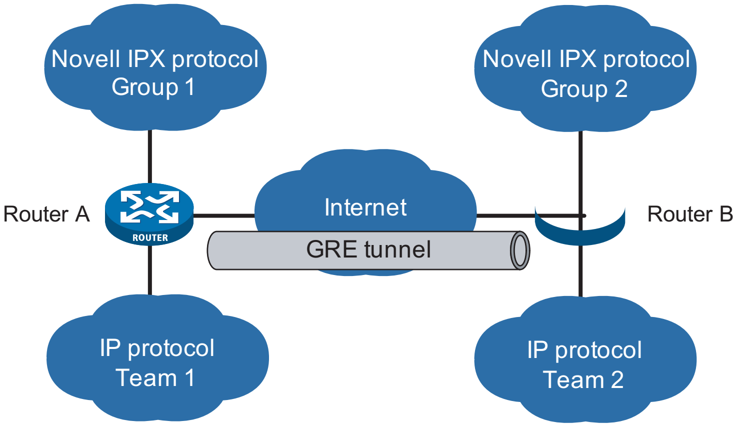
GRE Applications

GRE supports these types of applications:

- "Multi-protocol communications through a single-protocol backbone" on page 407
- "Scope enlargement of the network running a hop-limited protocol" on page 407
- "VPN creation by connecting discontinuous subnets" on page 408

Multi-protocol communications through a single-protocol backbone

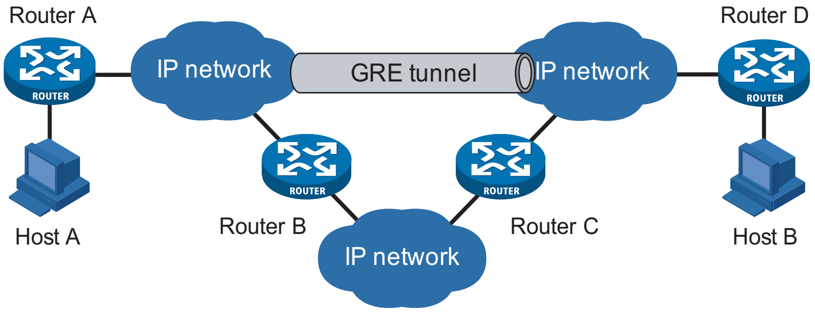
Figure 125 Multi-protocol communications through a single-protocol backbone



In the example as shown in Figure 125, Group1 and Group2 are local networks running Novell IPX, while Team1 and Team2 are local networks running IP. Through the GRE tunnel between Router A and Router B, Group1 can communicate with Group2 and Team1 can communicate with Team2. They will not interfere with each other.

Scope enlargement of the network running a hop-limited protocol

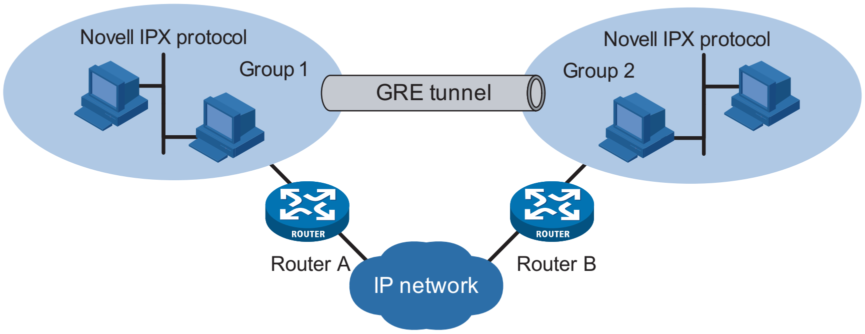
Figure 126 Scope enlargement of the network



When the hop count between two terminals exceeds 15, the terminals cannot communicate with each other. Using GRE, you can hide some hops so as to enlarge the scope of the network.

VPN creation by connecting discontinuous subnets

Figure 127 Connect discontinuous subnets with a tunnel to form a VPN



In the example as shown in Figure 127, Group1 and Group2 running Novell IPX are deployed in different cities. They can constitute a trans-WAN virtual private network (VPN) through the tunnel.

Configuring a GRE over IPv4 Tunnel

Configuration Prerequisites

Interfaces on a device, such as VLAN interfaces and loopback interfaces, are configured with IPv4 addresses and can communicate. These interfaces can be used as the source of a virtual tunnel interface to ensure the reachability of the tunnel destination address.

Configuration Procedure

Follow these steps to configure a GRE over IPv4 tunnel:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable IPv6 packet forwarding	**ipv6**	Optional By default, the IPv6 packet forwarding function is disabled. On IPv6 over IPv4 GRE tunnels, this function is mandatory.
Create a tunnel interface and enter tunnel interface view	**interface tunnel** *interface-number*	Required Not created by default

To do...		Use the command...	Remarks
Configure an IPv4 address for the tunnel interface		**ip address** *ip-address* { *mask* | *mask-length* }	Any of the three must be selected. By default, no IPv4 address is configured on a tunnel interface. Whether to configure an IPv4 or IPv6 address on a tunnel interface depends on the actual needs. By default, no IPv6 global unique address or site-local address is configured on a tunnel interface.
Configure an IPv6 address for the tunnel interface	Configure an IPv6 global unique address or a site-local address	**ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* } **ipv6 address** *ipv6-address*/*prefix-length* **eui-64**	
	Configure an IPv6 link-local address	**ipv6 address auto link-local**	Optional
		ipv6 address *ipv6-address* **link-local**	By default, when an interface is configured with an IPv6 global unique address or a site-local address, a link-local address is created automatically.
Set the tunnel mode to GRE over IPv4		**tunnel-protocol gre**	Optional GRE over IPv4 by default Note that both ends of a tunnel must be configured with the same tunnel mode. Otherwise, packet delivery will fail.
Configure the source address for the tunnel interface		**source** { *ip-address* | *interface-type interface-number* }	Required By default, no source address is configured for a tunnel interface.
Configure the destination address for the tunnel interface		**destination** *ip-address*	Required By default, no destination address is configured for a tunnel interface.
Specify the service loop group for the tunnel interface to reference		**aggregation-group** *aggregation-group-ID*	Required
Configure a route through the tunnel		Refer to *"IP Routing Overview" on page 187.*	Optional Each end of the tunnel must have a route (static or dynamic) through the tunnel to the other end.
Enable the expedite termination function for a tunnel interface		**expediting enable**	Optional Disabled by default Moreover, this function has no effect on GRE IPv4 over IPv4 tunnels
Set the MTU value for the tunnel interface		**mtu** *mtu-size*	Optional

Note that:

- For a tunnel interface that is configured with any of the above features, all the configuration disappears once that interface is deleted.
- The source address and destination address of a tunnel uniquely identify a path. They must be configured at both ends of the tunnel and are mutually the source address and the destination address.
- Two or more tunnel interfaces using the same encapsulation protocol must have different source addresses and destination addresses.
- If you configure a source interface for a tunnel interface, the source address of the tunnel interface is the primary IP address of the source interface.
- The source and destination addresses of a tunnel must be different from each other. Moreover, for static routes configured on a tunnel interface, their destination addresses cannot be in the same network segment as the address of that tunnel interface.
- When you configure a route through the tunnel, you can configure a static route, whose destination address is the destination address of the packet not encapsulated in GRE and next hop is the address of the tunnel interface at the remote end. Or, you can enable the dynamic routing protocol on both the tunnel interface and the router interface connecting the private network so that the dynamic routing protocol can establish a routing entry that allows the tunnel to forward packets through the tunnel.
- IPv6 over IPv4 GRE tunnel interfaces support such dynamic routing protocols as OSPFv3, RIPng and BGP4+.
- To run dynamic routing protocols (apart from BGP4+) on a tunnel interface, you need to enable the expediting function on the relevant tunnel first.
- It is not allowed to set up on a tunnel interface static routes whose destination addresses are in the network segment as that tunnel interface.
- For GRE IPv6 over IPv4 packets, due to restrictions of match conditions for the expediting function, only the physical ports with the link type as Access or Hybrid can be bound to the VLAN interface that acts as the source interface of a tunnel. Moreover, when the link type of a port is Hybrid, the *untagged* attribute must be specified for the VLAN that sends GRE tunnel packets.
- The 3C17526 and 3C17532 modules do not support the expediting function on GRE IPv6 over IPv4 tunnels.

Displaying and Maintaining GRE

To do...	Use the command...	Remarks
Display information about a specified or all tunnel interfaces	**display interface tunnel** [*number*]	Available in any view
Display IPv6 information about a tunnel interface	**display ipv6 interface tunnel** *number*	Available in any view

GRE Tunnel Configuration Example

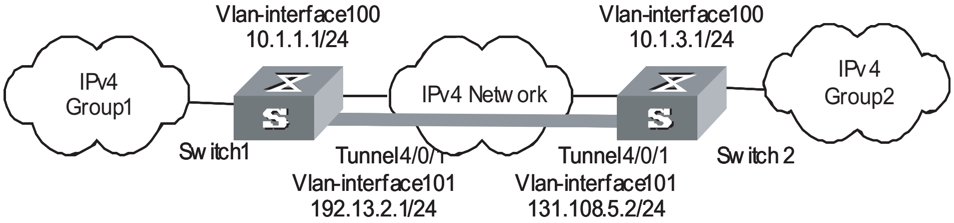
GRE IPv4 over IPv4 Tunnel Configuration Example

Network requirements

Switch 1 and Switch 2 are interconnected through an IPv4 network. Two private IPv4 subnets Group1 and Group2 are interconnected through a GRE tunnel between the two switches.

Network diagram

Figure 128 Network diagram for GRE application



Configuration procedure

1 Configure Switch 1

# Configure vlan-interface 100.

```
<Sysname1> system-view
[Sysname1] vlan 100
[Sysname1-vlan100] port GigabitEthernet 4/1/1
[Sysname1-vlan100] quit
[Sysname1] interface vlan-interface 100
[Sysname1-Vlan-interface100] ip address 10.1.1.1 255.255.255.0
[Sysname1-Vlan-interface100] quit
```

# Configure vlan-interface 101, the physical interface for the tunnel.

```
[Sysname1] vlan 101
[Sysname1-vlan101] port GigabitEthernet 4/1/2
[Sysname1-vlan101] quit
[Sysname1] interface vlan-interface 101
[Sysname1-Vlan-interface101] ip address 192.13.2.1 255.255.255.0
[Sysname1-Vlan-interface101] quit
```

# Create an interface named Tunnel 4/0/1.

```
[Sysname1] interface tunnel 4/0/1
```

# Configure an IPv4 address for interface Tunnel 4/0/1.

```
[Sysname1-Tunnel4/0/1] ip address 10.1.2.1 255.255.255.0
```

# Configure the tunnel encapsulation mode.

```
[Sysname1-Tunnel4/0/1] tunnel-protocol gre
```

# Configure the source address of interface Tunnel 4/0/1 to be the IP address of the VLAN interface of interface GigabitEthernet 4/1/2.

```
[Sysname1-Tunnel4/0/1] source vlan-interface 101
```

# Configure the destination address for interface Tunnel 4/0/1 (IP address of the VLAN interface to which GigabitEthernet 4/1/2 of Switch 2 belongs).

```
[Sysname1-Tunnel4/0/1] destination 131.108.5.2
[Sysname1-Tunnel4/0/1] expediting enable
[Sysname1-Tunnel4/0/1] quit
```

# Create service loop group 1, setting the configuration mode to manual and the service type to tunnel.

```
[Sysname1] link-aggregation group 1 mode manual
[Sysname1] link-aggregation group 1 service-type tunnel
```

# Add interface Ethernet4/1/3 to service loop group 1.

```
[Sysname1] interface GigabitEthernet 4/1/3
[Sysname1-GigabitEthernet4/1/3] stp disable
[Sysname1-GigabitEthernet4/1/3] port link-aggregation group 1
```

# Apply service loop group 1 to the tunnel in tunnel interface view.

```
[Sysname1-GigabitEthernet4/1/3] quit
[Sysname1] interface tunnel 4/0/1
[Sysname1-Tunnel4/0/1] aggregation-group 1
[Sysname1-Tunnel4/0/1] quit
```

# Configure a static route from Switch 1 through interface Tunnel 4/0/1 to Group2.

```
[Sysname1] ip route-static 10.1.3.0 255.255.255.0 tunnel 4/0/1
```

2 Configure Switch 2

# Configure vlan-interface 100.

```
<Sysname2> system-view
[Sysname2] vlan 100
[Sysname2-vlan100] port GigabitEthernet 4/1/1
[Sysname2-vlan100] quit
[Sysname2] interface vlan-interface 100
[Sysname2-Vlan-interface100] ip address 10.1.3.1 255.255.255.0
[Sysname2-Vlan-interface100] quit
```

# Configure vlan-interface 101, the physical interface for the tunnel.

```
[Sysname2] vlan 101
[Sysname2-vlan101] port GigabitEthernet 4/1/2
[Sysname2-vlan101] quit
[Sysname2] interface vlan-interface 101
[Sysname2-Vlan-interface101] ip address 131.108.5.2 255.255.255.0
[Sysname2-Vlan-interface101] quit
```

# Create an interface named Tunnel 4/0/1.

```
[Sysname2] interface tunnel 4/0/1
```

# Configure an IPv4 address for interface Tunnel 4/0/1.

```
[Sysname2-Tunnel4/0/1] ip address 10.1.2.2 255.255.255.0
```

# Configure the tunnel encapsulation mode.

```
[Sysname2-Tunnel4/0/1] tunnel-protocol gre
```

# Configure the source address for interface tunnel4/0/1 (IP address of the VLAN interface to which GigabitEthernet 4/1/2 belongs).

```
[Sysname2-Tunnel4/0/1] source vlan-interface 101
```

# Configure the destination address for interface Tunnel 4/0/1 (IP address of the VLAN interface to which GigabitEthernet 4/1/2 of Switch 1 belongs). Moreover, enable the expediting function.

```
[Sysname2-Tunnel4/0/1] destination 192.13.2.1
[Sysname2-Tunnel4/0/1] expediting enable
[Sysname2-Tunnel4/0/1] quit
```

# Create service loop group 1, setting the configuration mode to manual and the service type to tunnel.

```
[Sysname2] link-aggregation group 1 mode manual
[Sysname2] link-aggregation group 1 service-type tunnel
```

# Add interface GigabitEthernet 4/1/3 to service loop group 1.

```
[Sysname2] interface GigabitEthernet 4/1/3
[Sysname2-GigabitEthernet4/1/3] stp disable
[Sysname2-GigabitEthernet4/1/3] port link-aggregation group 1
```

# Apply service loop group 1 to the tunnel in tunnel interface view.

```
[Sysname2-GigabitEthernet4/1/3] quit
[Sysname2] interface tunnel 4/0/1
[Sysname2-Tunnel4/0/1] aggregation-group 1
[Sysname2-Tunnel4/0/1] quit
```

# Configure a static route from Switch 2 through interface Tunnel 4/0/1 to Group1.

```
[Sysname2] ip route-static 10.1.1.0 255.255.255.0 Tunnel 4/0/1
```

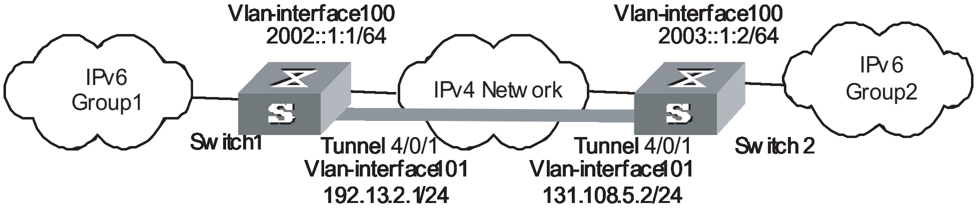
GRE IPv6 over IPv4 Tunnel Configuration Example

Network requirements

Switch 1 and Switch 2 are interconnected through an IPv4 network. Two IPv6 subnets Group1 and Group2 are interconnected through a GRE tunnel between Switch 1 and Switch 2.

Network diagram

Figure 129 Network diagram for GRE application



Configuration procedure

1 Configure Switch 1

# Enter system view.

```
<Sysname1> system-view
```

# Enable IPv6.

```
[Sysname1] ipv6
```

# Configure Vlan-interface100.

```
[Sysname1] vlan 100
[Sysname1-vlan100] port GigabitEthernet 4/1/1
[Sysname1-vlan100] quit
[Sysname1] interface vlan-interface 100
[Sysname1-Vlan-interface100] ipv6 address 2002::1:1 64
[Sysname1-Vlan-interface100] quit
```

# Configure Vlan-interface101, the physical interface for the tunnel.

```
[Sysname1] vlan 101
[Sysname1-vlan101] port GigabitEthernet 4/1/2
[Sysname1-vlan101] quit
[Sysname1] interface vlan-interface 101
[Sysname1-Vlan-interface101] ip address 192.13.2.1 255.255.255.0
[Sysname1-Vlan-interface101] quit
```

# Create an interface named Tunnel 4/0/1.

```
[Sysname1] interface tunnel 4/0/1
```

# Configure an IPv6 address for interface Tunnel 4/0/1.

```
[Sysname1-Tunnel4/0/1] ipv6 address 2001::1:1 64
```

# Configure the tunnel encapsulation mode.

```
[Sysname1-Tunnel4/0/1] tunnel-protocol gre
```

# Configure the source address of interface Tunnel 4/0/1 to be the IP address of the Vlan interface to GigabitEthernet 4/1/2 belongs.

```
[Sysname1-Tunnel4/0/1] source vlan-interface 101
```

# Configure the destination address of interface Tunnel 4/0/1 to be the IP address of the Vlan interface to which GigabitEthernet 4/1/2 of Switch 2 belongs. Additionally, enable the expediting function.

```
[Sysname1-Tunnel4/0/1] destination 131.108.5.2
[Sysname1-Tunnel4/0/1] expediting enable
[Sysname1-Tunnel4/0/1] quit
```

# Create service loop group 1, setting the configuration mode to manual and the service type to tunnel.

```
[Sysname1] link-aggregation group 1 mode manual
[Sysname1] link-aggregation group 1 service-type tunnel
```

# Add GigabitEthernet 4/1/3 to service loop group 1.

```
[Sysname1] interface GigabitEthernet 4/1/3
[Sysname1-GigabitEthernet4/1/3] stp disable
[Sysname1-GigabitEthernet4/1/3] port link-aggregation group 1
```

# Apply service loop group 1 to the tunnel in tunnel interface view.

```
[Sysname1-GigabitEthernet4/1/3] quit
[Sysname1] interface tunnel 4/0/1
[Sysname1-Tunnel4/0/1] aggregation-group 1
[Sysname1-Tunnel4/0/1] quit
```

# Configure a static route from Switch 1 through interface Tunnel 4/0/1 to Group2.

```
[Sysname1] ipv6 route-static 2003::0 64 tunnel 4/0/1
```

2 Configure Switch 2

# Enter system view.

```
<Sysname2> system-view
```

# Enable IPv6.

```
[Sysname2] ipv6
```

# Configure interface Vlan-interface100.

```
[Sysname2] vlan 100
[Sysname2-vlan100] port GigabitEthernet 4/1/1
[Sysname2-vlan100] quit
[Sysname2] interface vlan-interface 100
[Sysname2-Vlan-interface100] ipv6 address 2003::1:2 64
[Sysname2-Vlan-interface100] quit
```

# Configure interface Vlan-interface101, the physical interface for the tunnel.

```
[Sysname2] vlan 101
[Sysname2-vlan101] port GigabitEthernet 4/1/2
[Sysname2-vlan101] quit
[Sysname2] interface vlan-interface 101
```

```
[Sysname2-Vlan-interface101] ip address 131.108.5.2 255.255.255.0
[Sysname2-Vlan-interface101] quit
```

# Create an interface named Tunnel 4/0/1.

```
[Sysname2] interface tunnel 4/0/1
```

# Configure an IPv6 address for interface Tunnel 4/0/1.

```
[Sysname2-Tunnel4/0/1] ipv6 address 2001::1:2 64
```

# Configure the tunnel encapsulation mode.

```
[Sysname2-Tunnel4/0/1] tunnel-protocol gre
```

# Configure the source address of interface Tunnel 4/0/1 to be the IP address of the Vlan interface to which GigabitEthernet 4/1/2 belongs.

```
[Sysname2-Tunnel4/0/1] source vlan-interface 101
```

# Configure the destination address of interface Tunnel 4/0/1 to be the IP address of the Vlan interface to which GigabitEthernet 4/1/2 of Switch 1 belongs. Moreover, enable the expediting function.

```
[Sysname2-Tunnel4/0/1] destination 192.13.2.1
[Sysname2-Tunnel4/0/1] expediting enable
[Sysname2-Tunnel4/0/1] quit
```

# Create service loop group 1, setting the configuration mode to manual and the service type to tunnel.

```
[Sysname2] link-aggregation group 1 mode manual
[Sysname2] link-aggregation group 1 service-type tunnel
```

# Add GigabitEthernet 4/1/3 to service loop group 1.

```
[Sysname2] interface GigabitEthernet 4/1/3
[Sysname2-GigabitEthernet4/1/3] stp disable
[Sysname2-GigabitEthernet4/1/3] port link-aggregation group 1
```

# Apply service loop group 1 to the tunnel in tunnel interface view.

```
[Sysname2-GigabitEthernet4/1/3] quit
[Sysname2] interface tunnel 4/0/1
[Sysname2-Tunnel4/0/1] aggregation-group 1
[Sysname2-Tunnel4/0/1] quit
```

# Configure a static route from Switch 2 through interface Tunnel 4/0/1 to Group1.

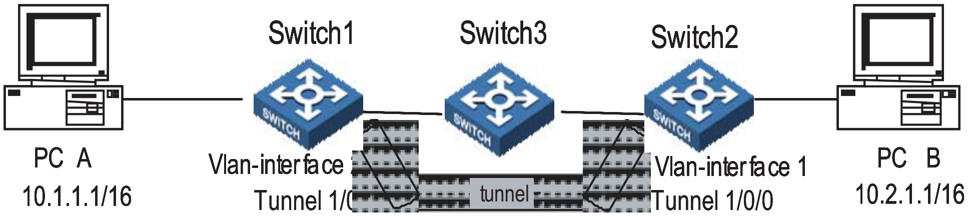
```
[Sysname2] ipv6 route-static 2002::0 64 Tunnel 4/0/1
```

Troubleshooting GRE

The GRE configurations are relatively simple. The key is to keep the configurations consistent. Most faults can be located by using the **debugging gre** or **debugging tunnel** command. This section analyzes only one type of fault, as shown in Figure 130. Switch 1 connects to Switch 2 via an IPv4 network. PC A and

PC B run IPv4 and they are connected to each other via a GRE tunnel between Switch 1 and Switch 2.

Figure 130 Troubleshoot GRE



Symptom: The interfaces at both ends of the tunnel are configured correctly and can ping each other, but PC A and PC B cannot ping each other.

Solution:

- On Switch 1 and Switch 2, carry out the **display ip routing-table** command in any view respectively. On Switch 1, observe whether there is a route from itself through Tunnel 1/0/0 to 10.2.0.0/16. On Switch 2, observe whether there is a route from itself through Tunnel 1/0/0 to 10.1.0.0/16.
- For any missing static routes, use the **ip route-static** command in system view to configure.

35 BGP CONFIGURATION

Border Gateway Protocol (BGP) is a dynamic inter-AS route discovery protocol.

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

The term "router" refers to a router in a generic sense or a Layer 3 switch, and BGP refers to BGP-4 in this document.

BGP Overview

Three early versions of BGP are BGP-1 (RFC1105), BGP-2 (RFC1163) and BGP-3 (RFC1267). The current version in use is BGP-4 (RFC1771). BGP-4 is rapidly becoming the defacto Internet exterior routing protocol standard and is commonly 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 discovery and calculation of routes, which makes BGP, an exterior routing protocol different from interior routing protocols such as OSPF and RIP
- Using TCP as its transport layer protocol to enhance reliability
- Supporting CIDR
- Substantially reducing bandwidth occupation by advertising updating routes only and applicable to advertising a great amount of routing information on the Internet
- Eliminating route loops completely by adding AS path information to BGP routes

- Providing abundant routing policies, allowing for implementing flexible route filtering and selection
- Easy to extend, satisfying new network developments

A router advertising BGP messages is called a BGP speaker, which exchanges new routing information with other BGP speakers. 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 speakers in the local AS.

BGP speakers call each other peers, and several associated peers form a peer group.

BGP runs on a router in one of the following two modes:

- IBGP (Interior 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

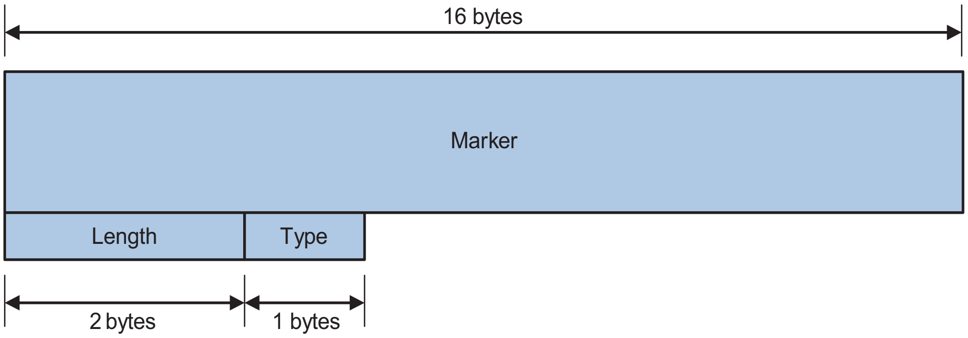
Header

BGP message involves five types:

- Open message
- Update message
- Notification message
- Keep-alive message
- Route-refresh message

They have the same header, as shown below:

Figure 131 BGP message header

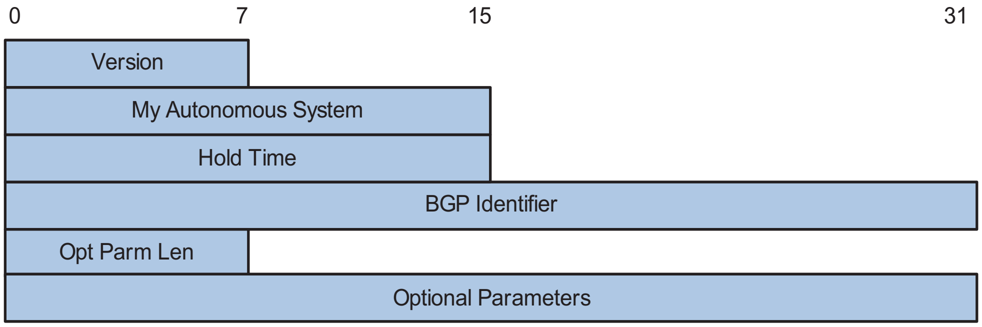


- Marker: The 16-octet field is used for BGP authentication calculation. If no authentication information is available, then the Marker must be all ones.
- Length: The 2-octet unsigned integer indicates the total length of the message.
- Type: This 1-octet unsigned integer indicates the type code of 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, the last one 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. The Open message contains the following fields:

Figure 132 BGP open message format

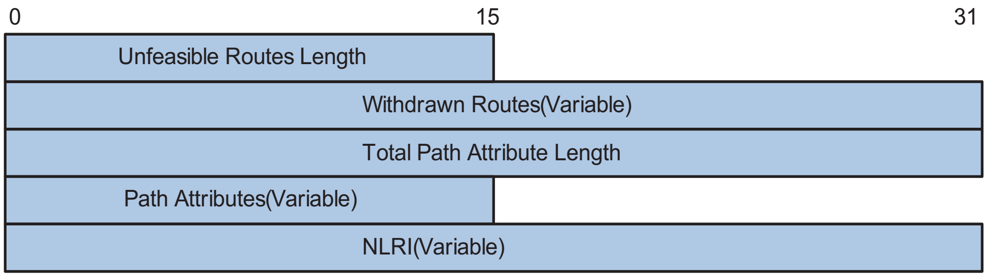


- Version: This 1-octet unsigned integer indicates the protocol version number of the message. The current BGP version is 4.
- My Autonomous System: This 2-octet unsigned integer indicates the Autonomous System number of the sender.
- Hold Time: When establishing peer relationship, two parties negotiate an identical Hold time. If no Keepalive or Update is received from a peer after the Hold time, the BGP connection is considered down.
- BGP Identifier: In IP address format, identifying the BGP router
- Opt Parm Len (Optional Parameters Length): Length of optional parameters, set to 0 if no optional parameter is available

Update

Update message is used to exchange routing information between peers. It can advertise a feasible route or remove multiple unfeasible routes. Its format is shown below:

Figure 133 BGP Update message format


0
15
31
Unfeasible Routes Length
Withdrawn Routes(Variable)
Total Path Attribute Length
Path Attributes(Variable)
NLRI(Variable)

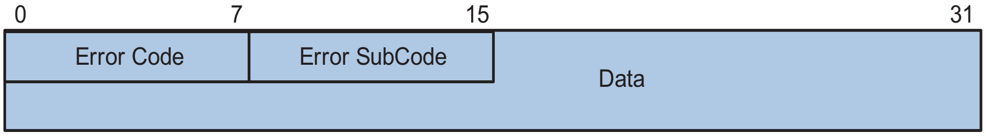
Each Update message can advertise a group of feasible routes with similar attributes, which are contained in the network layer reachable information (NLRI) field. The Path Attributes field carries attributes of these routes that are used by BGP for routing. Each message can also carry multiple withdrawn routes in the Withdrawn Routes field.

- Unfeasible Routes Length: The total length of the Withdrawn Routes field in octets. A value of 0 indicates neither any route is being withdrawn from service, nor Withdrawn Routes field is present in this Update message.
- Withdrawn Routes: This is a variable length field that contains a list of IP prefixes of routes that are being withdrawn from service.
- Total Path Attribute Length: Total length of the Path Attributes field in octets. 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, perform routing and protocol extension.
- NLRI (Network Layer Reachability Information): Reachability information is encoded as one or more 2-tuples of the form <length, prefix>.

Notification

A Notification message is sent when an error is detected. The BGP connection is closed immediately after sending it. Notification message format is shown below:

Figure 134 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.

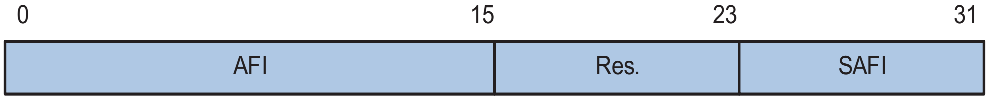
Keepalive

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 135 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 error occurs 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 attributes is described in the following table.

Table 23 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

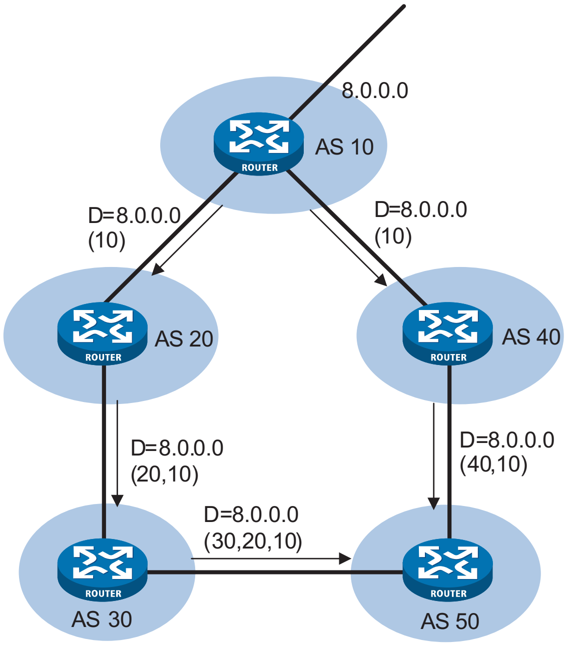
ORIGIN is a well-known mandatory attribute and defines the origin of routing information and how a route becomes a BGP route. It involves three types:

- 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 unavailable. 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 massages back. The number of the AS closest with the receiver's AS is leftmost, as shown below:

Figure 136 AS_PATH attribute



In general, a BGP router does not receive routes containing the local AS number to avoid routing loops.

The current implementation supports using the ***peer allow-as-loop*** *command to receive routes containing the local AS number to meet special requirements.*

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 information 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.

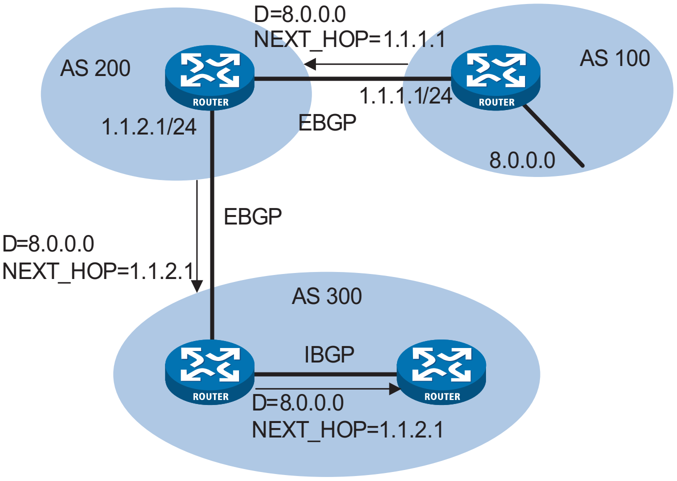
3 NEXT_HOP

Different from IGP, the NEXT_HOP attribute of BGP may not be the IP address of a neighboring router. It involves three types of values, as shown in Figure 137.

- 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” on page 426.

Figure 137 NEXT_HOP attribute

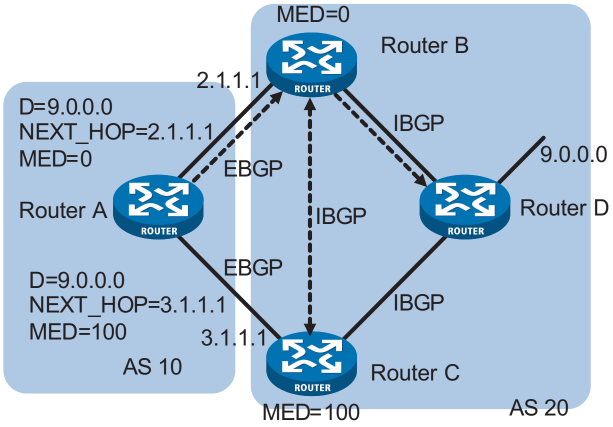


4 MED (MULTI_EXIT_DISC)

The MED attribute is exchanged between two neighboring ASs, each of which will 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 138 MED attribute



In general, BGP compares MEDs of routes to the same AS only.

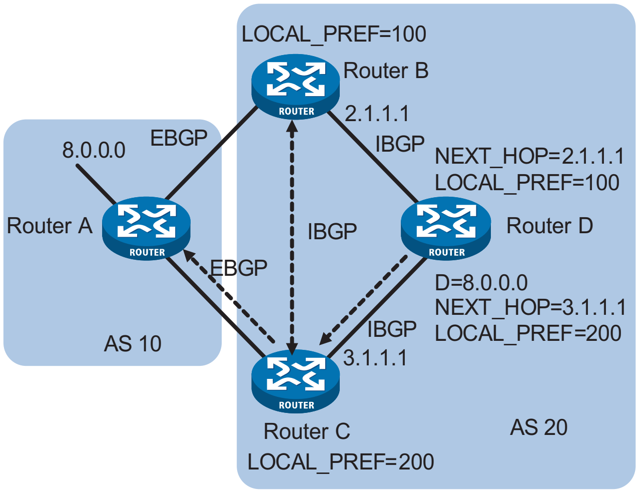
The current implementation supports using the ***compare-different-as-med*** *command to force BGP to compare MED values of routes to different ASs.*

5 LOCAL_PREF

This attribute is exchanged between IBGP peers only, thus 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 139 LOCAL_PREF attribute



6 COMMUNITY

The COMMUNITY attribute is used to simplify routing policy usage and ease management and maintenance. It is a collection of destination addresses having identical attributes, without physical boundaries in between, having nothing to do with local AS. Well known community attributes include:

- 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 Caused by Large Scale BGP Networks" on page 429).
- 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 rule

The current BGP implementation supports the following route selection rule:

- Discard routes with unreachable NEXT_HOP 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 ORIGIN 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

- *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 loop.*
- *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 a directly connected neighbor. One of the reasons is next hops in routing information exchanged between IBGPs are not modified. In this case, the router finds the direct route via IGP route entries to reach the next hop. The direct route is called reliable route. The process of finding a reliable route to reach a next hop is route recursion.

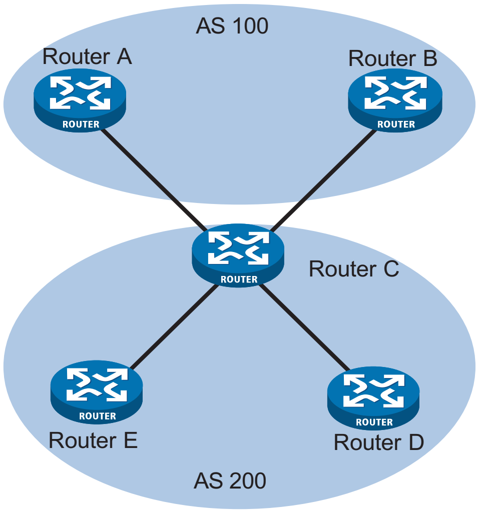
Currently, the system supports BGP load balancing based on route recursion, namely if reliable routes are load balanced (suppose three next hop addresses), BGP generates the same number of next hops to forward packets. Note that BGP load balancing based on route recursion is always enabled by the system rather than configured using command.

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 on 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.

- *BGP implements load balancing only on routes that have the same AS_PATH attribute, ORIGIN attribute, LOCAL_PREF and MED.*
- *BGP load balancing is applicable between EBGPs, IBGPs and between confederations.*
- *If multiple routes to the same destination are available, BGP selects routes for load balancing according to the configured maximum number of load balanced routes.*

Figure 140 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 adds both the two routes to its route table for load balancing. After that, Router C forwards routes to Router D and Router E only once, with AS_PATH unchanged, NEXT_HOP changed to Router C's address. Other BGP transitive attributes apply according to route selection rules.

BGP route advertisement rule

The current BGP implementation supports the following route advertisement rules:

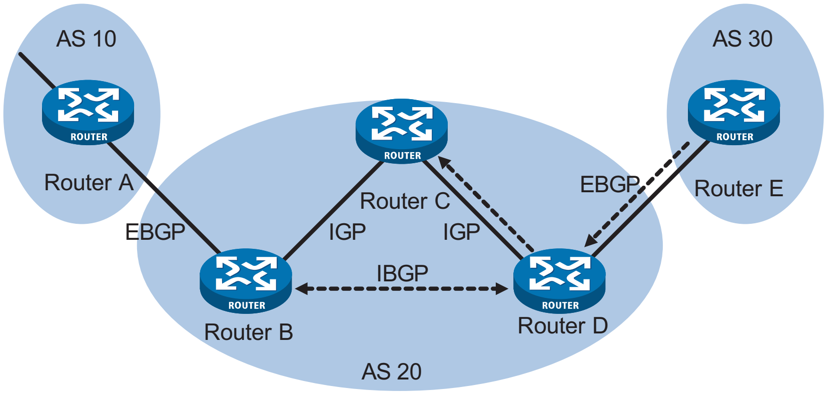
- When multiple available routes exist, a BGP speaker advertises only the best route to its peers.
- A BGP speaker advertises only routes used by itself.
- A BGP speaker advertises routes learned from EBGPs to all BGP peers, including both EBGP and IBGP peers.
- A BGP speaker does not advertise routes learned from IBGPs to IBGP peers.
- A BGP speaker advertises routes learned from IBGPs to EBGP peers. Note that if information synchronization is disabled between BGP and IGP, IBGP routes are advertised to EBGP peers. If enabled, only IGP advertises the IBGP routes can these routes be advertised to EBGP peers.
- A BGP speaker advertises all routes to a newly connected peer.

IBGP and IGP Information Synchronization

The routing Information synchronization between IBGP and IGP is for avoidance of giving wrong directions to routers outside of the local AS.

If a non-BGP router works in an AS, a packet forwarded via the router may be discarded due to unreachable destination. As shown in Figure 141, Router E learned a route of 8.0.0.0/8 from Router D via BGP. Then Router E sends a packet to Router A through Router D, which finds from its routing table that Router B is the next hop (configured using the **peer next-hop-local** command). Since Router D learned the route to Router B via IGP, it forwards the packet to Router C using route recursion. Router C has no idea about the route 8.0.0.0/8, so it discards the packet.

Figure 141 IBGP and IGP synchronization



If synchronization is configured in this example, the IBGP router (Router D) checks the learned IBGP route from its IGP routing table first. Only the route is available in the IGP routing table can the IBGP router 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).
- IBGP routers in the local AS are fully meshed.

Settlements for Problems Caused by Large Scale BGP Networks

Route summarization

The size of BGP routing tables on a large network is very large. Using route summarization can reduce the routing table size.

By summarizing multiple routes with one route, a BGP router advertises only the summary route rather than all routes.

Currently, the system supports both manual and automatic summarization. The latter provides for controlling the attribute of a summary route and deciding 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.

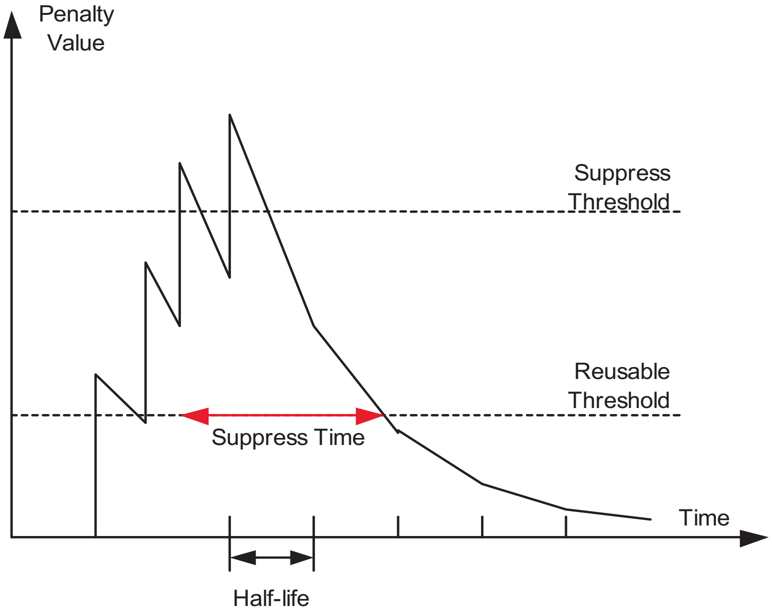
When a route flap occurs, the routing protocol sends an update to its neighbor, and then the neighbor needs to recalculate routes and modify the routing table. Therefore, frequent route flaps consume large bandwidth and CPU resources even affect normal operation of the network.

In most cases, BGP is used in complex networks, where route changes are very frequent. To solve the problem caused by route flaps, BGP uses route dampening 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 (the state change of a route from active to inactive is a route flap.), 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, that is, it is neither added into the routing table, nor advertised to other BGP peers.

The penalty value of the suppressed route will reduce to 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 in update packets.

Figure 142 BGP route dampening



Peer group

A peer group is a collection of peers with the same attributes. When a peer joins the peer group, the peer obtains the same configuration as the peer group. If configuration of the peer group is changed, configuration of group members is also changed.

There are many peers in a large BGP network. Some of these peers may be configured with identical commands. The peer group feature simplifies configuration of this kind.

When a peer is added into a peer group, the peer enjoys the same route update policy as the peer group, improving route distribution efficiency.

CAUTION: *If an option is configured both for a peer and for the peer group, the latest 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 the well-known community attribute, you can define the extended community attribute using a community list to help define a routing policy.

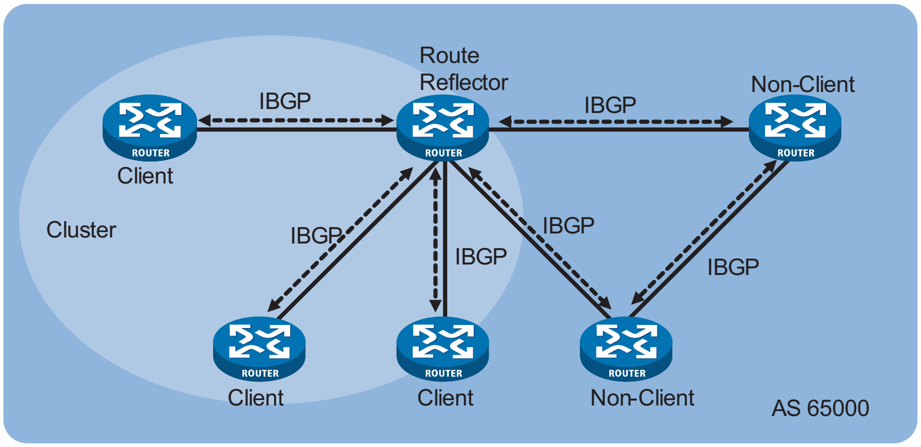
Route reflector

IBGP peers should be fully meshed to maintain connectivity. Suppose there are n routers in an AS, the number of IBGP connections is n(n-1)/2. If there are many IBGP peers, most network and CPU resources will be consumed.

Using route reflectors can solve the 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 (reflects) routing information between clients. BGP connections between clients need not be established.

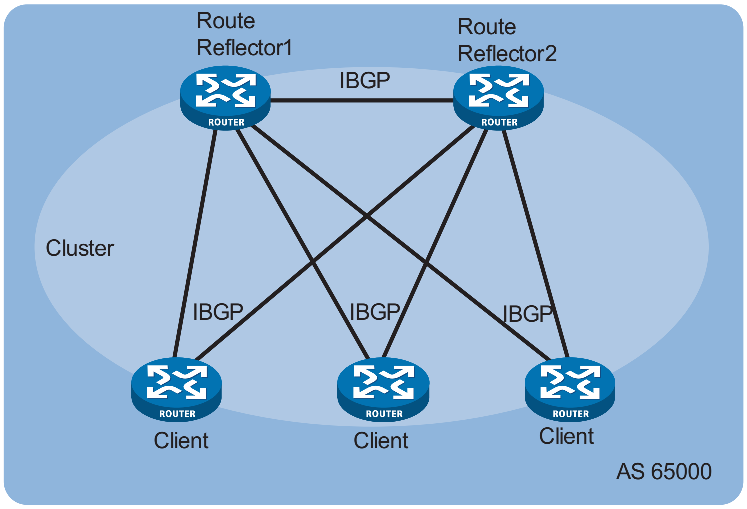
The router neither a route reflector nor a client is a non-client, which has to establish connections to all the route reflector and non-clients, as shown below.

Figure 143 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 144 Network diagram for route reflectors



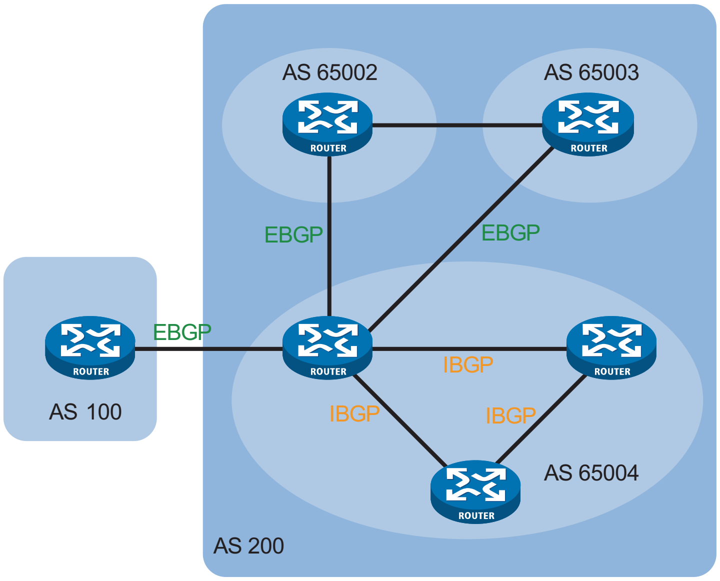
When clients of a route reflector are fully meshed, route reflection is unnecessary because it consumes more bandwidth resources. The system supports using related commands to disable route reflection in this case.

After route reflection is disabled between clients, routes between clients and non-clients can still be reflected.

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 EBGP connections are established between sub ASs, as shown below:

Figure 145 Confederation network diagram



From the perspective of a non-confederation 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, AS200 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.

MP-BGP

Overview

The legacy BGP-4 supports IPv4, but does not support some 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), which is defined in RFC2858.

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, namely NLRI, NEXT_HOP and 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 available 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, not forwarding them to peers.

Address family

MP-BGP employs address family 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, IPv6 extension. Different extensions are configured in respective address family view.

- *For information about the IPv6 extension application, refer to "IPv6 BGP Overview" on page 469.*
- *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

Features in draft stage include Graceful Restart and extended community attributes.

BGP Configuration Task List

To configure BGP, perform the tasks described in the following sections:

Task		Description
"Configuring BGP Basic Functions" on page 434		Required
"Controlling Route Distribution and Reception" on page 436	"Configuring BGP Route Redistribution" on page 436	Optional
	"Configuring BGP Route Summarization" on page 437	Optional
	"Advertising a Default Route to a Peer or Peer Group" on page 437	Optional
	"Configuring BGP Route Distribution Policy" on page 438	Optional
	"Configuring BGP Route Reception Policy" on page 438	Optional
	"Enabling BGP and IGP Route Synchronization" on page 439	Optional
	"Configuring BGP Route Dampening" on page 440	Optional
"Configuring BGP Routing Attributes" on page 440		Required
"Tuning and Optimizing BGP Networks" on page 442		Required
"Configuring a Large Scale BGP Network" on page 444	"Configuring BGP Peer Groups" on page 444	Optional
	"Configuring BGP Community" on page 445	Optional
	"Configuring a BGP Route Reflector" on page 446	Optional
	"Configuring a BGP Confederation" on page 446	Optional

Configuring BGP Basic Functions

The section describes BGP basic configuration.

- *This section does not differentiate between BGP and MP-BGP.*
- *Since BGP employs TCP, you need to specify IP addresses of peers, which may not be neighboring routers.*
- *Using logical links can also establish BGP peer relationships.*

- *In general, IP addresses of loopback interfaces are used to improve stability of BGP connections.*

Prerequisites

The neighboring nodes are accessible to each other at the network layer.

Configuration Procedure

To configure BGP basic functions, use the following commands:

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
Specify a Router ID		**Router-id** *ip-address*	Optional If no IP addresses are configured for loopback interface and other interfaces, the task becomes required
Specify the AS number for a peer or a peer group		**peer** { *group-name* | *ip-address* } **as-number** *as-number*	Required Not specified by default
Configure a description for a peer or a peer group		**peer** { *group-name* | *ip-address* } **description** *description-text*	Optional Not configured by default
Enable IPv4 unicast address family for all peers		**default ipv4-unicast**	Optional Enabled by default
Enable a peer		**peer** *ip-address* **enable**	Optional Enabled by default
Disable session with a peer or peer group		**peer** { *group-name* | *ip-address* } **ignore**	Optional Not disabled by default
Enable the logging on peer state changes	Enable BGP logging globally	**log-peer-change**	Optional Enabled by default
	Enable logging for a peer or peer group	**peer** { *group-name* | *ip-address* } **log-change**	Optional Enabled by default
Specify a preferred value for routes from a peer or peer group		**peer** { *group-name* | *ip-address* } **preferred-value** *value*	Optional The preferred value defaults to 0
Specify the source interface for route updates to a peer/peer group		**peer** { *group-name* | *ip-address* } **connect-interface** *interface-type interface-number*	Optional By default, BGP employs the source interface of the best routing updates
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. By specifying *hop-count,* you can specify the max hops for the EBGP connection

CAUTION:

- *It is required to specify for a BGP router a router ID, a 32-bit unsigned integer and the unique identifier of the router in the AS.*
- *You must create a peer group before configuring basic functions for it. For information about creating a peer group, refer to "Configuring BGP Peer Groups" on page 444.*
- *You can specify a router ID manually. If not, the system selects an IP address as the router ID. The selection sequence is the highest IP address among loopback interface addresses; if not available, then the highest IP address of interfaces. It is recommended to specify a loopback interface address as the router ID to enhance network reliability. Only when the interface with the selected Router ID or the manual Router ID is deleted will the system select another ID for the router.*
- *To guarantee updates sending in case of interface failure, you can specify the source interface of updates as a loopback interface.*
- *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. You do not need to use this command for directly connected EBGP peers, which employ loopback interfaces for peer relationship establishment.*

Controlling Route Distribution and Reception

Prerequisites

Before configuring this task, you have completed BGP basic configuration.

Configuring BGP Route Redistribution

BGP can advertise the routing information of the local AS to peering ASs, but it redistributes routing information from IGP into BGP routing table rather than self-finding. During route redistribution, BGP can filter routing information according to different routing protocols.

To configure BGP route redistribution, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	-
Enable BGP to redistribute default route into the BGP routing table	**default-route imported**	Optional Not enabled by default
Redistribute routes from another routing protocol for advertisement	**import-route** *protocol* [*process-id* [**med** *med-value* | **route-policy** *route-policy-name*] *]	Required Not redistributed by default
Advertise a network to the BGP routing table	**network** *ip-address* [*mask* | *mask-length*] [**short-cut** | **route-policy** *route-policy-name*]	Optional Not advertised by default

- *The ORIGIN attribute of routes redistributed using the* ***import-route*** *command is Incomplete.*
- *The ORIGIN attribute of networks advertised into the BGP routing table with the* ***network*** *command is IGP and these networks must exist in the local IP routing table. Using a routing policy makes routes control more flexible.*

Configuring BGP Route Summarization

To reduce the routing table size on medium and large BGP networks, you need to configure route summarization on peers. BGP supports two summarization types: automatic and manual.

- Automatic summarization: Summarizes redistributed IGP subnets. With the feature configured, BGP advertises only summary natural networks rather than subnets. The default route and routes imported using the **network** command can not be summarized.
- Manual summarization: Summarizes BGP local routes. The manual summary routes have higher priority than automatic ones.

To configure BGP route summarization, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter BGP view		**bgp** *as-number*	-
Configure BGP route summarization	Configure automatic route summarization	**summary automatic**	Required No route summarization is configured by default Choose either as needed; if both are configured, the manual route summarization takes effect.
	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*]*	

Advertising a Default Route to a Peer or Peer Group

To advertise a default route to a peer or peer group, use the following commands:

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

With the ***peer default-route-advertise*** *command executed, the 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.*

Configuring BGP Route Distribution Policy

To configure BGP route distribution policy, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	-
Filter redistributed routes when advertising them	**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; The filtering is 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 passing the first policy, can they go through the next; and only routes passing all the configured policies, can they be advertised.
Reference a routing policy to filter routes to a peer/peer group	**peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* **export**	
Reference an ACL to filer routing information to a peer/peer group	**peer** { *group-name* | *ip-address* } **filter-policy** *acl-number* **export**	
Reference an AS path ACL to filer routing information 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 to a peer/peer group	**peer** { *group-name* | *ip-address* } **ip-prefix** *ip-prefix-name* **export**	

CAUTION: *Only routes passing the filters can be advertised.*

Configuring BGP Route Reception Policy

To configure BGP routing reception policy, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	-

To do...	Use the command...	Remarks
Filter incoming routes	**filter-policy** { *acl-number* | **ip-prefix** *ip-prefix-name* } **import**	Required to choose any; No inbound 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 through 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** *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**	
Specify the maximum number of routes that can be received from a peer/peer group	**peer** { *group-name* | *ip-address* } **route-limit** *limit* [*percentage*]	The number is unlimited by default.

CAUTION:

- *Only routes permitted by the specified filter policy can be added into the local BGP routing table.*
- *Members of a peer group can have different inbound route filter policies from the peer group.*

Enabling BGP and IGP Route Synchronization

With this feature enabled, if a non BGP router is responsible for forwarding packets in the AS, the BGP speaker in the AS cannot advertise routes to external ASs unless all the 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. With BGP and IGP synchronization enabled, the BGP router cannot advertise the route to EBGP peers unless the route is also advertised by the IGP.

To configure BGP and IGP synchronization, use the following commands:

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

Configuring BGP Route Dampening

Through configuring BGP route dampening, you can suppress unstable routes to neither add them to the local routing table nor advertise them to BGP peers.

To configure BGP route dampening, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	-
Configure BGP route dampening	**dampening** [*half-life-reachable half-life-unreachable reuse suppress ceiling* | **route-policy** *route-policy-name*] *	Optional Not configured by default

Configuring BGP Routing Attributes

Prerequisites

Before configuring this task, you have configured BGP basic functions.

Configuration Procedure

You can use BGP route attributes to adjust BGP route selection policy.

To configure BGP route attributes, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter BGP view		**bgp** *as-number*	-
Configure preferences for external, internal, local routes		**preference** { *external-preference internal-preference local-preference* | **route-policy** *route-policy-name* }	Optional The default preferences of external, internal and local routes are 255, 255, 130 respectively.
Configure the default value of local preference		**default local-preference** *value*	Optional The value defaults to 100
Configure the MED attribute	Configure the default MED value	**default med** *med-value*	Optional The value defaults to 0
	Enable to compare MED values of routes from different ASs	**compare-different-as-med**	Optional Not enabled by default
	Enable to compare MED values of routes from each AS	**bestroute compare-med**	Optional Not enabled by default
	Enable to compare MED values of routes from confederation peers	**bestroute med-confederation**	Optional Not enabled by default

To do...		**Use the command...**	**Remarks**
Specify the router as the next hop of routes to a peer/peer group		**peer** { *group-name* | *ip-address* } **next-hop-local**	Optional By default, routes to an EBGP peer/peer group take the router as the next hop, while routes to an IBGP peer/peer group do not take the local router as the next hop.
Configure the AS_PATH attribute	Configure repeating times of local AS number in routes from a peer/peer group	**peer** { *group-name* | *ip-address* } **allow-as-loop** [*number*]	Optional The local AS number can not be repeated in routes from the peer/peer group.
	Disable the router from taking AS_PATH as a factor for best route selection	**bestroute as-path-neglect**	Optional By default, the router takes AS_PATH as a factor for best route selection
	Specify a fake AS number for a peer/peer group	**peer** { *group-name* | *ip-address* } **fake-as** *as-number*	Optional Not specified by default This command is only applicable to an EBGP peer or peer group.
	Substitute local AS number for the AS number of a peer/peer group in the AS_PATH attribute	**peer** { *group-name* | *ip-address* } **substitute-as**	Optional The substitution is not configured by default.
	Configure to not keep private AS number in AS_PATH of updates to a peer/peer group	**peer** { *group-name* | *ip-address* } **public-as-only**	Optional By default, BGP updates carry private AS number.

CAUTION:

- *Using a routing policy can set a preference for routes meeting its filtering conditions. Routes not meeting the conditions use the default preference.*
- *If other conditions are identical, the route with the smallest MED value is selected as the best external route of the AS.*
- *Using the* ***peer next-hop-local*** *command can specify the router as the next hop for a peer/peer group. If BGP load balancing is configured, the router specify itself as the next hop for routes to a 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 EBGP peers reside in a common broadcast subnet, the BGP router does not specify itself as the next hop for routes to the EBGP peer, unless the* ***peer next-hop-local*** *command is configured.*
- *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.*

- *You can specify a fake AS number to hide the real one as needed. The fake AS number applies to EBGP peers only, that is, EBGP peers in other ASs can only find the fake AS number.*
- *The **peer substitute-as** command is used only in specific networking environments. Inappropriate use of the command may cause routing loops.*

Tuning and Optimizing BGP Networks

This task involves the following parts:

1 Configure BGP timers

After establishing a 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 a BGP connection, the two parties compare their holdtimes, taking the shorter one as the common holdtime.

2 Reset BGP connections

After modifying a route selection policy, you have to reset BGP connections to make the new one take effect, causing a short time disconnection. The current BGP implementation supports the route-refresh capability. With this capability enabled on all BGP routers in a network, when a policy is modified on a router, the router advertises a route-refresh message to its peers, which then resend their routing information to the router. Therefore, the local router can perform dynamic route update and apply the new policy without tearing down BGP connections.

If a router not supporting route-refresh exists in the network, you need to configure the **peer keep-all-routes** command to save all route updates, and then use the **refresh bgp** command to soft reset BGP connections, which can refresh the BGP routing table and apply the new policy without tearing down BGP connections.

3 Configure BGP authentication

BGP employs TCP as the transport protocol. To enhance security, you can configure BGP to perform MD5 authentication when establishing a TCP connection. BGP MD5 authentication is not for BGP packets. It is used to set passwords for TCP connections. If the authentication fails, the TCP connection can not be established.

Prerequisites

Before configuring this task, you have configured BGP basic functions

Configuration Procedure

To tune and optimize BGP networks, use the following commands:

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 timers	Configure 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 a peer/peer group	**peer** { *group-name* | *ip-address* } **timer keepalive** *keepalive* **hold** *holdtime*	
Configure the interval for sending the same update to a peer/peer group		**peer** { *group-name* | *ip-address* } **route-update-interval** *seconds*	Optional The intervals for sending the same update to an IBGP peer and an EBGP peer default to 15 seconds and 30 seconds respectively.
Configure BGP soft reset	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
	Enable BGP route refresh for a peer/peer group	**peer** { *group-name* | *ip-address* } **capability-advertise route-refresh**	Optional Enabled by default
	Keep all original routes imported from a peer/peer group regardless of whether they pass the inbound filtering policy	**peer** { *group-name* | *ip-address* } **keep-all-routes**	Optional Not kept 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
	Enter system view	**system-view**	-
	Enter BGP view	**bgp** *as-number*	-
Clear the direct EBGP session on any interface that becomes down		**ebgp-interface-sensitive**	Optional The function is enabled by default
Perform MD5 authentication when establishing a TCP connection		**peer** { *group-name* | *ip-address* } **password** { **cipher** | **simple** } *password*	Optional Not performed by default
Configure the number of BGP load balanced routes		**balance** *number*	Optional Load balancing is not enabled by default.

CAUTION:

- *The maximum keepalive interval should be 1/3 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.*
- *Use of the **peer keep-all-routes** command saves all routing updates from the peer regardless of whether the filtering policy is configured. The system uses these updates to rebuild the routing table after a soft reset is triggered.*
- *Performing BGP soft reset can refresh the routing table and apply the new policy without tearing down BGP sessions.*
- *BGP soft reset requires all routers in the network have the route-refresh capability. If not, you need use the **peer keep-all-routes** command to keep all routing information from a BGP peer to perform soft reset.*
- *Configured in BGP view, MD5 authentication also applies to the MP-BGP VPNv4 extension, because the same TCP connection is used.*

Configuring a Large Scale BGP Network

In a large-scale BGP network, configuration and maintenance become difficult due to so 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.

Configuring a BGP community can also help simplify routing policy management, and a community has much larger management range than a peer group by controlling routing policies of multiple BGP routers.

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 a route reflector or confederation can solve it. In a large-scale AS, both of them can be used.

Configuration Prerequisites

Before configuring this task, you have made network layer accessible on peering nodes.

Configuring BGP Peer Groups

To do so, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter BGP view		**bgp** *as-number*	-
Configure an IBGP peer group	Create an IBGP peer group	**group** *group-name* [**internal**]	Optional You can add multiple peers into the group. The system will create these peers automatically and specify the local AS number as their AS in BGP view.
	Add a peer into the IBGP peer group	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	

To do...		Use the command...	Remarks
Configure a pure EBGP peer group	Create an EBGP peer group	**group** *group-name* **external**	Optional You can add multiple peers into the group. The system will create these peers automatically and specify the local AS number as their AS in BGP view.
	Specify the AS number for the group	**peer** *group-name* **as-number** *as-number*	
	Add a peer into the group	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	
Configure a mixed EBGP peer group	Create an EBGP peer group	**group** *group-name* **external**	Optional You can add multiple peers into the group.
	Specify a peer and the AS number for the peer respectively	**peer** *ip-address* **as-number** *as-number*	
	Add a peer into the group	**peer** *ip-address* **group** *group-name* [**as-number** *as-number*]	

CAUTION:

- *You need not specify the AS number when creating an IBGP peer group.*
- *If there are peers in a peer group, you can neither change the AS number of the group nor use the **undo** command to remove the AS number*
- *You need to specify the AS number for each peer in a mixed EBGP peer group respectively.*

Configuring BGP Community

To configure BGP community, use the following commands:

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

CAUTION:

- *When configuring 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" on page 243.*

Configuring a BGP Route Reflector

To configure a BGP route reflector, use the following commands:

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 to avoid routing loops.*

Configuring a BGP Confederation

To configure a BGP confederation, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter BGP view		**bgp** *as-number*	-
Configure a BGP confederation	Configure a confederation ID	**confederation id** *as-number*	Required Not configured by default
	Specify sub ASs contained in the confederation	**confederation peer-as** *as-number-list*	
Enable compatibility with AS confederation not compliant with RFC 3065		**confederation nonstandard**	Optional By default, a confederation complies with RFC 3065.

CAUTION:

- *A confederation contains 32 sub ASs at most. The as-number of a sub AS takes effect in the confederation only.*
- *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.*

Displaying and Maintaining BGP Configuration

Displaying BGP Configuration

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 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 dampening routing information and release suppressed routes	**reset bgp** [**vpn-instance** *vpn-instance-name*] **dampening** [*network-address* [*mask* | *mask-length*]]	Available in user view
Clear route flap information	**reset bgp** [**vpn-instance** *vpn-instance-name*] **flap-info** { **regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number* | *ip-address* [*mask* | *mask-length*] }	

BGP Configuration Examples

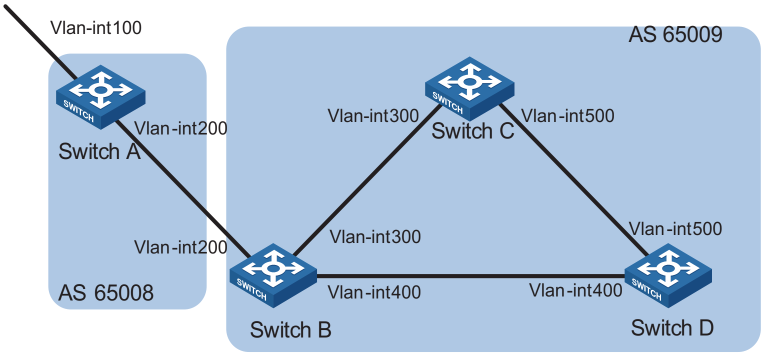
BGP Basic Configuration

Network requirements

In the following figure are all BGP switches. Between Switch A and Switch B is an EBGP connection. Switch B, Switch C and Switch D are IBGP fully meshed.

Network diagram

Figure 146 Network diagram for BGP basic configuration (on switches)



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	8.1.1.1/8	Switch D	Vlan-int400	9.1.1.2/24

	Vlan-int200	200.1.1.2/24		Vlan-int500	9.1.2.2/24
Switch B	Vlan-int400	9.1.1.1/24	Switch C	Vlan-int500	9.1.2.1/24
	Vlan-int200	200.1.1.1/24		Vlan-int300	9.1.3.2/24
	Vlan-int300	9.1.3.1/24			

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
```

# Advertise 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 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 routers.

# 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 routing table information 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
```

From the above outputs, you can find Switch A learned no route to ASwitch 7750 Family9, and Switch C learned network 8.0.0.0 but the next hop 200.1.1.2 is unreachable, thus 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
```

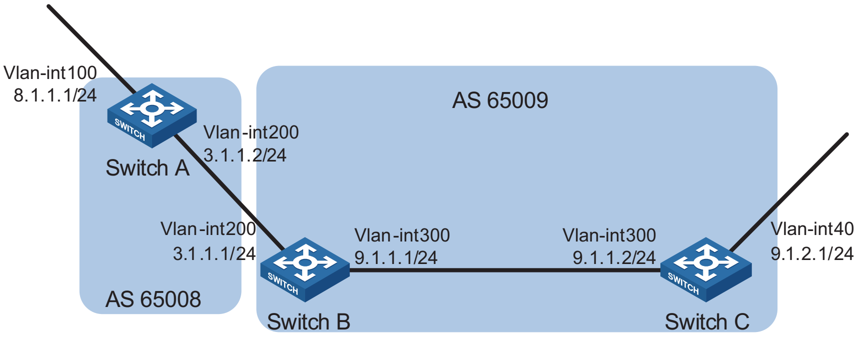
BGP and IGP Interaction Configuration

Network requirements

As shown below, OSPF is used as the IGP protocol in ASwitch 7750 Family9, where Switch C is a non-BGP switch. Between Switch A and Switch B is an EBGP connection.

Network diagram

Figure 147 Network diagram for BGP and IGP interaction



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
```

# Advertise 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 interaction

# 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?
 *>   9.1.2.0/24        3.1.1.1       2                   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       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
```

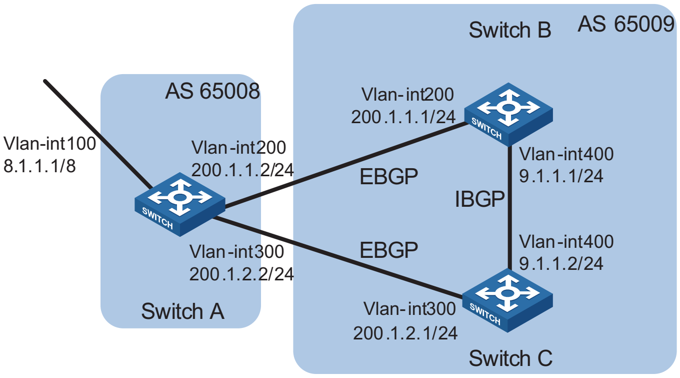
BGP Load Balancing and MED Attribute Configuration

Network requirements

- Configure BGP on all switches; Switch A is in ASwitch 7750 Family8, Switch B and C in ASwitch 7750 Family9.
- Between Switch A and B, Switch A and C are EBGP connections, and IBGP runs between Switch B and C.

Network diagram

Figure 148 Network diagram for BGP path selection configuration



Configuration procedure

1 Configure IP addresses for interfaces (omitted)

2 Configure BGP connections

# Configure SwitchA

```
<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
```

# Advertise 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 SwitchB

```
<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 SwitchC

```
<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
```

Two routes to 9.1.1.0/24 are available, and the one with the next hop being 200.1.1.1 is the optimal because the ID of SwitchB is smaller.

3 Configure loading balancing

# Configure SwitchA

```
[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
```

The route 9.1.1.0/24 has two next hops 200.1.1.1 and 200.1.2.1, and both are the optimal.

4 Configure MED

# Configure the default MED of SwitchB.

```
[SwitchB] bgp 65009
[SwitchB-bgp] default med 100
```

# Display the routing table on SwitchA.

```
[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         200.1.2.1     0                     0       65009i
 *                       200.1.1.1     100                   0       65009i
```

From the above information, you can find the route with the next hop 200.1.2.1 is the best route, because its MED (0) is smaller than the MED (100) of the other route with the next hop 200.1.1.1 (Switch B).

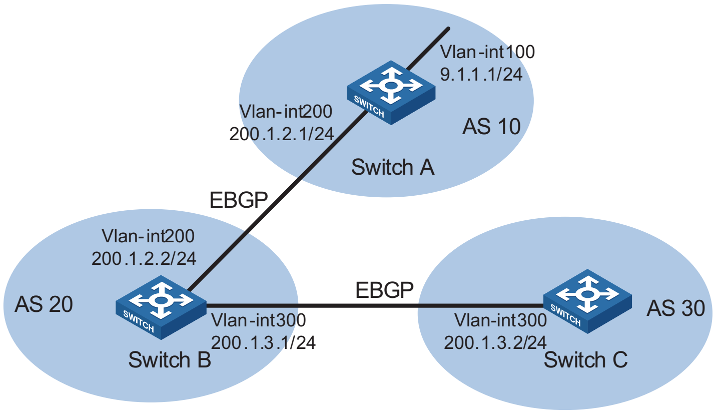
BGP Community Configuration

Network requirements

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.

Network diagram

Figure 149 Network diagram for BGP community configuration (on switches)



Configuration procedure

1 Configure IP addresses for interfaces (omitted)

2 Configure EBGP

# Configure SwitchA

```
<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 SwitchB

```
<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 SwitchC

```
<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 SwitchB.

```
[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 SwitchC in AS30.

# Display the routing table on SwitchC.

```
[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       20 10i
```

Switch C learned route 9.1.1.0/24 from SwitchB.

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 SwitchB.

```
[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 SwitchC.

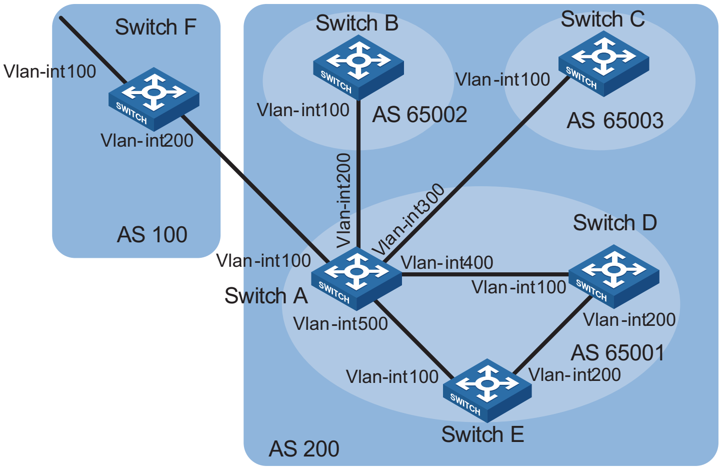
BGP Confederation Configuration

Network requirements

To reduce IBGP connections in AS 200, split it into three sub ASs, ASwitch 7750 Family1, ASwitch 7750 Family2 and ASwitch 7750 Family3. Switches in ASwitch 7750 Family1 are fully meshed.

Network diagram

Figure 150 Network diagram for BGP confederation configuration (on switches)



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	200.1.1.1/24	Switch D	Vlan-int100	10.1.3.2/24
	Vlan-int200	10.1.1.1/24		Vlan-int200	10.1.5.1/24
	Vlan-int300	10.1.2.1/24	Switch E	Vlan-int100	10.1.4.2/24
	Vlan-int400	10.1.3.1/24		Vlan-int200	10.1.5.2/24
	Vlan-int500	10.1.4.1/24	Switch F	Vlan-int100	9.1.1.1/24
Switch B	Vlan-int100	10.1.1.2/24		Vlan-int200	200.1.1.2/24
Switch C	Vlan-int100	10.1.2.2/24			

Configuration procedure

1. Configure IP addresses for interfaces (omitted)
2. Configure BGP confederation

# Configure SwitchA

```
<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 SwitchB

```
<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 SwitchC

```
<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 ASwitch 7750 Family1.

# Configure SwitchA

```
[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 SwitchD

```
<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 SwitchE

```
<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 SwitchA

```
[SwitchA] bgp 65001
[SwitchA-bgp] peer 200.1.1.2 as-number 100
[SwitchA-bgp] quit
```

# Configure SwitchF

```
<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 of SwitchB.

```
[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 SwitchD.

```
[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
```

BGP Route Reflector Configuration

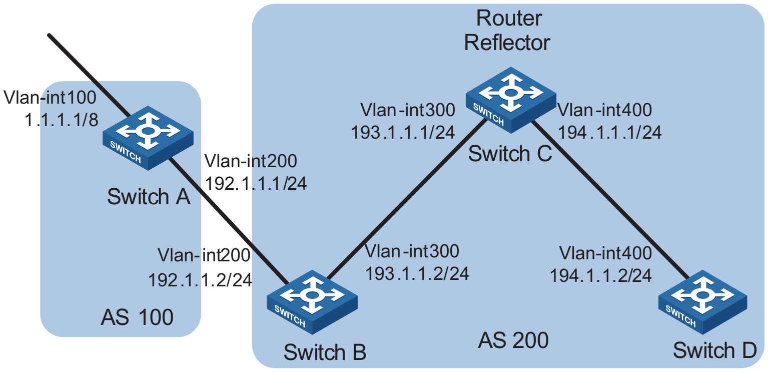
Network requirements

In the following figure, all switches run BGP.

- Between SwitchA and SwitchB is an EBGP connection, between SwitchC and SwitchB, SwitchC and SwitchD are IBGP connections.
- SwitchC is a route reflector with clients SwitchB and D.
- SwitchD can learn route 1.0.0.0/8 from SwitchC.

Network diagram

Figure 151 Network diagram for BGP route reflector configuration



Configuration procedure

1. Configure IP addresses for interfaces (omitted)
2. Configure BGP connections

# Configure SwitchA

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 192.1.1.2 as-number 200
```

# Advertise network 1.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp] network 1.0.0.0
[SwitchA-bgp] quit
```

# Configure SwitchB

```
<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 SwitchC

```
<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 SwitchD

```
<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 SwitchC

```
[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 of SwitchB.

```
[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 of SwitchD.

```
[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
```

SwitchD learned route 1.0.0.0/8 from SwitchC.

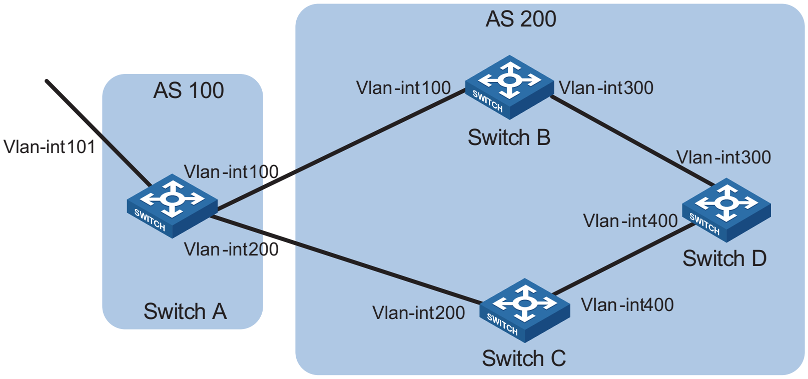
BGP Path Selection Configuration

Network requirements

- In the figure below, all switches run BGP. Between Switch A and Switch B, Switch A and Switch C are EBGP connections. Between Switch B and Switch D, Switch D and Switch C are IBGP connections.
- OSPF is the IGP protocol in AS 200.
- Configure routing policies, making Switch D give priority to the route 1.0.0.0/8 from Switch C.

Network diagram

Figure 152 Network diagram for BGP path selection configuration (on switches)



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
```

# Advertise network 1.0.0.0/8 to the BGP routing table of 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 SwitchD give priority to the route learned from SwitchC.

- 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 of SwitchD.

```
[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 learned from Switch C is the optimal.

- Configure different local preferences on Switch B and C for route 1.0.0.0/8, making SwitchD 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, set 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 the 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 learned from SwitchC is the optimal.

Troubleshooting BGP Configuration

No BGP Peer Relationship Established

Symptom

Display BGP peer information using the **display bgp peer** command. The state of the connection to the 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.

Processing steps

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.

36 IPv6 BGP CONFIGURATION

- *The term "router" refers to a router in a generic sense or an Ethernet switch running routing protocols in this document.*
- *This chapter describes only configuration specific to IPv6 BGP. For BGP related information, refer to “BGP Configuration” on page 419.*

When configuring IPv6 BGP, go to these sections for information you are interested in:

IPv6 BGP Overview

BGP-4 manages only IPv4 routing information, thus other network layer protocols such as IPv6 are not supported.

To support multiple network layer protocols, IETF extended BGP-4 by introducing IPv6 BGP that is defined in RFC 2858 (Multiprotocol Extensions for BGP-4).

To implement IPv6 support, IPv6 BGP puts IPv6 network layer information into the attributes of Network Layer Reachable Information (NLRI) and NEXT_HOP.

NLRI attribute of IPv6 BGP involves:

- MP_REACH_NLRI: Multiprotocol Reachable NLRI, for advertisement of next hop information of reachable routes.
- 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 utilizes BGP multiprotocol extensions for application in IPv6 networks. The original messaging and routing mechanisms of BGP are not changed.

Configuration Task List

Task		Description
"Configuring IPv6 BGP Basic Functions" on page 471	"Configuring an IPv6 Peer" on page 471	Required
	"Advertising a Local IPv6 Route" on page 471	Optional
	"Configuring a Preferred Value for Routes from a Peer/Peer Group" on page 472	Optional
	"Specifying a Local Update Source Interface to a Peer/Peer Group" on page 472	Optional
	"Configuring a Non Direct EBGP Connection to a Peer/Peer Group" on page 472	Optional
	"Configuring Description for a Peer/Peer Group" on page 473	Optional
	"Disabling Session Establishment to a Peer/Peer Group" on page 473	Optional
	"Logging Session State and Event Information of a Peer/Peer Group" on page 473	Optional
"Controlling Route Distribution and Reception" on page 473	"Configuring IPv6 BGP Route Redistribution" on page 474	Optional
	"Advertising a Default Route to a Peer/Peer Group" on page 474	Optional
	"Configuring Route Distribution Policy" on page 474	Optional
	"Configuring Route Reception Policy" on page 475	Optional
	"Configuring IPv6 BGP and IGP Route Synchronization" on page 476	Optional
	"Configuring Route Dampening" on page 476	Optional
"Configuring IPv6 BGP Route Attributes" on page 476	"Configuring IPv6 BGP Preference and Default LOCAL_PREF and NEXT_HOP Attributes" on page 477	Optional
	"Configuring the MED Attribute" on page 477	Optional
	"Configuring the AS_PATH Attribute" on page 477	Optional

Task		Description
"Tuning and Optimizing IPv6 BGP Networks" on page 478	"Configuring IPv6 BGP Timers" on page 479	Optional
	"Configuring IPv6 BGP Soft Reset" on page 479	Optional
	"Configuring the Maximum Number of Equal Cost Routes" on page 480	Optional
"Configuring a Large Scale IPv6 BGP Network" on page 480	"Configuring IPv6 BGP Peer Group" on page 480	Optional
	"Configuring IPv6 BGP Community" on page 482	Optional
	"Configuring an IPv6 BGP Route Reflector" on page 482	Optional

Configuring IPv6 BGP Basic Functions

Prerequisites

Before configuring this task, you need to:

- Specify IP addresses for interfaces.
- Enable IPv6.

You need create a peer group before configuring basic functions for it. For related information, refer to "Configuring IPv6 BGP Peer Group" on page 480.

Configuring an IPv6 Peer

To configure an IPv6 peer, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required Not enabled by default
Specify a router ID	**router-id** *router-id*	Optional Required if no IP addresses configured for Loopback interface and other 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

Advertising a Local IPv6 Route

To advertise a local route into the routing table, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-

To do...	Use the command...	Remarks
Add a local route into 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

To configure a preferred value for routes received from a peer/peer group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure a preferred value for routes received from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value*	Optional By default, the preferred value is 0.

Specifying a Local Update Source Interface to a Peer/Peer Group

To specify a local update source interface connected to a peer, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Specify a local update source interface connected to a peer	**peer** { *ipv6-group-name* | *ipv6-address* } **connect-interface** *interface-type interface-number*	Required By default, the source interface of the optimal updates is used.

To improve stability and reliability, you can specify the local interface of an IPv6 BGP connection as loopback interface. By doing so, a connection failure upon redundancy availability will not affect IPv6 BGP connection.

Configuring a Non Direct EBGP Connection to a Peer/Peer Group

To configure an EBGP connection to a peer not directly connected, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure a non direct EBGP connection to a 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 connection with loopback interfaces.*

Configuring Description for a Peer/Peer Group

To configure description for a peer/peer group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure description for a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **description** *description-text*	Optional Not configured by default

The peer group for which to configure a description must have been created.

Disabling Session Establishment to a Peer/Peer Group

To disable session establishment to a peer/peer group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Disable session establishment to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ignore**	Optional Not disabled by default

Logging Session State and Event Information of a Peer/Peer Group

To log the session and event information of a peer/peer group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enable global logging	**log-peer-change**	Optional Enabled by default
Enter IPv6 address family view	**ipv6-family**	-
Enable to log session and event information of a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **log-change**	Optional Enabled by default

*Refer to the Switch 8800 Command Reference Guide 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 have:

- Enabled the IPv6 function
- Configured the IPv6 BGP basic functions

Configuring IPv6 BGP Route Redistribution

To configure IPv6 BGP route redistribution and filtering, use the following commands:

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

*If the **default-route imported** command is not configured, using the **import-route** command cannot redistribute any IGP default route.*

Advertising a Default Route to a Peer/Peer Group

To advertise default route to a peer/peer group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Advertise a default route to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **default-route-advertise** [**route-policy** *route-policy-name*]	Required Not advertised by default

*With the **peer default-route-advertise** command used, the local router advertises a default route with itself as the next hop to the specified peer/peer group, regardless of whether the default route is available in the routing table.*

Configuring Route Distribution Policy

To configure policies for route distribution, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure outbound route filtering	**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 a 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 a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **export**	Required Not specified by default

To do...	Use the command...	Remarks
Specify an AS path ACL to filer routes advertised to a 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 a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **ipv6-prefix** *ipv6-prefix-name* **export**	Required Not specified by default

- *After configuring the filtering of routes to a peer group, you can also configure the filtering of routes to a member of the peer group, and the last configuration takes effect.*
- *IPv6 BGP advertises routes passing the specified policy to peers. Using the protocol argument can filter only the specified protocol routes. If no protocol specified, IPv6 BGP filters all routes to be advertised, including redistributed routes and routes imported using the **network** command.*

Configuring Route Reception Policy

To configure route reception policy, use the following commands:

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 a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **import**	Required Not applied by default
Reference an ACL to filter routes imported from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **filter-policy** *acl6-number* **import**	Required Not specified by default
Reference an AS path ACL to filter routing information imported from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **as-path-acl** *as-path-acl-number* **import**	Required Not specified by default
Reference an IPv6 prefix list to filter routing information imported from a 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 address prefixes imported from a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-limit** *limit* [*percentage*]	Optional By default, the number of prefixes is unlimited. If the received IPv6 prefixes exceed the upper limit, the neighbor is still maintained but the exceeding routes will be discarded.

- *Only routes passing the specified policy can be added into the local IPv6 BGP routing table.*
- *Members of a peer group can have different inbound route 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.

To configure IPv6 BGP and IGP route synchronization, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Enable route synchronization between IPv6 BGP and IGP	**synchronization**	Required Not enabled by default

Configuring Route Dampening

To configure BGP route dampening, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
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

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
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 value for local preference	**default local-preference** *value*	Optional The *value* defaults to 100
Advertise routes to a peer/peer group with the local router as the next hop	**peer** { *ipv6-group-name* | *ipv6-address* } **next-hop-local**	Required By default, the feature is available for routes advertised to the EBGP peer/peer group, but not available to the IBGP peer/peer group

- *To make sure an IBGP peer can find the correct next hop, you can configure routes advertised to the peer 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 outbound routes to a 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 EBGP peers reside in a common broadcast subnet, the router does not specify itself as the next hop for routes to the EBGP peer by default, unless the **peer next-hop-local** command is configured.*

Configuring the MED Attribute

To configure the MED attribute, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure a default MED value	**default med** *med-value*	Optional Defaults to 0
Enable to compare MED values of routes from different EBGP peers	**compare-different-as-med**	Optional Not enabled by default
Prioritize MED values of routes from each AS	**bestroute compare-med**	Optional Not configured by default
Prioritize MED values of routes from confederation peers	**bestroute med-confederation**	Optional Not configured by default

Configuring the AS_PATH Attribute

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
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 a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **fake-as** *as-nmber*	Optional Not specified by default
Neglect the AS_PATH attribute for best route selection	**bestroute as-path-neglect**	Optional Not neglected 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, BGP updates carry private AS number
Substitute local AS number for the AS number of a peer/peer group indicated 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 router not supporting route-refresh exists in the network, you need to configure the **peer keep-all-routes** command on the router to save all route updates, and then use the **refresh bgp ipv6** command to soft-reset IPv6 BGP connections.

Prerequisites

Before configuring IPv6 BGP timers, you have:

- Enabled IPv6 function
- Configured IPv6 BGP basic functions

Configuring IPv6 BGP Timers

To do so, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter BGP view		**bgp** *as-number*	Required
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 a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **timer keepalive** *keepalive* **hold** *holdtime*	
Configure the interval for sending the same update to a peer/peer group		**peer** { *ipv6-group-name* | *ipv6-address* } **route-update-interval** *seconds*	Optional The interval for sending the same update to an IBGP peer or an EBGP peer defaults to 15 seconds or 30 seconds

- *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

To enable route refresh, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
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

To perform manual soft reset, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-

To do...	Use the command...	Remarks
Save all routes from a 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** }	

Configuring the Maximum Number of Equal Cost Routes

Perform these commands 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*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure the maximum number of equal cost routes for load balancing	**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.

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 BGP, so it is not mentioned here. The following describes:

- Configuring an IPv6 BGP peer group
- Configuring the IPv6 BGP community
- Configuring an IPv6 BGP route reflector

Prerequisites

Before configuring an IPv6 BGP peer group, you have:

- Made peer nodes accessible at the network layer
- Enabled BGP and configured router ID.

Configuring IPv6 BGP Peer Group

Create an IBGP peer group

To create an IBGP group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required Not enabled by default
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

After you add an IPv6 IBGP peer to the peer group, the system will automatically create the peer in BGP view and enable the IPv6 peer in IPv6 address family view.

Create a pure EBGP peer group

To configure a pure EBGP group, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required Not enabled by default
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

- *After you add an IPv6 EBGP peer to the peer group, the system will automatically create the EBGP peer in BGP view and enable the EBGP peer in IPv6 address family view.*
- *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.*

Create a mixed EBGP peer group

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required Not enabled by default
Enter IPv6 address family view	**ipv6-family**	-
Create an EBGP peer group	**group** *ipv6-group-name* **external**	Required

To do...	Use the command...	Remarks
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

- *After you add an IPv6 EBGP peer to the peer group, the system will automatically create the EBGP peer in IPv6 address family view.*
- *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 a peer/peer group

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required Not enabled by default
Enter IPv6 address family view	**ipv6-family**	-
Advertise community attribute to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **advertise-community**	Required Not advertised by default
Advertise extended community attribute to a 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

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Apply a routing policy to routes advertised to a peer/peer group	**peer** { *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* **export**	Required Not applied by default

- *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" on page 243.*

Configuring an IPv6 BGP Route Reflector

To configure an IPv6 BGP route reflector, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter BGP view	**bgp** *as-number*	Required
Enter IPv6 address family view	**ipv6-family**	-
Configure the router as a route reflector and specify a 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

- *In general, since the route reflector forwards routing information between clients, it is not required to make clients 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 Configuration

Displaying IPv6 BGP Configuration

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** [*ipv6-address* { **log-info** | **verbose** } | *ipv6-group-name* **log-info** | **verbose**]	
Display IPv6 BGP routing table information	**display bgp ipv6 routing-table** [*ipv6-address prefix-length*]	
Display IPv6 BGP routing information matching a AS path ACL	**display bgp ipv6 routing-table as-path-acl** *as-path-acl-number*	

To do...	Use the command...	Remarks
Display IPv6 BGP community routing information	**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 labeled IPv6 BGP routing information	**display bgp ipv6 routing-table label**	
Display IPv6 BGP routing information to or from an IPv6 BGP peer	**display bgp ipv6 routing-table peer** *ipv6-address* { **advertised-routes** | **received-routes** } [*network-address prefix-length* | **statistic**]	
Display IPv6 BGP routing information matching the 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
Reset all IPv6 BGP connections	**reset bgp ipv6 all**	Available in user view
Reset IPv6 BGP connections to an AS	**reset bgp ipv6** *as-number*	
Reset the IPv6 BGP connection to a peer	**reset bgp ipv6** *ipv6-address* [**flap-info**]	
Reset all IPv6 EBGP connections	**reset bgp ipv6 external**	
Reset the IPv6 BGP connections to an IPv6 peer group	**reset bgp ipv6 group** *ipv6-group-name*	
Reset all IPv6 IBGP connections	**reset bgp ipv6 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* | **regexp** *as-path-regexp* | **as-path-acl** *as-path-acl-number*]	

IPv6 BGP Configuration Examples

Some examples for IPv6 BGP configuration are similar to those of BGP4, so refer to "BGP Configuration" on page 419 for related information.

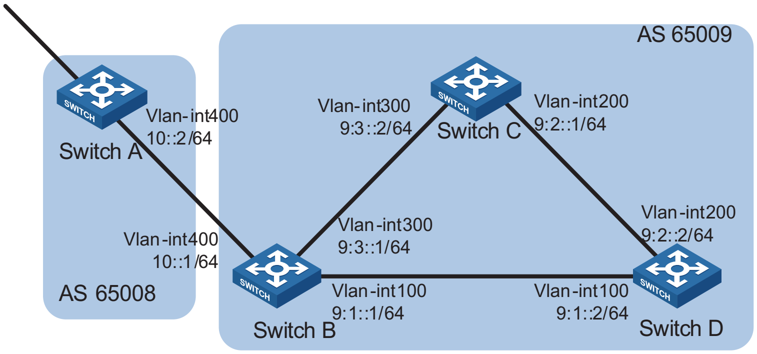
IPv6 BGP Basic Configuration

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 IBGP fully meshed.

Network diagram

Figure 153 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 established an EBGP connection; Switch B, C and D established IBGP connections with each other.

IPv6 BGP Route Reflector Configuration

Network requirements

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.

Network diagram

Figure 154 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] ipv6-family
[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.

37 MULTICAST OVERVIEW

- *The term "router" in this document refers to a router in a generic sense or a Switch 8800 running the multicast routing protocol.*
- *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.

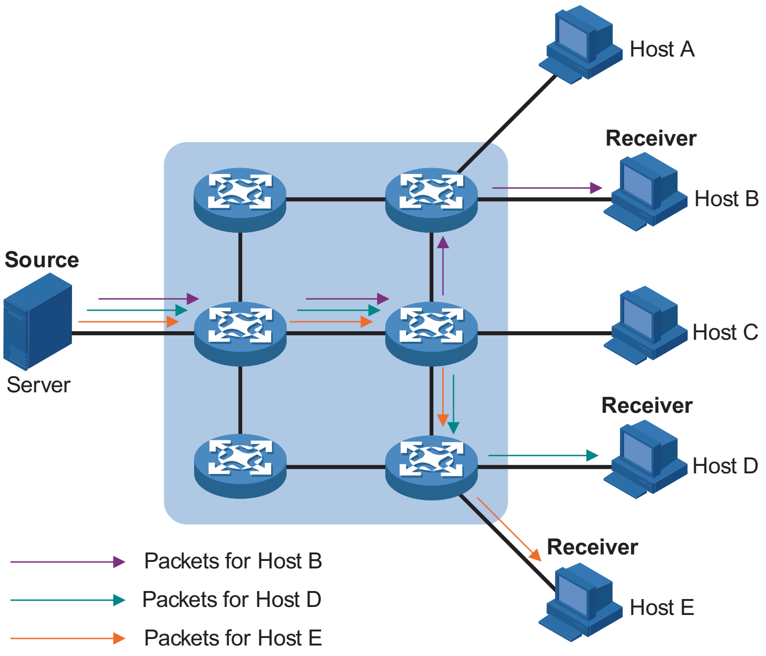
With the multicast technology, a network operator can easily provide new value-added services, such as live Webcasting, Web TV, distance learning, telemedicine, Web radio, real-time videoconferencing, and other information services that have high demands on the bandwidth and real-time data communication.

Comparison of Information Transmission Techniques

Unicast

In unicast, the information source sends a separate copy of information to each host that needs the information, as shown in Figure 155.

Figure 155 Unicast transmission



Assume that Hosts B, D and E need this information. The information source establishes a separate transmission channel for each of these hosts.

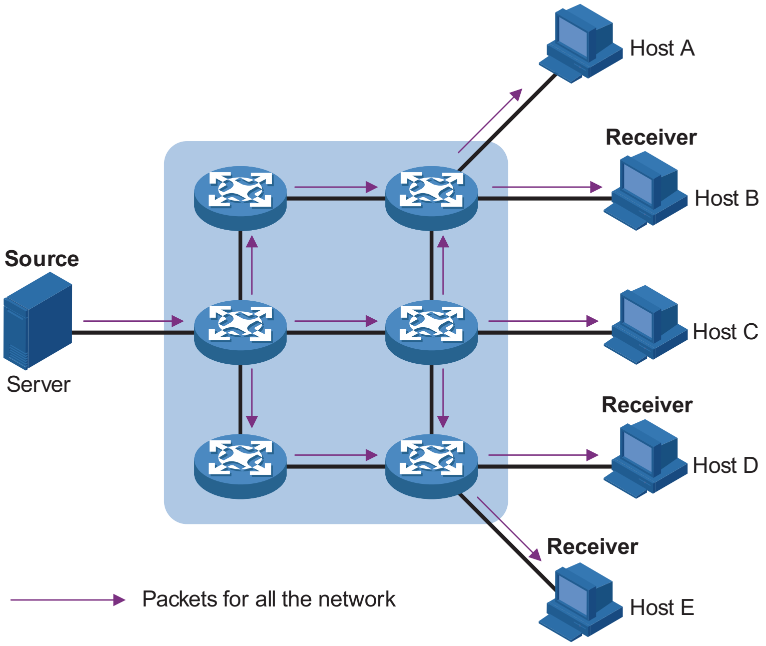
In unicast transmission, the traffic over the network is proportional to the number of hosts that need the information. If a large number of users need the information, the information source needs to send a copy of the same information to each of these users. This means a tremendous pressure on the 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 network, even if some hosts do not need the information, as shown in Figure 156.

Figure 156 Broadcast transmission



Assume that only Hosts B, D, and E need the information. If the information source broadcasts the information, Hosts A and C also receive it. In addition to information security issues, this also causes traffic flooding on the same network.

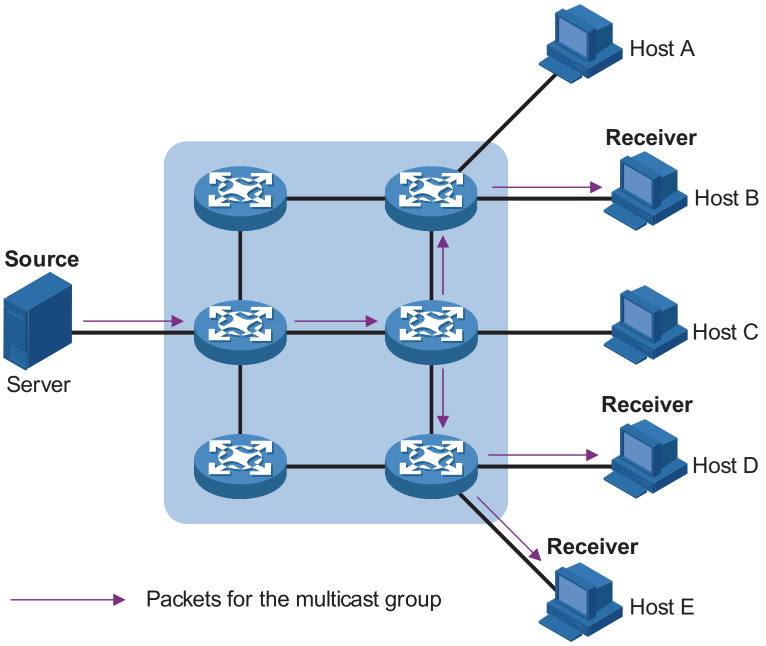
Therefore, broadcast is disadvantageous in transmitting data to specific hosts; moreover, broadcast transmission is a significant usage of network resources.

Multicast

As discussed above, the unicast and broadcast techniques are unable to provide point-to-multipoint data transmissions with the minimum network consumption.

The multicast technique has solved this problem. When some hosts on the network need the information, the multicast source (namely, the information source) sends only one copy of the information. With tree-type routes established for multicast packets through multicast routing protocols, the packets are replicated only where the tree branches, as shown in Figure 157:

Figure 157 Multicast transmission



Assume that Hosts B, D and E need the information. To receive the information correctly, these hosts need to join a receiver set, which is known as a multicast group. The multicast routers on the network duplicate and forward the information based on the distribution of the receivers in this set. Finally, the information is correctly delivered to Hosts B, D, and E.

To sum up, multicast has the following advantages:

- 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 remarkably add to the network load.
- Over broadcast: As multicast data is sent only to the receivers that need it, multicast uses the network bandwidth reasonably and brings no waste of network resources, and enhances network security.

Roles in Multicast

The following roles are involved in multicast transmission:

- An information sender is referred to as a Multicast Source ("Source" in Figure 157).
- Each receiver is a Multicast Group Member ("Receiver" in Figure 157).
- All receivers interested in the same information form a Multicast Group. Multicast groups are not subject to geographic restrictions.
- A router that supports Layer 3 multicast is called multicast router or Layer 3 multicast device. In addition to providing the multicast routing function, a multicast router can also manage multicast group members. In practice, switches that support Layer 3 multicast also act as Layer 3 multicast devices.

For a better understanding of the multicast concept, you can assimilate multicast transmission to the transmission of TV programs, as shown in Table 24.

Table 24 An analogy between TV transmission and multicast transmission

Step	TV transmission	Multicast transmission
1	A TV station transmits a TV program through a channel.	A multicast source sends multicast data to a multicast group.
2	A user tunes the TV set to the channel.	A receiver joins the multicast group.
3	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 sends to the multicast group.
4	The user turns off the TV set or tunes to another channel.	The receiver leaves the multicast group or joins another group.

- *A multicast source does not necessarily belong to a multicast group. Namely, a multicast source is not necessarily a multicast data receiver.*
- *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.*

Advantages and Applications of Multicast

Advantages of multicast

Advantages of the multicast technique include:

- Enhanced efficiency: reduces the CPU load of information sources and network devices.
- Optimal performance: reduces redundant traffic.
- Distributive application: Enables point-to-multiple-point applications at the price of the 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 point-to-multiple-point data distribution application.

Multicast Models

Based on the multicast source processing modes, there are three multicast models:

- Any-Source Multicast (ASM)
- Source-Filtered Multicast (SFM)
- Source-Specific Multicast (SSM)

ASM model

In the ASM model, any sender can become a multicast source and send information to a multicast group; 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 ware of the position of a multicast source in advance. However, they can join or leave the multicast group at any time.

SFM model

The SFM model is derived from the ASM model. From the view of a sender, the two models have the same multicast group membership architecture.

Functionally, the SFM model is an extension of the ASM model. In the SFM model, the upper layer software checks the source address of received multicast packets so as to permit or deny 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 practice, 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 module, and dedicated multicast forwarding paths are established between receivers and the specified multicast sources.

For details about the concepts of SPT and RPT, refer to "PIM Configuration" on page 563.

Multicast Architecture

Multicast Mechanism

The purpose of an IPv6 multicast technology is to carry information, by multicast, from a multicast source to the receivers.

IP multicast involves the following questions:

- Where should the multicast source transmit information to? (multicast addressing)
- What receivers exist on the network? (host registration)
- 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.

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, as shown in Table 25.

Table 25 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.
224.0.1.0 to 231.255.255.255 233.0.0.0 to 238.255.255.255	ASM/SFM multicast addresses available for users (IP addresses for temporary groups). They are valid across the internet.
232.0.0.0 to 232.255.255.255	SSM multicast addresses available for users (IP addresses of temporary groups). They are valid across the internet.
239.0.0.0 to 239.255.255.255	User-available administratively scoped multicast addresses for ASM/SFM.

Like the 10.0.0.0/8 block that IANA has reserved for IP unicast, 239.0.0.0/8 is an IP multicast address block reserved by IANA. These addresses are administratively scoped addresses. The use of the administratively scoped multicast addresses allows flexible definition of the ranges of multicast domains to isolate addresses between different multicast domains, so that the same multicast address can be used in different multicast domains without causing collisions.

The membership of a group is dynamic. Hosts can join or leave multicast groups at any time.

A multicast group is identified by a multicast address. There are two types of multicast addresses:

- Permanent group addresses: Multicast addresses reserved by IANA for routing protocols. Such an address identifies a group of specific network devices (also known as reserved multicast groups). For detail, see Table 26. A permanent group address will never change. There can be any number of, or even 0, members in a permanent multicast group.
- Temporary group addresses: Group addresses that are temporarily assigned for user multicast groups. Once the number of members of a group comes to 0, the address is released.

Table 26 Some reserved multicast addresses

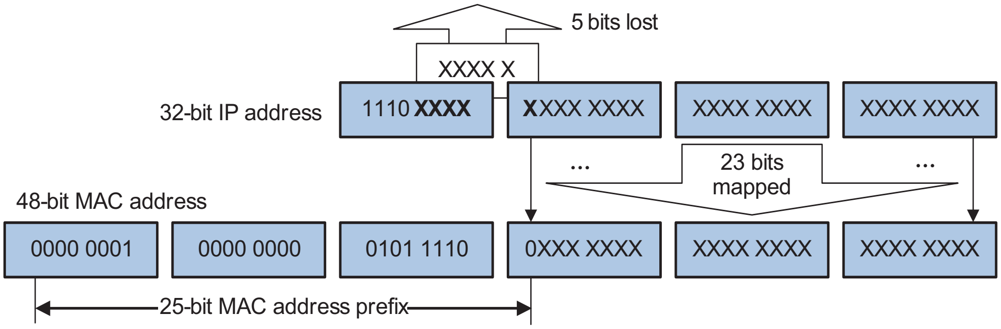
Address	Description
224.0.0.1	All systems on this subnet, including hosts and routers
224.0.0.2	All multicast routers on this subnet
224.0.0.3	Unassigned
224.0.0.4	Distance Vector Multicast Routing Protocol (DVMRP) routers
224.0.0.5	Open Shortest Path First (OSPF) routers
224.0.0.6	OSPF designated routers/backup 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)

Multicast MAC addresses

When a unicast IP packet is transmitted over an Ethernet network, the destination MAC address is the MAC address of the receiver. When a multicast packet is transmitted over an Ethernet network, however, a multicast MAC address is used as the destination address because the packet is directed to a group formed by a number of receivers, rather than to a specific receiver.

As defined by IANA, the high-order 24 bits of a multicast MAC address are 0x01005e, bit 25 is 0x0, and the low-order 23 bits are the low-order 23 bits of a multicast IP address. The IPv4-to-MAC mapping relation is shown in Figure 158.

Figure 158 IPv4-to-MAC address mapping

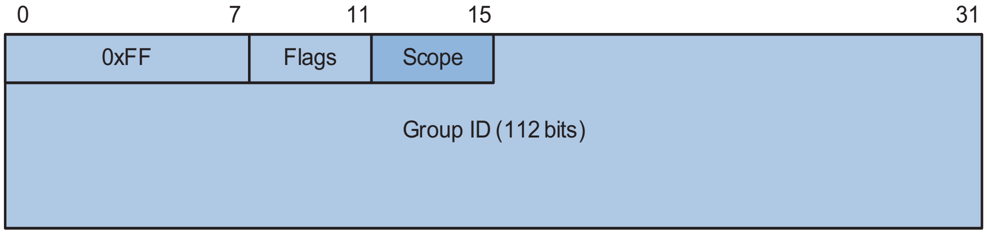


The high-order four bits of a multicast IPv4 address are 1110, indicating that this address is a multicast address, and only 23 bits of the remaining 28 bits are mapped to a MAC address, so five bits of the multicast IPv4 address are lost. As a result, 32 multicast IPv4 addresses map to the same MAC 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.

IPv6 Multicast Addresses

As defined in RFC 4291, the format of an IPv6 multicast is as follows:

Figure 159 IPv6 multicast format



- 0xFF: 8 bits, indicating that this address is an IPv6 multicast address.
- Flags: 4 bits, of which the high-order bits are reserved and set to 0; the lowest-order bit is the Transient (T) flag. When set to 0, the T flag indicates a permanently-assigned (well-known) multicast address assigned by IANA; when set to 1, the T flag indicates a transient, or dynamically assigned 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 27.
- Reserved: 80 bits, all set to 0 currently.
- Group ID: 112 bits, identifying the multicast group. For details about this field, refer to RFC 3306.

Table 27 Values of the Scope field

Value	Meaning
0, 3, F	Reserved
1	Node-local scope
2	Link-local scope
4	Admin-local scope

Table 27 Values of the Scope field

Value	Meaning
5	Site-local scope
6, 7, 9 through D	Unassigned
8	Organization-local scope
E	Global scope

Multicast Protocols

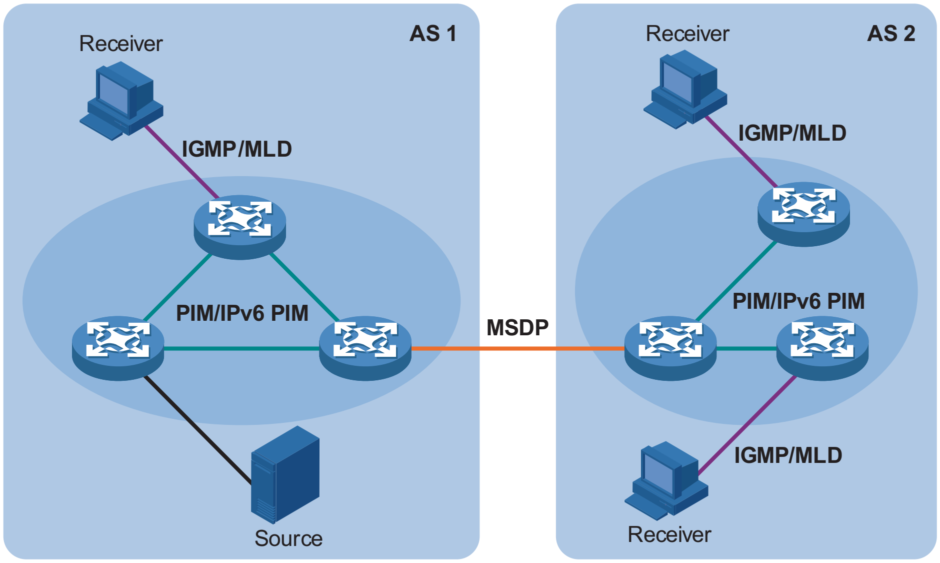
- *Generally, we refer to IP multicast working at the network layer as Layer 3 multicast and the corresponding multicast protocols as Layer 3 multicast protocols, which include IGMP/MLD, PIM/IPv6 PIM, and MSDP; we refer to IP multicast working at the data link layer as Layer 2 multicast and the corresponding multicast protocols as Layer 2 multicast protocols, which include IGMP Snooping/MLD Snooping, and multicast VLAN.*
- *IGMP Snooping, IGMP, PIM and MSDP are for IPv4, MLD Snooping, MLD, and IPv6 PIM are for IPv6. Multicast VLAN are for both IPv4 and 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 160 describes where these multicast protocols are in a network.

Figure 160 Positions of Layer 3 multicast protocols



1 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 multicast devices.

2 Multicast routing protocols

A multicast routing protocol runs between Layer 3 multicast devices to establish and maintain multicast routes and forward multicast packets correctly and efficiently. A multicast route is a loop-free data transmission path from a data source to multiple receivers. Namely, it is a multicast distribution tree.

In the ASM model, multicast routes come in intra-domain routes and inter-domain routes.

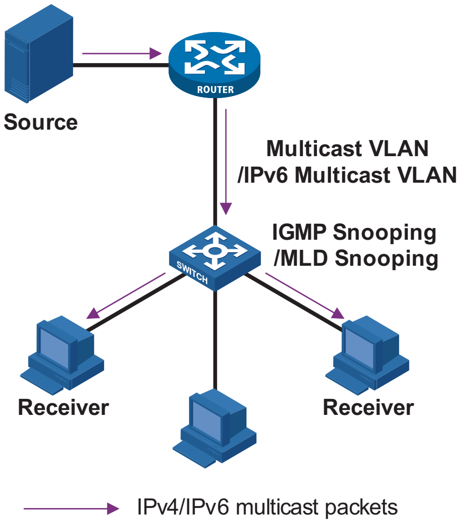
- An intra-domain multicast routing protocol is used to discover multicast sources and build multicast distribution trees with an autonomous system (AS) so as to deliver multicast data to receivers. Among a variety of mature intra-domain multicast routing protocols, protocol independent multicast (PIM) is a most popular one. It delivers information to receivers by discovering the multicast source and establishing multicast distribution trees. Based on the forwarding mechanism, PIM comes in two modes - dense mode (PIM-DM) and sparse mode (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).

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-SD are sufficient for multicast information transport.

Layer 2 multicast protocols

Layer 2 multicast protocols include IGMP Snooping/MLD Snooping and multicast VLAN. Figure 161 shows where these protocols are in the network.

Figure 161 Positions 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

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 feature enabled on the Layer 2 device, the Layer 3 multicast device needs to send only one copy of multicast to the multicast VLAN on the Layer 2 device. This avoids waste of network bandwidth and extra burden on the Layer 3 device.

Multicast Packets Forwarding Mechanism

In a multicast model, a multicast source sends information to a host group, which is identified by a multicast group address in the destination address field of the IP 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 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.

For details about RPF, refer to "RPF Mechanism" on page 503 or "RPF Mechanism" on page 515.

38 MULTICAST ROUTING AND FORWARDING CONFIGURATION

When configuring multicast routing and forwarding, go to the following sections for information you are interested in:

The term "router" in this document refers to a router in a generic sense or a Switch 8800 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 device supports multiple multicast protocols, its multicast routing table will include routes generated by multiple protocols. The device chooses the optimal route from the multicast routing table based on the configured multicast routing and forwarding policy and installs the route entry into its multicast forwarding table.

RPF Mechanism

When creating multicast routing table entries, a multicast routing protocol uses the reverse path forwarding (RPF) mechanism to ensure multicast data delivery along the correct path.

The RPF mechanism enables devices to correctly forward multicast packets based on the multicast route configuration. In addition, the RPF mechanism also helps avoid data loops caused by various reasons.

Implementation of the RPF mechanism

Upon receiving a multicast packet that a multicast source S sends to a multicast group G, the device first searches its multicast forwarding table:

1. If the corresponding (S, G) entry exists, and the interface on which the packet actually arrived is the incoming interface in the multicast forwarding table, the device forwards the packet to all the outgoing interfaces.
2. 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 result of the RPF check shows that the RPF interface is the incoming interface of the existing (S, G) entry, this means that the (S, G) entry is correct but the packet arrived from a wrong path and is to be discarded.
   - If the result of the RPF check shows that the RPF interface is not the incoming interface of the existing (S, G) entry, this means that the (S, G) entry is no longer valid. The device replaces the incoming interface of the (S, G) entry with the interface on which the packet actually arrived and forwards the packet to all the outgoing interfaces.
3. If no corresponding (S, G) entry exists in the multicast forwarding table, the packet is also subject to an RPF check. The device creates an (S, G) entry based on the relevant routing information and using the RPF interface as the incoming interface, and installs the entry into the multicast forwarding table.
   - If the interface on which the packet actually arrived is the RPF interface, the RPF check is successful and the device 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 device discards the packet.

RPF check

The basis for an RPF check is a unicast route or a multicast static route. A unicast routing table contains the shortest path to each destination address, while a multicast static routing table lists the RPF routing information defined by the user through static configuration. A multicast routing protocol does not independently maintain any type of unicast route; instead, it relies on the existing unicast routing information or multicast static routes in creating multicast routing entries.

When performing an RPF check, the device searches its unicast routing table and multicast static routing table at the same time. The specific process is as follows:

1. The device first chooses an optimal route from the unicast routing table and multicast static routing table:
   - The device 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 device 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 device 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 device selects one from these two optimal routes as the RPF route. The selection is as follows:

- If configured to use the longest match principle, the device selects the longest match route from the two; if these two routes have the same mask, the route selects the route with a higher priority; if the two routes have the same priority, the device selects the multicast static route.
- If not configured to use the longest match principle, the device selects the route with a higher priority; if the two routes have the same priority, the device selects the multicast static route.

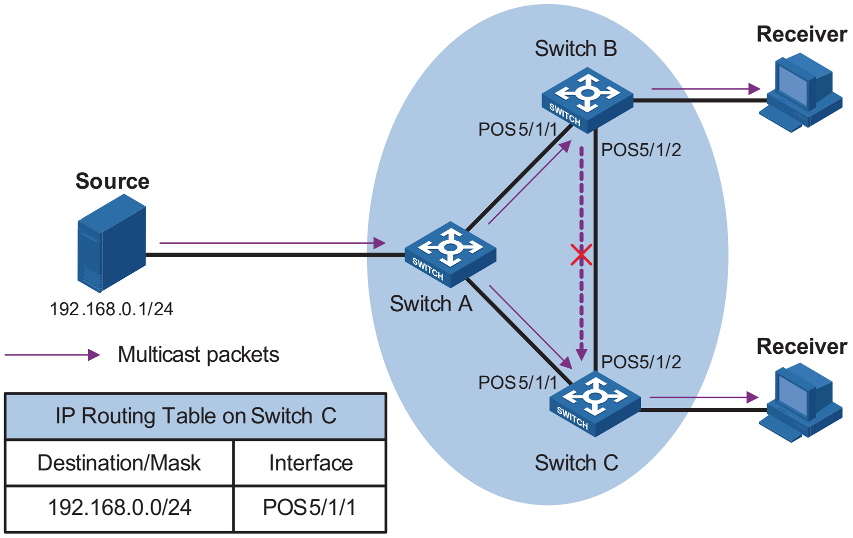
The above-mentioned "packet source" can mean different things in different situations:

- *For a packet traveling along the shortest path tree (SPT) from the multicast source to the receivers or the source-based tree from the multicast source to the rendezvous point (RP), "packet source" means the multicast source.*
- *For a packet traveling along the rendezvous point tree (RPT) from the RP to the receivers, "packet source" means the RP.*
- *For a bootstrap message from the bootstrap router (BSR), "packet source" means the BSR.*

For details about the concepts of SPT, RPT and BSR, refer to *“PIM Configuration” on page 563*.

Assume that unicast routes exist in the network and no multicast static routes have been configured on Switch C, as shown in Figure 162. Multicast packets travel along the SPT from the multicast source to the receivers.

Figure 162 RPF check process



IP Routing Table on Switch C	
Destination/Mask	Interface
192.168.0.0/24	POS5/1/1

- A multicast packet from Source arrives to POS5/1/2 of Switch C, and the corresponding forwarding entry does not exist in the multicast forwarding table of Switch C. Switch C performs an RPF check, and finds in its unicast routing table that the outgoing interface to 192.168.0.0/24 is POS5/1/1. This means that the interface on which the packet actually arrived is not the RPF interface. The RPF check fails and the packet is discarded.

- A multicast packet from Source arrives to POS5/1/1 of Switch C, and the corresponding forwarding entry does not exist in the multicast forwarding table of Switch C. The device performs an RPF check, and finds in its unicast routing table that the outgoing interface to 192.168.0.0/24 is the interface on which the packet actually arrived. The RPF check succeeds and the packet is forwarded.

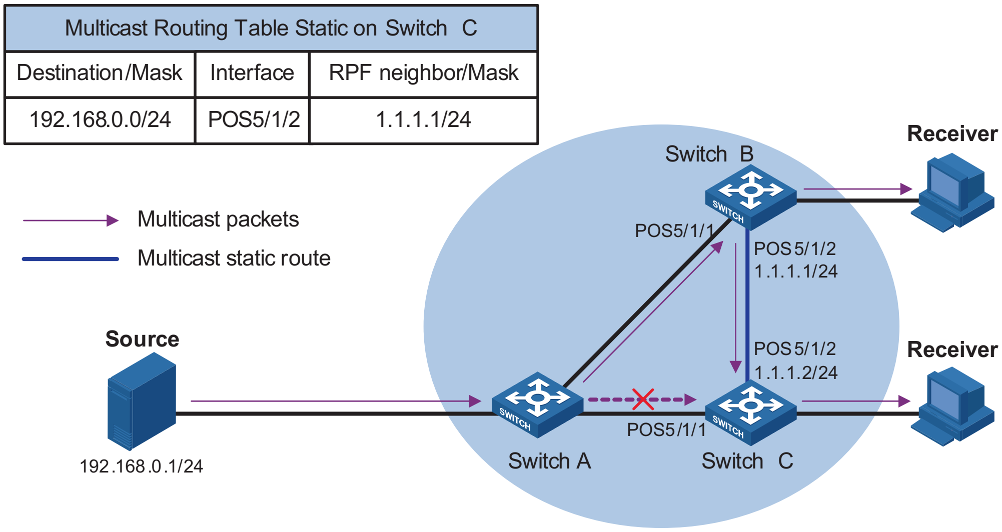
Multicast Static Route

If the topology structure of a multicast network is the same as that of a unicast network, receivers can receive multicast data via unicast routes. However, the topology structure of a multicast network may differ from that of a unicast network, and some devices may support only unicast but not multicast. In this case, you can configure multicast static routes to provide multicast transmission paths that are different from those for unicast traffic. Note the following two points:

- A multicast static route only affects RPF checks, rather than guides multicast forwarding, so it is also called an RPF static route.
- A multicast static route is effective on the multicast device on which it is configured, and will not be broadcast throughout the network or injected to other devices.

A multicast static route is an important basis for RPF checks. With a multicast static route configured on a device, the device searches the unicast routing table and the multicast static routing table simultaneously in a RPF check, chooses the optimal unicast RPF route and the optimal multicast static route respectively from the routing tables, and uses one of them as the RPF route after comparison.

Figure 163 Multicast static route



Multicast Routing Table Static on Switch C		
Destination/Mask	Interface	RPF neighbor/Mask
192.168.0.0/24	POS5/1/2	1.1.1.1/24

As shown in Figure 163, when no multicast static route is configured, Switch C's RPF neighbor on the path back to Source is Switch A and the multicast information from Source travels along the path from Switch A to Switch C, which is the unicast route between the two devices; with a static route configured on Switch C and Switch B as Switch C's RPF neighbor on the path back to Source, the multicast information from Source travels from Switch A to Switch B and then to Switch C.

Configuration Task List

Complete these tasks to configure multicast routing and forwarding:

Task	Remarks
"Enable IP Multicast Routing" on page 507"Enable IP Multicast Routing" on page 507	Required
"Configuring Multicast Static Routes" on page 508"Configuring Multicast Static Routes" on page 508	Required
"Configuring a Multicast Route Match Rule" on page 508"Configuring a Multicast Route Match Rule" on page 508	Optional
"Configuring Multicast Load Splitting" on page 509"Configuring Multicast Load Splitting" on page 509	Optional
"Configuring Multicast Forwarding Range" on page 509"Configuring Multicast Forwarding Range" on page 509	Optional
"Configuring Multicast Forwarding Table Size" on page 510"Configuring Multicast Forwarding Table Size" on page 510	Optional
"Displaying and Maintaining Multicast Routing and Forwarding" on page 510	Optional

Configuring Multicast Routing and Forwarding

Enable IP Multicast Routing

Before configuring any Layer 3 multicast functionality, you must enable IP multicast routing.

Follow these steps to enable IP multicast routing:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable IP multicast routing	**multicast routing-enable**	Required Disable by default

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 *"Assigning an IP Address to an Interface" on page 208*.

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 maximum number of downstream nodes for a single route in a multicast forwarding table
- The maximum number of routing entries in a multicast forwarding table
- The multicast forwarding range

Configuring Multicast Static Routes

Based on the application environment, a multicast static route has the following two functions:

- Changing an RPF route. If the multicast topology structure is the same as the unicast topology in a network, the delivery path of multicast traffic is the same as in unicast. By configuring a multicast static route, you can change the RPF route so as to create a transmission path that is different from the unicast traffic transmission path.
- Creating an RPF route. When a unicast route is interrupted, multicast traffic forwarding is stopped due to lack of an RPF route. By configuring a multicast static route, you can create an RPF route so that a multicast routing entry is created to guide multicast traffic forwarding.

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 is configured by default.

CAUTION: *When configuring a multicast static route, you cannot designate an RPF neighbor by specifying an interface (by means of the interface-type interface-number command argument combination) if the interface type of that device is VLAN-interface; instead, you can designate an RPF neighbor only by specifying an address (rpf-nbr-address).*

Configuring a Multicast Route Match Rule

In RPF route selection, the device chooses an optimal route from the multicast static routing table and the unicast routing table respectively, and then selects the superior route from these two routes. RPF route selection falls in two cases:

- If the device is configured to use longest match for route selection, then:

1 The device selects the longest match route from the two routes;

2 If these routes have the same mask, the route with a higher priority will be selected;

3 If these routes have the same priority, the multicast static route will be selected.

- If the device is not configured to use longest match for route selection, then:

1. The route with a higher priority will be selected;
2. If these routes have the same priority, the multicast static route will be selected.

Follow these steps to configure a multicast route match rule:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the device to select a route based on the longest match	**multicast longest-match**	Optional In order of routing table entries by default

Configuring Multicast Load Splitting

With the load splitting feature enabled, multicast traffic will be evenly distributed among different routes.

Follow these steps to configure multicast load splitting:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configuring multicast load splitting	**multicast load-splitting** { **source** | **source-group** }	Required Disabled by default

Configuring 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.

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 methods 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*	-
Configure a multicast forwarding boundary	**multicast boundary** *group-address* { *mask* | *mask-length* }	Required No forwarding boundary by default

Currently, Switch 8800s support multicast forwarding boundary configuration in VLAN interface view and POS interface view.

Configuring Multicast Forwarding Table Size

Too many multicast routing entries can exhaust the device's memory and thus result in lower device performance. Therefore, the number of multicast routing entries should be limited. You can set a limit on the number of entries in the multicast routing table based on the actual networking situation and the performance requirements. In any case, the number of route entries must not exceed the maximum number allowed by the system. This maximum value varies with different device models.

If the configured maximum number of downstream nodes (namely, the maximum number of outgoing interfaces) for a routing entry in the multicast forwarding table is smaller than the current number, the downstream nodes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the multicast routing protocol. In addition, newly added downstream nodes cannot be installed to the routing entry into the forwarding table.

If the configured maximum number of routing entries in the multicast forwarding table is smaller than the current number, the routes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the multicast routing protocol. In addition, newly added route entries cannot be installed to the forwarding table.

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 downstream nodes for a single route in the multicast forwarding table	**multicast forwarding-table downstream-limit** *limit*	Optional 128 by default
Configure the maximum number of routing entries in the multicast forwarding table	**multicast forwarding-table route-limit** *limit*	Optional 512 by default

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** [*group-address* [**mask** { *mask* | *mask-length* }] | *source-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **include** | **exclude** | **match** } { *interface-type interface-number* | **register** } } | **statistics** | **slot** *slot-id*] * [**port-info**]	Available in any view
		Available in any view

To do...	Use the command...	Remarks
View the multicast routing table information	**display multicast routing-table** [*group-address* [**mask** { *mask* | *mask-length* }] | *source-address* [**mask** { *mask* | *mask-length* }] | **incoming-interface** { *interface-type interface-number* | **register** } | **outgoing-interface** { { **include** | **exclude** | **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** { { *group-address* [**mask** { *mask* | *mask-length* }] | *source-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** { { *group-address* [**mask** { *mask* | *mask-length* }] | *source-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

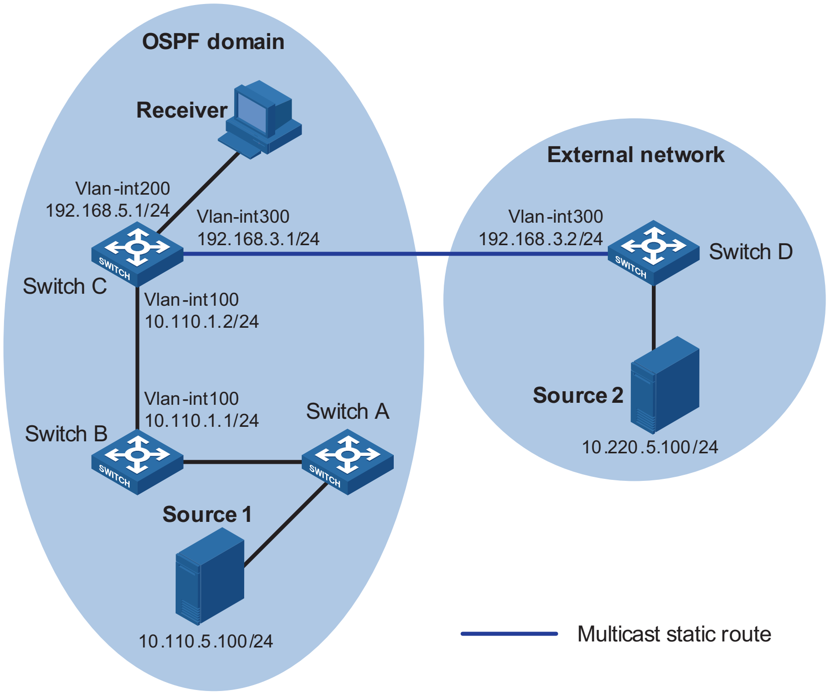
Multicast Static Route Configuration

Network requirements

- All switches in the network support IP multicast.
- Switch A, Switch B and Switch C run OSPF, and have no unicast routes to Switch D.
- Receiver can receive the multicast data from Source 1 through the path Switch A - Switch B - Switch C.
- Perform the following configuration so that Receiver can receive multicast data from Source 2, which is out of the OSPF domain, through Switch C.

Network diagram

Figure 164 Network diagram for multicast route configuration



Configuration procedure

Only the commands related to multicast static route configuration are listed in the configuration procedure.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the IP address and subnet mask for each interface as per Figure 164. The detailed configuration steps are omitted here.

Enable OSPF on Switch A, Switch B and Switch C. Ensure the network-layer interoperation among the switches. 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 on each interface

# Enable IP multicast routing on Switch C, and enable PIM-DM on each interface.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] pim dm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] pim dm
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] pim dm
[SwitchC-Vlan-interface300] 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 a multicast static route

# Configure a multicast static route on Switch C, specifying Switch D as its RPF neighbor on the route to Source 2.

```
[SwitchC] ip rpf-route-static 10.220.5.0 255.255.255.0 192.168.3.2
```

4 Verify the configuration

# Before the above-mentioned multicast static route is configured, Receiver can multicast data from Source 1, but cannot receive multicast data from Source 2. Use the **display multicast rpf-info** command to view the RPF routes to Source 1 and Source 2 respectively on Switch C.

```
[SwitchC] display multicast rpf-info 10.110.5.100
RPF information about source 10.110.5.100:
     RPF interface: Vlan-interface100, RPF neighbor: 10.110.1.1
     Referenced route/mask: 10.110.5.0/24
     Referenced route type: igp
     Route selection rule: preference-preferred
     Load splitting rule: disable
[SwitchC] display multicast rpf-info 10.220.5.100
```

As shown above, Switch C does not have an RPF route to Source 2.

# After the multicast static route is configured, use the **display multicast rpf-info** command to view the RPF route to Source 2 again on Switch C.

```
[SwitchC] display multicast rpf-info 10.220.5.100
RPF information about source 10.220.5.100:
     RPF interface: Vlan-interface300, RPF neighbor: 192.168.3.2
     Referenced route/mask: 10.220.5.0/24
     Referenced route type: multicast static
     Route selection rule: preference-preferred
     Load splitting rule: disable
```

As shown above, an RPF route to Source 2 has been established on Switch C.

Troubleshooting Multicast Routing and Forwarding

Multicast Static Route Failure

Symptom

No dynamic routing protocol is enabled on the devices, 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 does not exist in the multicast route configuration table and multicast routing table.
- If the optimal route if 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 a VLAN interface, you can specify an RPF neighbor only by providing its IP address (*rpf-nbr-address*) rather than an interface type and interface number (*interface-type interface-number*).
4. Check that the multicast static route matches the specified routing protocol. If a protocol was specified when the multicast static route was configured, 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 intermediate devices but fails to reach the last hop device.

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 device. If so, the device has received the multicast data; otherwise, the device 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.

39 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" on page 515
- "Configuring IPv6 Multicast Routing and Forwarding" on page 518
- "Displaying and Maintaining IPv6 Multicast Routing and Forwarding" on page 520
- "Troubleshooting IPv6 Multicast Policy Configuration" on page 521

- *The term "router" in this document refers to a router in a generic sense or a Switch 8800 running an IPv6 multicast routing protocol.*
- *Currently, POS interfaces of Switch 8800s do not support IPv6 multicast. The commands to be executed in interface view are not executable in POS interface view.*

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. The IPv6 multicast forwarding table is consistent with the IPv6 routing table.

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 Mechanism

When creating IPv6 multicast routing table entries, an IPv6 multicast routing protocol uses the reverse path forwarding (RPF) to ensure IPv6 multicast data delivery along the correct path.

The RPF mechanism enables routers to correctly forward IPv6 multicast packets based on the multicast route configuration. In addition, the RPF mechanism also helps avoid data loops caused by various reasons.

Implementation of the RPF mechanism

Upon receiving an IPv6 multicast packet sent from a multicast source S to an IPv6 multicast group G, the device first searches its IPv6 multicast forwarding table:

1. If the corresponding (S, G) entry exists, and the interface on which the packet actually arrived is the incoming interface in the IPv6 multicast forwarding table, the device forwards the packet to all the outgoing interfaces.
2. 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 packet is subject to an RPF check.
   - If the result of the RPF check shows that the RPF interface is the incoming interface of the existing (S, G) entry, this means that the (S, G) entry is correct but the packet arrived from a wrong path and is to be discarded.
   - If the result of the RPF check shows that the RPF interface is not the incoming interface of the existing (S, G) entry, this means that the (S, G) entry is no longer valid. The device replaces the incoming interface of the (S, G) entry with the interface on which the packet actually arrived and forwards the packet to all the outgoing interfaces.
3. If no corresponding (S, G) entry exists in the multicast forwarding table, the packet is also subject to an RPF check. The device creates an (S, G) entry based on the relevant routing information and using the RPF interface as the incoming interface, and installs the entry into the IPv6 multicast forwarding table.
   - If the interface on which the packet actually arrived is the RPF interface, the RPF check is successful and the device 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 device discards the packet.

RPF Check

The basis for an RPF check is an IPv6 unicast route. The IPv6 unicast routing table contains the shortest path to each destination subnet. A multicast routing protocol does not independently maintain any type of unicast routes; instead, it relies on the existing unicast routing information in creating multicast routing entries.

When performing an RPF check, the device searches its IPv6 unicast routing table using the IP 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 device 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.

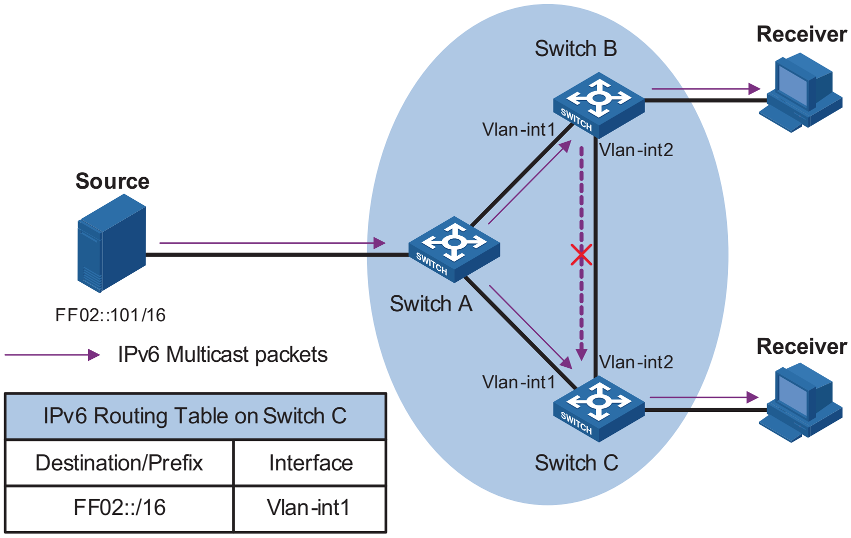
The above-mentioned "packet source" can mean different things in different situations:

- *For a packet traveling along the shortest path tree (SPT) from the multicast source to the receivers or the source-based tree from the multicast source to the rendezvous point (RP), "packet source" means the multicast source.*
- *For a packet traveling along the rendezvous point tree (RPT) from the RP to the receivers, "packet source" means the RP.*
- *For a bootstrap message from the bootstrap device (BSR), "packet source" means the BSR.*

For details about the concepts of SPT, RPT, RP and BSR, refer to *"PIM Configuration" on page 563*.

Assume that IPv6 unicast routes exist in the network, and IPv6 multicast packets travel along the SPT from the multicast source to the receivers, as shown in Figure 165.

Figure 165 RPF check process



IPv6 Routing Table on Switch C	
Destination/Prefix	Interface
FF02::/16	Vlan-int1

- An IPv6 multicast packet from Source arrives on VLAN-interface 2 of Switch C, and the IPv6 multicast forwarding table does not contain the corresponding forwarding entry. Switch C performs an RPF check, and finds in its IPv6 unicast routing table that the outgoing interface to the network subnet FF02::/16 is VLAN-interface 1. This means that the interface on which the packet actually arrived is not the RPF interface. The packet fails the FPR check and is discarded.
- An IPv6 multicast packet from Source arrives on VLAN-interface 1 of Switch C, and the IPv6 multicast forwarding table does not contain the corresponding forwarding entry. Switch C performs an RPF check, and finds in its IPv6 unicast routing table that the outgoing interface to the subnet FF02::/16 is the one on which the IPv6 multicast packet actually arrived. The RPF check succeeds and the packet is forwarded.

Configuration Task List

Complete these tasks to configure IPv6 multicast routing and forwarding:

Task	Remarks
"Enabling IPv6 Multicast Routing" on page 518	Required
"Configuring IPv6 Multicast Load Splitting" on page 518	Optional

Task	Remarks
"Configuring an IPv6 Multicast Forwarding Range" on page 518	Optional
"Configuring the IPv6 Multicast Forwarding Table Size" on page 519	Optional

Configuring IPv6 Multicast Routing and Forwarding

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

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 route in the IPv6 multicast forwarding table
- Maximum number of routing entries in the IPv6 multicast forwarding table

Configuring IPv6 Multicast Load Splitting

With the IPv6 multicast load splitting feature enabled, IPv6 multicast traffic will be evenly distributed among different routes.

Follow these steps to enable IPv6 multicast load splitting:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable IPv6 multicast load splitting	**multicast ipv6 load-splitting** {**source** | **source-group** }	Required 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.

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 VLAN 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

Too many IPv6 multicast routing entries can exhaust the device's memory and thus result in lower device performance. Therefore, the number of IPv6 multicast routing entries should be limited. You can set a limit on the number of entries in the IPv6 multicast routing table based on the actual networking situation and the performance requirements. In any case, the number of route entries must not exceed the maximum number allowed by the system.

If the configured maximum number of downstream nodes (namely the maximum number of outgoing interfaces) for a routing entry in the IPv6 multicast forwarding table is smaller than the current number, the downstream nodes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the IPv6 multicast routing protocol. In addition, newly added downstream nodes cannot be installed to the routing entry in the forwarding table.

If the configured maximum number of routing entries in the IPv6 multicast forwarding table is smaller than the current number, the routes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the IPv6 multicast routing protocol. In addition, newly added routing entries cannot be installed to the forwarding table.

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 downstream nodes for a single route in the IPv6 multicast forwarding table	**multicast ipv6 forwarding-table downstream-limit** *limit*	Optional 128 by default

To do...	Use the command...	Remarks
Configure the maximum number of routing entries in the IPv6 multicast forwarding table	**multicast ipv6 forwarding-table route-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-id*] * [**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
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

To do...	Use the command...	Remarks
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

Multicast Data Fails to Reach Receivers

Symptom

The multicast data can reach some devices but fails to reach the last-hop device.

Analysis

If a multicast forwarding boundary is configured on a VLAN interface through the **multicast ipv6 boundary** command, any multicast packet will be kept from crossing the boundary.

Solution

1. Use the **display pim ipv6 routing-table** command to check whether the corresponding (S, G) entries exist on the device. If so, the device has received the multicast data; otherwise, the device has not received the data..
2. Use the **display multicast ipv6 boundary** command to view the multicast boundary information on the interfaces. .Use the **multicast ipv6 boundary** command to change the multicast forwarding boundary setting.
3. In the case of IPv6 PIM-SM, check the BSR and RP information..

40 IGMP CONFIGURATION

When configuring IGMP, go to the following sections for the information you are interested in:

The term "router" in this document refers to a router in a generic sense or a Switch 8800 running IGMP.

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, IGMPv3 provides strong support to the Source-Specific Multicast (SSM) model.

Work Mechanism of IGMPv1

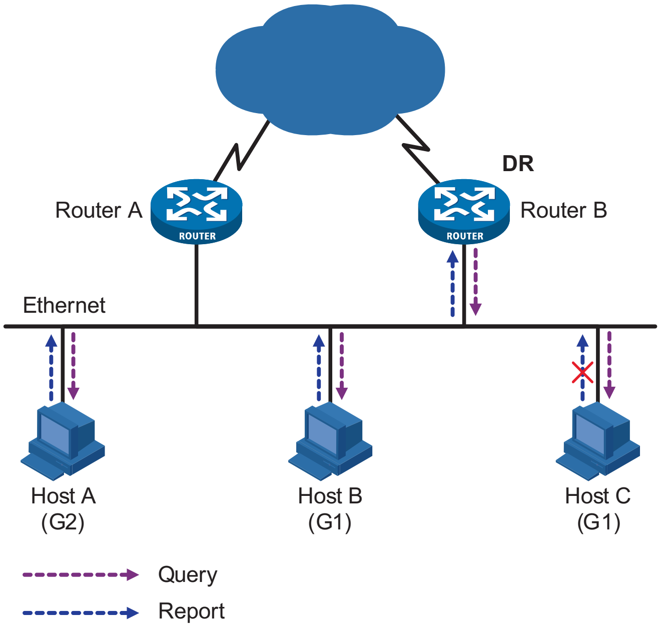
IGMPv1 manages multicast group memberships mainly based on the query and response mechanism.

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 a multicast routing protocol (such as PIM) serves as the IGMP querier, which sends periodical IGMP queries.

For more information about DR, refer to *"PIM Configuration" on page 563*.

Figure 166 Joining multicast groups



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 166. The basic process that the hosts join the multicast groups is as follows:

1. The IGMP querier (Router B in the figure) periodically multicasts IGMP queries (with the destination address of 224.0.0.1) to all hosts and routers on the local subnet.
2. 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 interest in G1. Assume it is Host B that sends the report message.
3. Host C, which is on the same subnet, hears the report from Host B for joining G1. Upon hearing the report, Host C will suppress itself from sending a report message for the same multicast group, 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 helps reduce traffic over 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 generate (*, 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 mechanism, upon leaving a multicast group, an IGMPv1 host stops sending reports with the destination address being the address of that multicast group. If no member of a multicast group exists on the subnet, the IGMP routers will not receive any report addressed to that multicast group, so the routers will delete the multicast forwarding entries corresponding to that multicast group after a period of time.

Enhancements Provided by IGMPv2

Compared with IGMPv1, IGMPv2 provides the querier election mechanism and 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 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

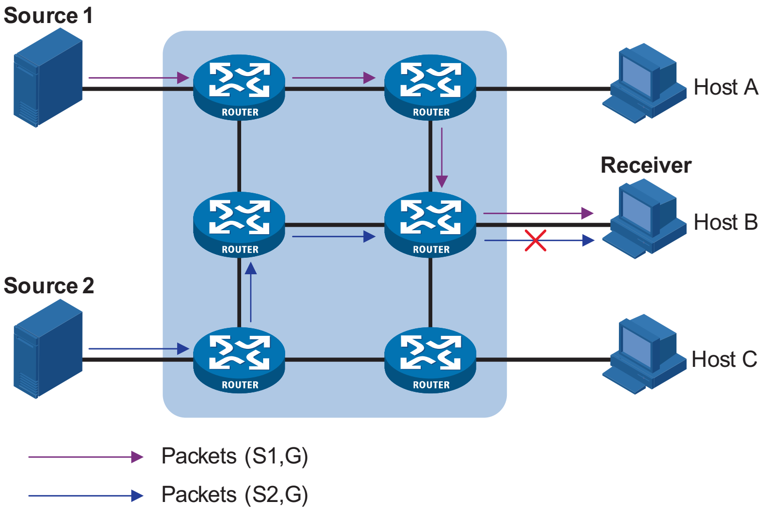
Enhancements in control capability of hosts

In addition to group-specific queries, IGMPv3 has introduced source filtering modes (Include and Exclude), so that a host not only can join a designated multicast group but also can specify to receive or reject multicast data from a designated multicast source. When a host joins a multicast group:

- If it needs to receive multicast data from specific sources like S1, S2, ..., it sends a report with the Filter-Mode denoted as "Include Sources (S1, S2, ...).
- If it needs to reject 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 167, 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 167 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, rather than the multicast data Source 2 sends to multicast group G. Thus, only multicast data from Source 1 will be delivered to Host B.

CAUTION: *Currently, Switch 8800s do not support the Exclude filtering mode.*

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 an uncertain number of source addresses.

Group record types include:

- Current-state record: Current-state record: The current-state record reports the current reception state of the interface on which the report is sent. The state can be either Include (the interface has a filter mode of Include for the specified multicast address list) or Exclude (the interface has a filter mode of Exclude for the specified multicast address list).
- Filter-mode-change record: A filter-mode-change record indicates that the interface filter mode has changed from Include to Exclude or from Include to Exclude for the specified multicast address list.
- Source-list-change record: A source-list-change record indicates that new source addresses are allowed or old source addresses are blocked.

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

Configuring IGMP

Complete these tasks to configure IGMP:

Task		Description
“IGMP Configuration” on page 523	“Enabling IGMP” on page 528	Required
	“Configuring IGMP Versions” on page 528	Optional
	“Configuring a Static Member of a Multicast Group” on page 529	Optional
	“Configuring a Multicast Group Filter” on page 529	Optional
“Configuring IGMP Performance Parameters” on page 530	“Configuring IGMP Message Options” on page 530	Optional
	“Configuring IGMP Timers” on page 531	Optional
	“Configuring IGMP Fast Leave” on page 533	Optional

- *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

Enabling IGMP

First, IGMP must be enabled on the interface on which the multicast group memberships are to be established and maintained.

Follow these steps to enable IGMP:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enable IP multicast routing	**multicast routing-enable**	Required Disabled by default
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
Enable IGMP	**igmp enable**	Required Disabled by default

CAUTION:

- *Hosts can join multicast groups only if IGMP is enabled on the receiver-side DR. For more information about a DR, refer to "PIM Configuration" on page 563.*
- *After IGMP is enabled on a VLAN interface, IGMP snooping cannot be enabled in the VLAN corresponding to the VLAN interface, and vice versa.*

Configuring IGMP Versions

Because messages vary with different IGMP versions, the same IGMP version should be configured for all devices 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...	Description
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 for an interface

Follow these steps to configure an IGMP version on an interface:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
Configure an IGMP version on the interface	**igmp version** *version-number*	Optional IGMPv2 by default

CAUTION: *All devices on the same subnet must run the same version of IGMP, and cannot automatically switch between different IGMP versions.*

Configuring a Static Member of a Multicast Group

After being configured as a static member of a multicast group, an interface 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.

Before you can configure an interface of a PIM-SM device as a static member of a multicast group, if the interface is PIM-SM enabled, it must be a PIM-SM DR; if this interface is IGMP enable but not PIM-SM enabled, it must an IGMP querier. For more information about PIM-SM and a DR, refer to "PIM Configuration" on page 563.

Follow these steps to configure an interface as a statically connected member of a multicast group:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter POS interface view	**interface** *interface-type interface-number*	-
Configure the interface as a static member of a multicast group	**igmp static-group** *group-address* [**source** *source-address*]	Required Not a static member of any multicast group by default

This command does not trigger any IGMP message, and a device that has statically joined a multicast group is not a real member of that 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 that the interface serves.

Follow these steps to configure a multicast group filter:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter POS interface view	**interface** *interface-type interface-number*	-
Configuring a Multicast Group Filter	**igmp group-policy** *acl-number* [*version-number*]	Required No multicast group filter configured by default

When you use an advanced ACL as a filter, the source address in the ACL rule is the address of the multicast source specified in the IGMPv3 reports, rather than the source address in the IP packets.

Configuring IGMP Performance Parameters

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 configuring IGMP performance parameters, 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 configuring IGMP performance parameters, prepare the following data:

- IGMP general query interval
- Maximum response time for IGMP general queries
- Other querier present interval
- IGMP last-member query interval and count

Configuring IGMP Message Options

As IGMPv2 and IGMPv3 involve group-specific and group-and-source-specific queries, and multicast groups change dynamically, a device cannot join all multicast groups. Therefore, 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. Depending on whether an IGMP message carries the Router-Alert option in the IP header, the device processes the message differently. For details about Router-Alert, refer to RFC 2113.

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...	Description
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
Enable the 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 for an interface

Follow these steps to configure IGMP packet options for an interface:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
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
Enable the 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 Timers

The IGMP querier periodically sends IGMP general queries to decide whether any multicast group member exists on the network. The network administrator 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 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 before the "other querier present interval" timer expires, 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 interval" timer.

Configuring IGMP timers globally

Follow these steps to configure IGMP timers globally:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter IGMP view	**igmp**	-
Configure IGMP general query interval	**timer query** *interval*	Optional 60 seconds 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 IGMP last-member query count	**robust-count** *robust-value*	Optional 2 times by default
Configure the other querier present interval	**timer other-querier-present** *interval*	Optional For the system default, see "Note" below

Configuring IGMP timers for an interface

Follow these steps to configure IGMP timers for an interface:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
Configure IGMP general query interval	**igmp timer query** *interval*	Optional 60 seconds 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 lastmember-queryinterval** *interval*	Optional 1 second by default
Configure the IGMP last-member query count	**igmp robust-count** *robust-value*	Optional 2 times by default

To do...	Use the command...	Description
Configure the other querier present interval	**igmp timer other-querier-present** *interval*	Optional For the system default, see "Note" below

- *If not statically configured, the other querier present interval is [IGMP general query interval] times [IGMP robustness variable] plus [maximum response time for IGMP general queries] divided by two. By default, the values of these three parameters are 60 (seconds), 2 and 10 (seconds) respectively, so the default value of the other querier present interval = 60 × 2 + 10 / 2 = 125 (seconds).*
- *If statically configured, the other querier present interval takes the configured value.*

CAUTION:

- *If the statically configured other querier present interval is shorter than the IGMP general query interval, the querier election may be triggered frequently.*
- *In configuration, make sure that the maximum response time for IGMP general queries is less than the IGMP general query interval; otherwise, multicast group members may be wrongly removed.*
- *The configurations of the maximum response time for IGMP general queries, IGMP last-member query interval and other querier present interval are effective only when the IGMP querier runs IGMPv2 or IGMPv3.*

Configuring IGMP Fast Leave

To enable fast response to leave messages of hosts, you can enable the IGMP fast leave feature.

With the fast leave function enabled, after an IGMP querier receives a leave message from a host, it will not send IGMP group-specific queries to the group being left; instead, it will send a leave notification to the upstream. As a result, the leave latency is reduced on one hand, and the network bandwidth is saved on the other.

Configuring IGMP fast leave globally

Follow these steps to enable IGMP fast leave globally:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter IGMP view	**igmp**	-
Enable IGMP fast leave	**fast-leave** [**group-policy** *acl-number*]	Required Disabled by default

Configured in IGMP view, the fast leave feature takes effect only on POS interfaces rather than VLAN interfaces.

Configuring IGMP fast leave for an interface

Follow these steps to configure IGMP fast leave for an interface:

To do...	Use the command...	Description
Enter system view	**system-view**	-
Enter POS interface view	**interface** *interface-type interface-number*	-
Enable IGMP fast leave	**igmp fast-leave** [**group-policy** *acl-number*]	Required Disabled by default

The IGMP fast leave feature is effective only if the device is running IGMPv2 or IGMPv3.

Displaying and Maintaining IGMP

To do...	Use the command...	
View IGMP multicast group information	**display igmp group** [*group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]	Available in any view
View IGMP Ethernet port information	**display igmp group port-info** [**vlan** *vlan-id*] [**slot** *slot-number*] [**verbose**]	Available in any view
View IGMP configuration and running information on the interface	**display igmp interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View routing information in the IGMP routing table	**display igmp routing-table** [*group-address* [**mask** { *mask* | *mask-length* }] | *source-address* [**mask** { *mask* | *mask-length* }]] *	
Clear IGMP forwarding entries	**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

CAUTION:

- *The **reset igmp group** command may cause an interruption of receivers' reception of multicast data.*
- *The **reset igmp group** command cannot clear the IGMP forwarding entries of static joins.*

IGMP Configuration Examples

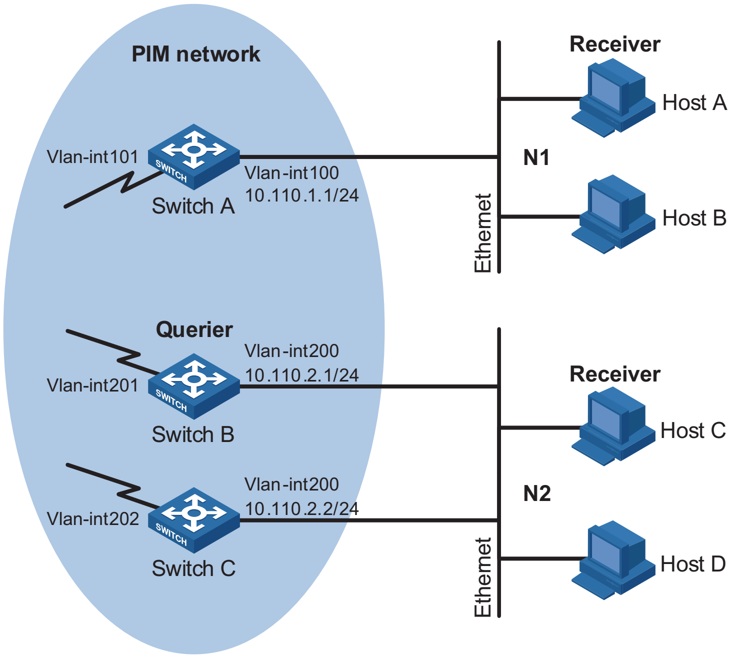
Network requirements

- Receivers receive VOD information through the multicast mode. 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-DM network connects to N1, and both Switch B and Switch C connect to N2.

- Switch A connects to N1 through Vlan-interface100, and to other devices in the PIM-DM network through Vlan-interface101.
- Switch B and Switch C connect to N2 through their respective Vlan-interface200, and to other devices in the PIM-DM network through Vlan-interface201 and Vlan-interface202 respectively.
- IGMPv2 is required between Switch A and N1. IGMPv2 is required between the other two switches and N2, with Switch B as the IGMP querier.

Network diagram

Figure 168 Network diagram for IGMP configuration



Configuration procedure

In the configuration procedure, only the commands related to IGMP configurations are listed.

1 Configure the IP addresses of the switch interfaces and configure a unicast routing protocol.

Configure the IP address and subnet mask of each interface as per Figure 168. The detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation among the switches. Ensure the network-layer interoperation among Switch A, Switch B and Switch C on the PIM-DM 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 IGMP on the host-side interfaces.

# Enable IP multicast routing on Switch A, and enable IGMP (version 2) and PIM-DM on Vlan-interface100.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] igmp version 2
[SwitchA-Vlan-interface100] pim dm
[SwitchA-Vlan-interface100] quit
```

# Enable IP multicast routing on Switch B, and enable IGMP (version 2) and PIM-DM on Vlan-interface200.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] igmp enable
[SwitchB-Vlan-interface200] igmp version 2
[SwitchB-Vlan-interface200] pim dm
[SwitchB-Vlan-interface200] quit
```

# Enable IP multicast routing on Switch C, and enable IGMP (version 2) and PIM-DM on Vlan-interface200.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] igmp version 2
[SwitchC-Vlan-interface200] pim dm
[SwitchC-Vlan-interface200] quit
```

3 Verify the configuration

Carry out the **display igmp interface** command to view the IGMP configuration and running status on each switch interface. For example:

# View IGMP information on Vlan-interface200 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
```

Troubleshooting IGMP

No Member Information on the Receiver-Side Device

Symptom

When a host sends a report for joining multicast group G, there is no member information of the multicast group G on the device closest to that host.

Analysis

- The correctness of networking and interface connections directly affects the generation of group member information.
- Multicast routing must be enabled on the device.
- If the **igmp group-policy** command has been configured on the POS interface, the POS interface cannot receive report messages that fail to pass filtering.

Solution

1. Check that the networking is correct and interface connections are correct.
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.
3. Check that the interface is in normal state and the correct IP address has been configured. Carry out the **display igmp interface** command to view the interface information. If no interface information is output, this means the interface is abnormal. Typically this is because the **shutdown** command has been executed on the interface, or the interface connection is incorrect, or no correct IP address has been configured on the interface.
4. Check that no ACL rule has been configured on the POS interface 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 Devices on the Same Subnet

Symptom

Different memberships are maintained on different IGMP devices on the same subnet.

Analysis

- A device running IGMP maintains multiple parameters for each interface, and these parameters influence one another, forming very complicated relationships. Inconsistent IGMP interface parameter configurations for devices on the same subnet will surely result in inconsistency of memberships.
- In addition, although IGMP routers are compatible with hosts, all devices on the same subnet must run the same version of IGMP. Inconsistent IGMP versions running on devices 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 devices 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 the devices are running the same version of IGMP.

41 IGMP SNOOPING CONFIGURATION

When configuration IGMP Snooping, go to the following sections for information you are interested in:

For details about IGMP and PIM, refer to "IGMP Configuration" on page 523 and "PIM Configuration" on page 563.

IGMP Snooping Overview

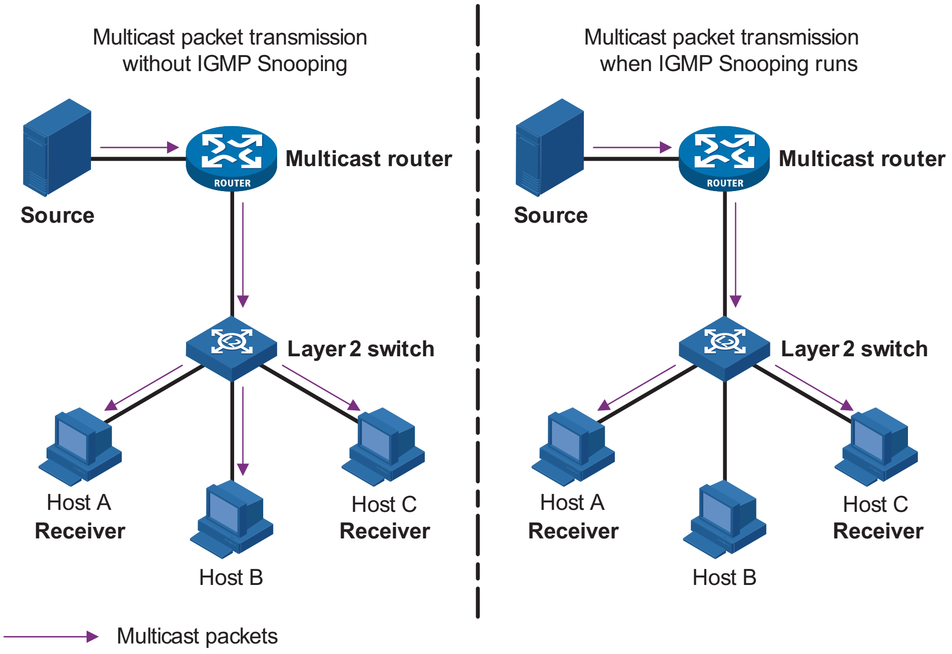
Internet Group Management Protocol Snooping (IGMP Snooping) is a multicast constraining mechanism that runs on Layer 2 devices to manage and control multicast groups.

Principle of IGMP Snooping

By analyzing received IGMP messages, a switch (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 169, 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 169 Before and after IGMP Snooping is enabled on Layer 2 device

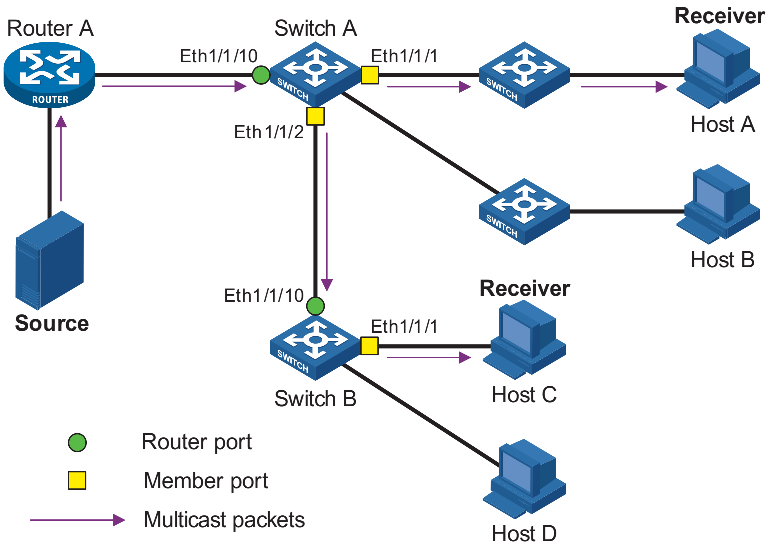


Basic Concepts in IGMP Snooping

IGMP Snooping related ports

As shown in Figure 170, 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 170 IGMP Snooping related ports



Ports involved in IGMP Snooping, as shown in Figure 170, are described as follows:

- Router port: A router port is a port on the Layer-3 multicast device or the IGMP querier side of the Ethernet switch. In the figure, Ethernet 1/1/10 of Switch A and Ethernet 1/0 of Switch B are router ports. A switch registers all its local router ports in its router port list.

- Member port: Also known as a listener port, a member port is a port on the multicast group member side of the Ethernet switch. In the figure, Ethernet 1/1/1 and Ethernet 1/1/2 of Switch A and Ethernet1/1/1 of Switch B are member ports. The switch records all member ports on the local device in the IGMP Snooping forwarding table.

- *Whenever mentioned in this document, a router port is a port on a switch that leads the switch to a Layer 3 multicast device, rather than a port on a router.*
- *An IGMP-snooping-enabled switch deems all its ports on which IGMP general queries with the source address other than 0.0.0.0 or PIM hello messages are received to be router ports. For details about PIM hello messages, see "Configuring PIM Hello Options" on page 589.*

Port aging timers in IGMP Snooping and related messages and actions

Table 28 Port aging timers in IGMP Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Router port aging timer	For each router port, the switch sets a timer initialized to the aging time of the route port	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
Member port aging timer	When a port joins an multicast group, the switch sets a timer for the port, which is initialized to the member port aging time	IGMP report	The switch removes this port from the multicast group forwarding table

Work Mechanism of IGMP Snooping

A switch running IGMP Snooping performs different actions when it receives different IGMP messages, as follows:

When receiving a general query

The IGMP querier periodically sends IGMP general queries to all hosts and routers 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 router port existing in its router port list, the switch resets the aging timer of this router port.
- If the receiving port is not a router port existing in its router port list, the switch adds it into its router port list and sets an aging timer for this router port.

When receiving a membership report

A host sends an IGMP report to the multicast router 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 multicast router 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 multicast group the host is interested in, and performs the following to the receiving port:

- If the port is already in the forwarding table, the switch resets the member port aging timer of the port.
- If the port is not in the forwarding table, the switch installs an entry for this port in the forwarding table and starts the member port aging timer of this port.

A switch will not forward an IGMP report through a non-router port.

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 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 to announce that it has left the multicast group.

Upon receiving an IGMP leave message on a member port, a switch forwards it to all router ports in the VLAN. Because the switch does not know whether any other member hosts of that multicast group still exists under the port to which the IGMP leave message arrived, the switch does not immediately delete the forwarding entry corresponding to that port from the forwarding table; instead, it resets the aging timer of the member port.

Upon receiving the IGMP leave message from a host, the IGMP querier resolves from the message the address of the multicast group that the host just left 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, a switch (non-querier) forwards it through all the router ports in the VLAN and all member ports of that multicast group, and performs the following to the receiving port:

- If any IGMP membership report in response to the group-specific query arrives to the member port before its aging timer expires, this means that some other members of that multicast group still exist under that port: the switch resets the aging timer of the member port.
- If no IGMP membership report in response to the group-specific query arrives to this member port before its aging timer expires as a response to the IGMP group-specific query, this means that no members of that multicast group still

exist under the port: the switch deletes the forwarding entry corresponding to the port from the forwarding table when the aging timer expires.

Processing of Multicast Protocol Messages

Under different conditions, an IGMP Snooping-capable switch processes multicast protocol messages differently, 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 router port.
3. When IGMP is disabled on the switch, or when IGMP forwarding entries are cleared (by using the **reset igmp group** command):
   - If PIM is disabled, the switch clears all its Layer 2 multicast entries and router ports.
   - If PIM is enabled, the switch clears only its Layer 2 multicast entries without deleting its router ports.
4. When PIM is disabled on the switch:
   - If IGMP is disabled, the switch clears all its router ports.
   - If IGMP is enabled, the switch maintains all its Layer 2 multicast entries and router ports.

IGMP Snooping Configuration Task List

Complete these tasks to configure IGMP Snooping:

Task		Remarks
"Configuring Basic Functions of IGMP Snooping" on page 544	"Enabling IGMP Snooping" on page 544	Required
	"Configuring the Version of IGMP Snooping" on page 545	Optional
	"Configuring the Function of Dropping Unknown Multicast Data" on page 545	Optional
"Configuring IGMP Snooping Port Functions" on page 546	"Configuring Static Ports" on page 546	Optional
	"Configuring Simulated Joining" on page 547	Optional
	"Enabling the Fast Leave Feature" on page 548	Optional
	"Configuring Port Aging Timers" on page 549	Optional

Task		Remarks
"Configuring IGMP-Related Functions" on page 549	"Enabling IGMP Snooping Querier" on page 550	Optional
	"Configuring IGMP Timers" on page 550	Optional
	"Configuring Source IP Address of IGMP Queries" on page 551	Optional
	"Configuring Port Aging Timers" on page 549	Optional
"Configuring a Multicast Group Policy" on page 552	"Configuring a Multicast Group Filter" on page 552	Optional
	"Configuring Maximum Multicast Groups that Can Be Joined on a Port" on page 553	Optional
	"Configuring Multicast Group Replacement" on page 554	Optional

- *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 interface view are effective only for the current interface; 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 interface view or port group view.*

Configuring Basic Functions of IGMP Snooping

Configuration Prerequisites

Before configuring the basic functions of IGMP Snooping, complete the following tasks:

- Configure the corresponding VLANs
- Configure the corresponding port groups

Before configuring the basic functions of IGMP Snooping, consider the following points:

- Version of IGMP Snooping
- Whether to enable the function of dropping unknown multicast data.

Enabling IGMP Snooping

Follow these steps to enable IGMP Snooping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
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

- *IGMP Snooping must be enabled globally before it can be enabled in a VLAN.*
- *When you enable IGMP Snooping in a specified VLAN, this function takes effect for Ethernet ports in this VLAN only.*
- *After enabling IGMP Snooping in a VLAN, you cannot enable IGMP and/or PIM on the corresponding VLAN interface, and vice versa.*

Configuring the Version of IGMP Snooping

By configuring the 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*	-
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 for 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” on page 546.

Configuring the Function of Dropping Unknown Multicast Data

Unknown multicast data refers to multicast data whose forwarding entries do not exist in the corresponding multicast forwarding table:

- With the function of dropping unknown multicast data enabled, the switch drops all the unknown multicast data received.
- With the function of dropping unknown multicast data disabled, the switch floods unknown multicast data in the VLAN to which the unknown multicast data belongs.

Follow these steps to configure the function of dropping unknown multicast data:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IGMP Snooping view	**igmp-snooping**	-
Enable the function of dropping unknown multicast data	**drop-unknown**	Required Disabled by default

ⓘ *A Switch 8800 still forwards unknown multicast data to other router ports in the VLAN even if enabled to drop unknown multicast data.*

Configuring IGMP Snooping Port Functions

Configuration Prerequisites

Before configuring IGMP Snooping port functions, complete the following task:

- Enable IGMP Snooping in the VLAN or enable IGMP on the desired VLAN interface

Before configuring IGMP Snooping port functions, consider the following points:

- Multicast group and multicast source addresses
- Whether to configure simulated joining
- Whether to enable the fast leave feature
- Router port aging time
- Member port aging time

Configuring Static Ports

If the host attached to a port is 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 this port to be a group-specific or source-and-group-specific static member port (static (*, G) or (S, G) joining).

In a network where the topology structure is unlikely to change, you can configure ports of a switch to be static router ports. The switch forwards multicast traffic to static router ports as well as to member ports.

Follow these steps to configure static ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	

To do...	Use the command...	Remarks
Configure a static member port	**igmp-snooping static-group** *group-address* [**source-ip** *source_address*] **vlan** *vlan-id*	Required Disabled by default
Configure a static router port	**igmp-snooping static-router-port vlan** *vlan-id*	Required Disabled by default

- *The static (S, G) joining function is available only if a valid multicast source address is specified and IGMP Snooping version 3 is currently running on the switch.*
- *When you configure or remove a port as a static (*, G) or (S, G) member, the port will not send an IGMP report or an IGMP leave message.*
- *Static member ports and static router ports never age out. To delete such a port, you need to use the corresponding command.*

Configuring Simulated Joining

Generally, a host running IGMP responds to IGMP queries from a multicast router. If a host fails to respond due to some reasons, the multicast router will 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 of the multicast group. When an IGMP query arrives, that member port will give a response. As a result, the switch can continue receiving multicast data.

Through this configuration, the following functions can be implemented:

- When an Ethernet port is configured as a simulated member host, it sends an IGMP report.
- When receiving an IGMP general query, the simulated host responds with an IGMP report just like a real host.
- When the simulated joining function is disabled on an Ethernet port, the simulated host sends an IGMP leave message.

Follow these steps to configure simulated joining:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	

To do...	Use the command...	Remarks
Configure simulated (*, G) or (S, G) joining	**igmp-snooping host-join** *group-address* [**source-ip** *source-address*] **vlan** *vlan-id*	Required Disabled by default

- *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.*
- *The IGMP version of a simulated host is the same as the IGMP Snooping version current running on the switch.*

Enabling the Fast Leave Feature

By default, when receiving a group-specific IGMP leave message on a port, the switch first sends an IGMP group-specific query message that port, rather than directly deleting the port from the multicast forwarding table. If the switch receives no IGMP reports within a certain period of waiting time, it deletes the port from the forwarding table.

With the fast leave feature enabled, when the switch receives a group-specific IGMP leave message on a port, the switch directly deletes this port from the forwarding table without first sending an IGMP group-specific query to the port. If only one host is attached to the port, enable the fast leave feature to improve bandwidth and resource usage.

Configuring the fast leave feature globally

Follow these steps to configure the fast leave feature globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IGMP Snooping view	**igmp-snooping**	-
Enable the fast leave feature	**fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Configuring the fast leave feature on a port or a group ports

Follow these steps to configure the fast leave feature on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the fast leave feature		**igmp-snooping fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

CAUTION: *If the fast leave feature 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 Port Aging Timers

If the switch does not receive an IGMP general query or a PIM hello message before the aging timer of a router port expires, the switch deletes this port from the router port list when the aging timer times out.

If the switch does not receive an IGMP report for a multicast group before the aging timer of a member port expires, the switch deletes this port from the forwarding table for that multicast group when the aging timers times out.

If multicast group memberships change frequently, you can set a relatively small value for the member port aging timer, and vice versa.

Configuring port aging timers globally

Follow these steps to configure port aging timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IGMP Snooping view	**igmp-snooping**	-
Configure router port aging time	**router-aging-time** *interval*	Optional 105 seconds by default
Configure member port aging time	**host-aging-time** *interval*	Optional 260 seconds by default

Configuring port aging timers in a VLAN

Follow these steps to configure port aging timers in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure router port aging time	**igmp-snooping router-aging-time** *interval*	Optional 105 seconds by default
Configure member port aging time	**igmp-snooping host-aging-time** *interval*	Optional 260 seconds by default

Configuring IGMP-Related Functions

Configuration Prerequisites

Before configuring IGMP-related functions, complete the following task:

- Enable IGMP Snooping in the VLAN

Before configuring IGMP-related functions, consider the following points:

- IGMP general query interval
- IGMP last-member query interval
- Maximum response time to IGMP general queries

- Source address of IGMP general queries
- Source address of IGMP group-specific queries
- Whether to enable IGMP report suppression

Enabling IGMP Snooping Querier

In an IP multicast network running IGMP, a Layer 3 multicast device acts as the IGMP querier, which periodically sends IGMP queries so that all Layer 3 multicast devices can create and maintain multicast forwarding entries at the network layer, thus to forward multicast traffic correctly at the network layer.

In a network that does not comprise Layer 3 multicast devices, however, it is a problem to implement an IGMP querier, because Layer 2 devices do not support IGMP. To solve this problem, you can enable the IGMP Snooping querier function on a Layer 2 device so that it can work as an IGMP Snooping querier to create and maintain multicast forwarding entries 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:

- *An IGMP Snooping querier does not take part in IGMP querier elections.*
- *It is meaningless to configure an IGMP Snooping querier in a multicast network running IGMP. It may affect IGMP querier elections because it sends IGMP general queries with a low source IP address.*

Configuring IGMP Timers

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 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 to the IGMP last-member query interval.

Configuring IGMP timers globally

Follow these steps to configure IGMP timers 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 timers in a VLAN

Follow these steps to configure IGMP timers in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure IGMP general query interval	**igmp-snooping query-interval** *interval*	Optional 60 second 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.*

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 will not set that port as a 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 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

To do...	Use the command...	Remarks
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 election within the segment.*

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 a Multicast Group Policy

Configuration Prerequisites

Before configuring a multicast group filtering policy, complete the following task:

- Enable IGMP Snooping in the VLAN or enable IGMP on the desired VLAN interface

Before configuring a multicast group filtering policy, consider the following points:

- ACL rule for multicast group filtering
- The maximum number of multicast groups that can pass the ports
- Whether to configure multicast group replacement

Configuring a Multicast Group Filter

On an IGMP Snooping-enabled switch, the configuration of a multicast group 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 IGMP report. Upon receiving this report message, the switch checks the report against the ACL rule configured on the receiving port. If the receiving port can join this multicast group, the switch adds this port to the IGMP Snooping multicast group list; 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 filter is configured by default, namely hosts can join any multicast group

Configuring a multicast group filter on a port or a group ports

Follow these steps to configuring a multicast group filter on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure a multicast group filter		**igmp-snooping group-policy** *acl-number* [**vlan** *vlan-list*]	Required No filter is configured by default, namely hosts can join any multicast group

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 that can be joined on a port or ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the maximum number of multicast groups that can be joined on the port(s)		**igmp-snooping group-limit** *limit* [**vlan** *vlan-list*]	Optional 512 by default..

- *When the number of multicast groups a port has joined reaches the maximum number configured, the system deletes this port from all the related IGMP Snooping forwarding entries, and hosts on this port need to join the multicast groups again.*
- *If you have configured a port to be a static member port or a simulated member of a multicast group, the system deletes this port from all the related IGMP Snooping forwarding entries and applies the configurations again, until the number of multicast groups the port has joined reaches the maximum number configured.*

Configuring Multicast Group Replacement

For some special reasons, the number of multicast groups that can be joined on the current switch or Ethernet 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, 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 Ethernet ports. When the number of multicast groups joined on the switch or an Ethernet 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**	-
Configure multicast group replacement	**overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Configuring multicast group replacement on a port or a group port

Follow these steps to configure multicast group replacement on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the correspon ding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter interface group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure 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 to be in the range of 1 to 511 before configuring multicast group replacement. Otherwise, the multicast group replacement functionality will not take effect.

Displaying and Maintaining IGMP Snooping

To do...	Use the command...
View the information of multicast groups learned by IGMP Snooping	**display igmp-snooping group** [**vlan** *vlan-id*] [**slot** *slot-id*] [**verbose**]
View the statistics information of IGMP messages learned by IGMP Snooping	**display igmp-snooping statistics**
Clear IGMP Snooping entries	**reset igmp-snooping group** { *group-address* | **all** } [**vlan** *vlan-id*]
Clear the statistics information of all kinds of IGMP messages learned by IGMP Snooping	**reset igmp-snooping statistics**

- *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 IGMP Snooping entries of static joins.*

IGMP Snooping Configuration Examples

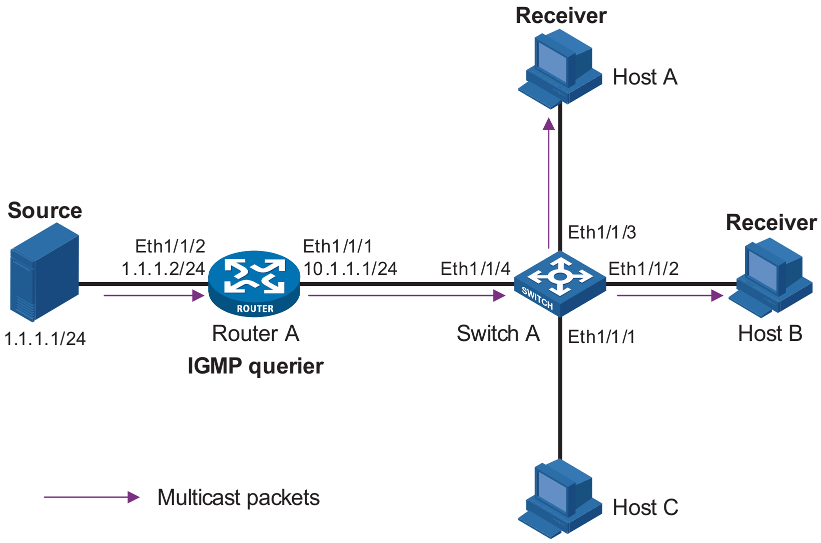
Configuring Simulated Joining

Network requirements

- As shown in Figure 171, Router A connects to the multicast source through Ethernet1/1/2 and to Switch A (Switch 8800) through Ethernet1/1/1. Router A is the IGMP querier on the subnet.
- IGMPv3 runs on Router A and Switch A. IGMPv3 Snooping runs on Switch A. PIM-SM runs on Router A. Ethernet1/1/2 is to be configured as C-BSR and C-RP.
- Perform the following configuration so that multicast data (1.1.1.1, 224.1.1.1) can be forwarded through Ethernet1/1/2 and Ethernet1/1/3 even if Host A and Host B temporarily stop receiving multicast data for some unexpected reasons.

Network diagram

Figure 171 Network diagram for simulated joining configuration



Configuration procedure

1 Configure the IP address of each interface

Configure an IP address and subnet mask for each interface as per Figure 171. The detailed configuration steps are omitted.

2 Configure Router A

# Enable IP multicast routing, enable PIM-SM on each interface, and enable IGMPv3 on Ethernet1/1/1 and configure Ethernet1/1/2 as C-BSR and C-RP.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface ethernet 1/1/1
[RouterA-Ethernet1/1/1] igmp enable
[RouterA-Ethernet1/1/1] igmp version 3
[RouterA-Ethernet1/1/1] pim sm
[RouterA-Ethernet1/1/1] quit
[RouterA] interface ethernet 1/1/2
[RouterA-Ethernet1/1/2] pim sm
[RouterA-Ethernet1/1/2] quit
[RouterA] pim
[RouterA-pim] c-bsr ethernet 1/1/2
[RouterA-pim] c-rp ethernet 1/1/2
[RouterA-pim] quit
```

The configurations mentioned above are for reference only. Specific configurations are subject to the actual implementations of the devices.

3 Configure Switch A

# Create VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
```

# Assign Ethernet1/1/1 through Ethernet1/1/4 to this VLAN.

```
[SwitchA-vlan100] port ethernet 1/1/1 to ethernet 1/1/4
[SwitchA-vlan100] quit
```

# Enable IGMP Snooping in VLAN 100 and set the version to 3.

```
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
[SwitchA] vlan 100
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping version 3
[SwitchA-vlan100] quit
```

# Enable simulated (S, G) joining on Ethernet 1/1/2 and Ethernet 1/1/3 respectively.

```
[SwitchA] interface ethernet 1/1/2
[SwitchA-Ethernet1/1/2] igmp-snooping host-join 232.1.1.1 source-ip
1.1.1.1 vlan 100
[SwitchA-Ethernet1/1/2] quit
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] igmp-snooping host-join 232.1.1.1 source-ip
1.1.1.1 vlan 100
[SwitchA-Ethernet1/1/3] quit
```

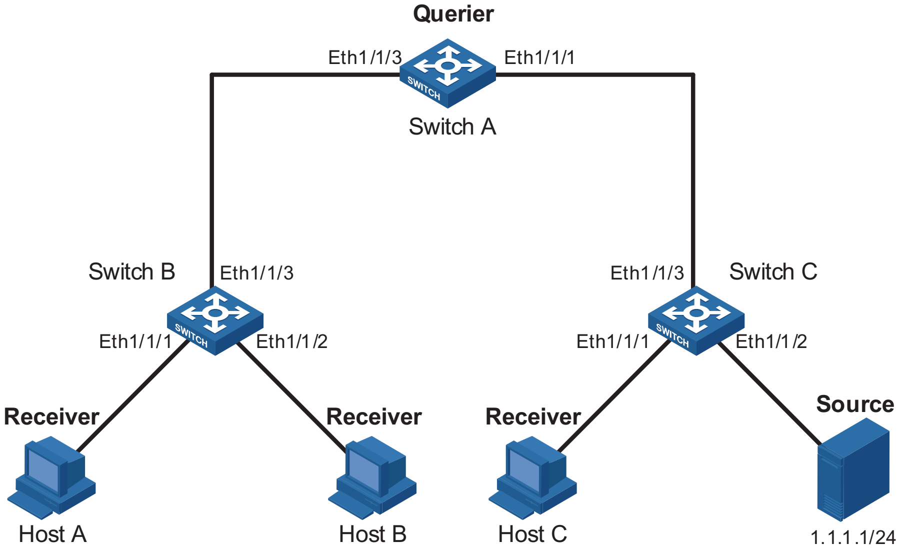
IGMP Snooping Querier Configuration

Network requirements

- As shown in Figure 172, in a Layer-2 network environment without Layer-3 devices, Switch C is connected to the multicast source (Source) through Ethernet1/1/2. At least one receiver is attached to Switch B and Switch C respectively.
- All the receiver hosts run IGMPv2. Switch A, Switch B, and Switch C run IGMP Snooping. Switch A is to be configured as the IGMP Snooping querier.
- Configure a non-all-zero IP address as the source IP address of IGMP queries to ensure normal creation of multicast forwarding entries.

Network diagram

Figure 172 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 add Ethernet1/1/3 and Ethernet1/1/1 to VLAN 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port ethernet 1/1/1
[SwitchA-vlan100] port ethernet 1/1/3
```

# Enable IGMP Snooping in VLAN 100 and enable the IGMP-Snooping querier feature.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping querier
```

# Set the source IP address of IGMP general queries and group-specific queries to 192.168.1.1.

```
[SwitchA-vlan100] igmp-snooping general-query source-ip 192.168.1.1
[SwitchA-vlan100] igmp-snooping special-query source-ip 192.168.1.1
```

2 Configure Switch B.

# Enable IGMP Snooping globally.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

# Create VLAN 100, add Ethernet1/1/1 through Ethernet1/1/3 to VLAN 100, and enable IGMP Snooping in this VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port ethernet 1/1/1 to ethernet 1/1/3
[SwitchB-vlan100] igmp-snooping enable
```

3 Configuration on Switch C.

# Enable IGMP Snooping globally.

```
<SwitchC> system-view
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

# Create VLAN 100, add Ethernet1/1/1 through Ethernet1/1/3 into VLAN 100, and enable IGMP Snooping in this VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port ethernet 1/1/1 to ethernet 1/1/3
[SwitchC-vlan100] igmp-snooping enable
```

Static Router Port Configuration

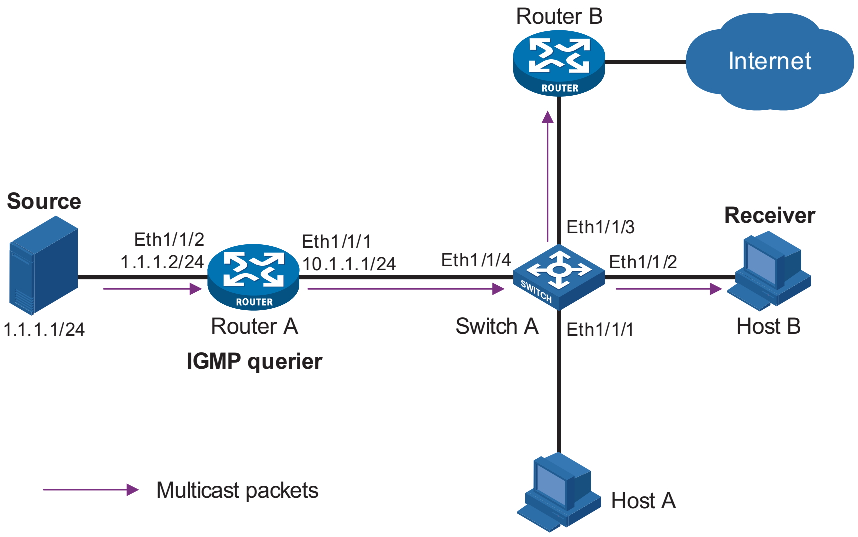
Network requirements

- As shown in Figure 173. Router A, which acts as the IGMP querier on the subnet, connects to the multicast source through Ethernet1/1/2 and to Switch A (Switch 8800) through Ethernet1/1/1.
- IGMPv2 runs on Router A and Switch A. IGMPv2 Snooping runs on Switch A. Router A works as the IGMP querier.

- While no multicast protocol runs on Router B, perform the following configuration so that Switch A can forward all received multicast data to Router B.

Network diagram

Figure 173 Network diagram for static router port configuration



Configuration procedure

1 Configure the IP address of each interface

Configure an IP address and subnet mask for each interface as per Figure 173. The detailed configuration steps are omitted.

2 Configure Router A

# Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMPv2 on Ethernet1/1/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface ethernet 1/1/1
[RouterA-Ethernet1/1/1] igmp enable
[RouterA-Ethernet1/1/1] igmp version 2
[RouterA-Ethernet1/1/1] pim dm
[RouterA-Ethernet1/1/1] quit
[RouterA] interface ethernet 1/1/2
[RouterA-Ethernet1/1/2] pim dm
[RouterA-Ethernet1/1/2] quit
```

The configurations mentioned above are for reference only. Specific configurations are subject to the implementations of the devices.

3 Configure Switch A

# Create VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
```

# Add Ethernet1/1/1 through Ethernet1/1/4 into VLAN 100.

```
[SwitchA-vlan100] port ethernet 1/1/1 to ethernet 1/1/4
[SwitchA-vlan100] quit
```

# Enable IGMP Snooping in VLAN 100 and configure the version as 2.

```
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
[SwitchA] vlan 100
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping version 2
[SwitchA-vlan100] quit
```

# Configure Ethernet1/1/3 as a static router port.

```
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] igmp-snooping static-router-port vlan 100
[SwitchA-Ethernet1/1/3] quit
```

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 applied.
- The function of dropping unknown multicast data is not enabled, so unknown multicast data is broadcast.
- Certain ports have been configured as static member ports of multicasts groups, and this configuration conflicts with the configured multicast group policy.

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 interface 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 **drop-unknown** command to enable the function of dropping unknown multicast data.

4 Use the **display igmp-snooping group** command to check whether any port has been configured as a static member port of any multicast group. If so, check whether this configuration conflicts with the configured multicast group policy. If any conflict exists, remove the port as a static member of the multicast group.

42 PIM CONFIGURATION

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

- "PIM Overview" on page 563
- "Configuring PIM-DM" on page 575
- "Configuring PIM-SM" on page 577
- "Configuring PIM-SSM" on page 586
- "Configuring PIM Common Information" on page 587
- "Displaying and Maintaining PIM" on page 592
- "PIM Configuration Examples" on page 593
- "Troubleshooting PIM Configuration" on page 604

The term "router" in this document refers to a router in a generic sense or a Switch 8800 running the PIM protocol.

PIM Overview

Protocol Independent Multicast (PIM) provides IP multicast forwarding by leveraging 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 *"RPF Mechanism" on page 503*.

Based on the forwarding mechanism, PIM falls into two modes:

- Protocol Independent Multicast-Dense Mode (PIM-DM), and
- Protocol Independent Multicast-Sparse Mode (PIM-SM).

Presently, the any-source multicast (ASM) model implementations include the PIM-DM and PIM-SM modes, while the source-specific multicast (SSM) model can be implemented by leveraging part of the PIM-SM technique.

To simplify 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).

Every activated interface on a router sends hello messages periodically, and thus learns the PIM neighboring information pertinent to the interface.

SPT building

The process of building an SPT is the process of "flood and prune".

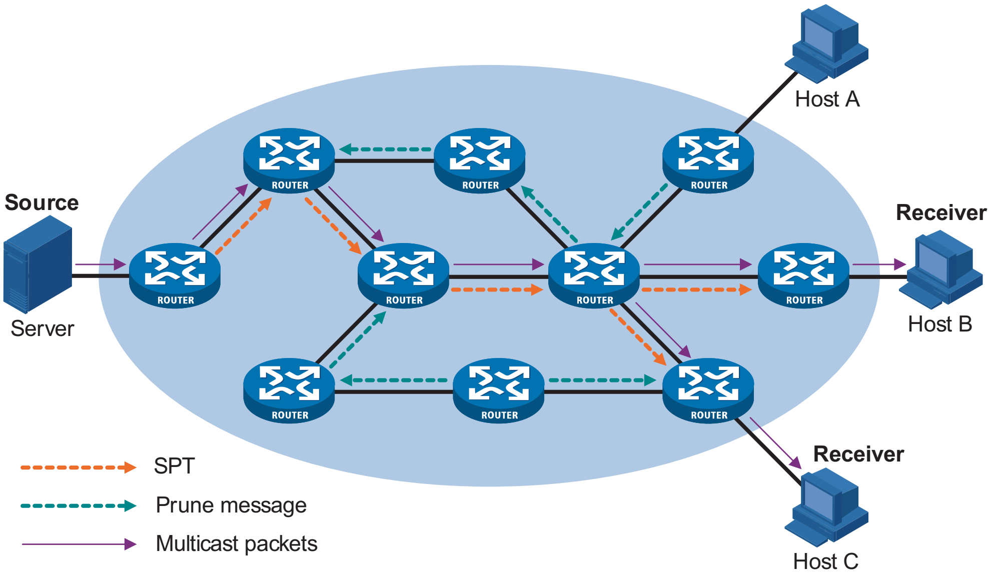
1 **In a PIM-DM domain, when a multicast source S sends multicast data to a multicast group G, the multicast packet is first flooded throughout the domain: The router first performs RPF check on the multicast packet. If the packet passes the RPF check, the router creates an (S, G) entry and forwards 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 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 multicast group down to this node.**

- *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 174, 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 174 SPT building



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.

Graft

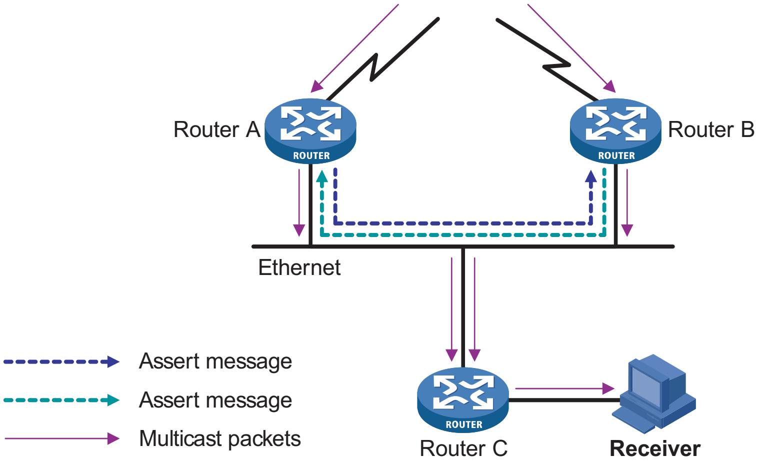
When a host attached to a pruned node joins a multicast group, to reduce the join latency, PIM-DM uses a graft mechanism to resume data forwarding to that branch. The process is as follows:

1. **The node that need to receive multicast data sends a graft message hop by hop toward the source, 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

If multiple multicast routers exist on a multi-access subnet, duplicate packets may flow to the same subnet. To shutoff duplicate flows, the assert mechanism is used for election of a single multicast forwarder on a multi-access network.

Figure 175 Assert mechanism



As shown in Figure 175, 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 mod 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 larger 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 a multicast packet to a multicast group, the router directly connected with the multicast source first registers the multicast source with the RP by sending a register message to the RP by unicast. The arrival of this 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.

Multicast traffic is duplicated only where the distribution tree branches. 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 formation
- Multicast source registration
- Switchover from RPT to SPT
- Assert

Neighbor discovery

PIM-SM uses exactly the same neighbor discovery mechanism as PIM-DM does. Refer to “Neighbor discovery” on page 564 described above.

DR election

PIM-SM also uses hello messages to elect a designated router (DR) for a multi-access network (such as an LAN). The elected DR will be the only multicast forwarder on this multi-access network.

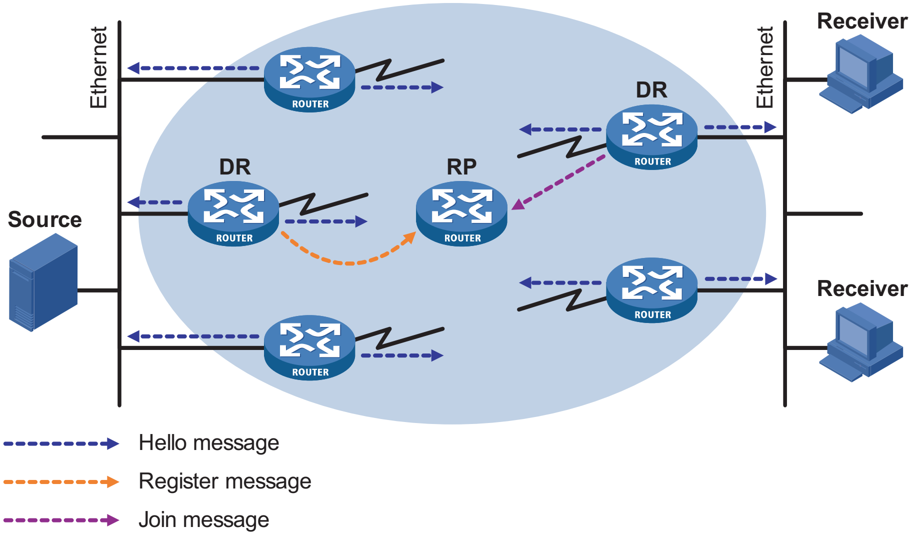
A DR must be elected in a multi-access network, no matter this network connects to multicast sources or to receivers. The DR at the receiver side sends join messages to the RP; the DR at the multicast source side sends register messages to the RP.

- *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 any IGMPv1 router exists in a PIM-DM domain, a DR must be elected to act as the IGMPv1 querier.*

- *IGMP must be enabled on a device that acts as a DR before receivers attached to this device can join multicast groups through this DR.*

For details about IGMP, refer to *"IGMP Configuration" on page 523*.

Figure 176 DR election



As shown in Figure 176, 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

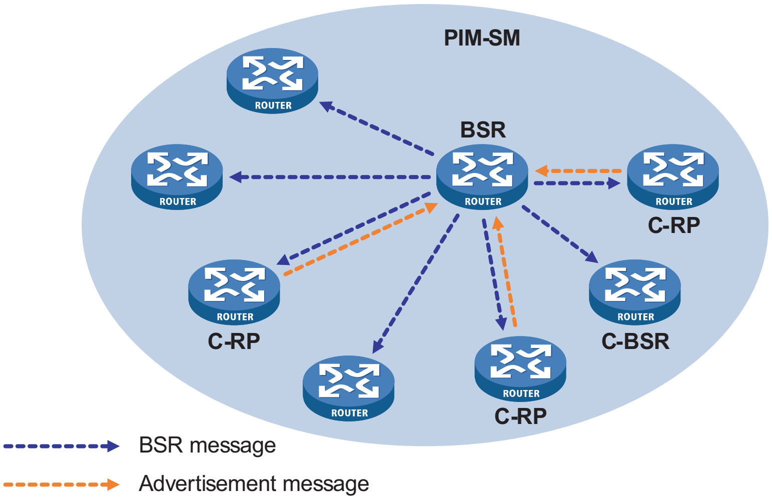
The RP is the core of a PIM-SM domain. For a small-sized, simple network, one RP is enough for forwarding information throughout the network, and the position of the RP can be statically specified on each router in the PIM-SM domain. In most cases, however, a PIM-SM network covers a wide area and a huge amount of multicast traffic needs to be forwarded through the RP. To lessen the RP burden and optimize the topological structure of the RPT, each multicast group should have its own RP. Therefore, a bootstrap mechanism is needed for dynamic RP election. For this purpose, a bootstrap router (BSR) should be configured.

As the administrative core of a PIM-SM domain, the BSR collects advertisement messages (C-RP-Adv messages) from candidate-RPs (C-RPs) 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 floods the RP-Set to the entire PIM-SM domain. Based on the information in these RP-Sets, all routers (including the DRs) in the network can calculate the location of the corresponding RPs.

A PIM-SM domain (or an administratively scoped region) 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 through the bootstrap mechanism to avoid service interruption. Similarly, multiple C-RPs can be configured in a PIM-SM domain, and the position of the RP corresponding to each multicast group is calculated through the BSR mechanism.

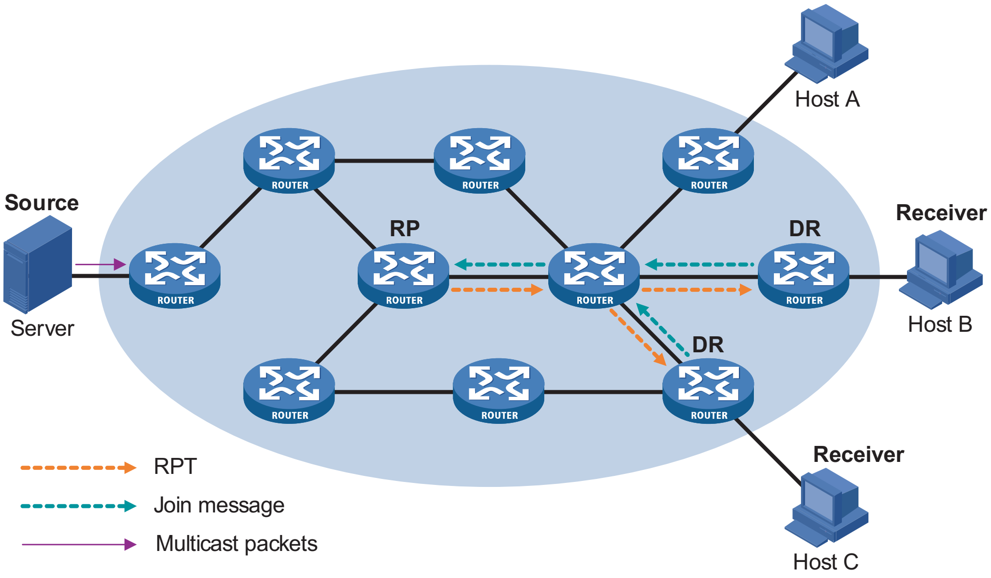
Figure 177 shows the positions of C-RPs and the BSR in the network.

Figure 177 BSR and C-RPs



RPT formation

Figure 178 Build an RPT in PIM-SM



As shown in Figure 178, host B and host C are receivers of multicast data. The process of building an RPT is as follows:

1 **When a receiver joins a 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.**

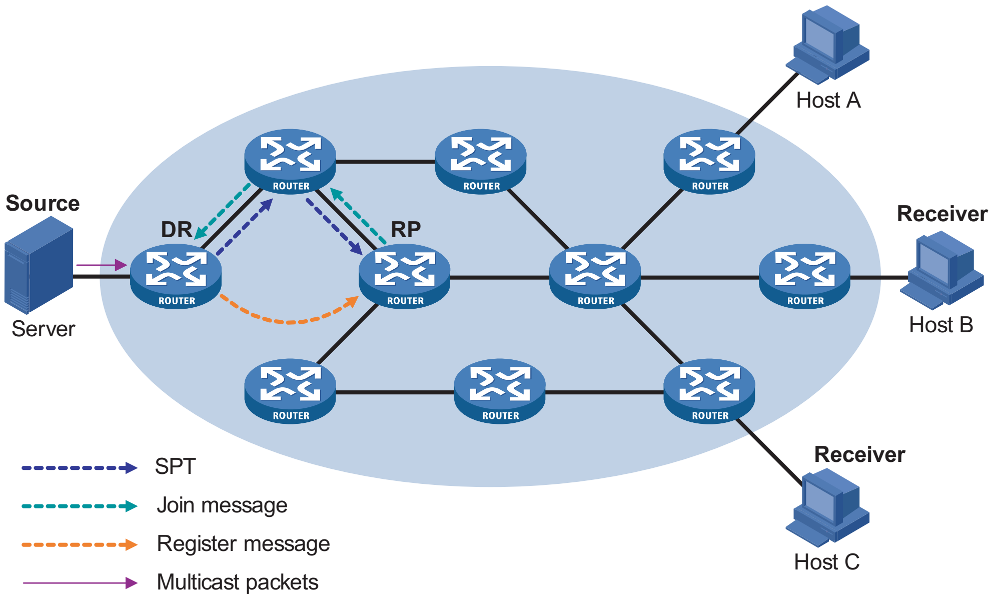
The multicast data addressed to the multicast group G flows through the RP, reaches the corresponding DR along the established RPT, and finally is delivered to the receiver.

When a receiver is no longer interested in the multicast data addressed to 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 its link 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 179 Multicast registration



As shown in Figure 179, the multicast source registers with the RP as follows:

1 **When the multicast source S sends the first multicast packet to a 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, on one hand, it extracts the multicast packet from the register message and forwards the multicast packet down the RPT, and, on the other hand, it 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 a (S, G) entry in its forwarding table. The multicast source 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.**

Switchover from RPT to SPT

Initially, multicast traffic flows along an RPT from the RP to the receivers. Because the RPT is not necessarily the tree that has the shortest path, the receiver-side DR initiates an RPT-to-SPT switchover process upon receiving the first multicast packet along the RPT by default.

1 **First, the receiver-side DR sends an (S, G) join message hop by hop to 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.**

2 **Subsequently, the receiver-side DR sends a prune message hop by hop to the RP. Upon receiving this prune message, the RP forwards it towards the multicast source, thus to implement RPT-to-SPT switchover.**

After the RPT-to-SPT switchover, multicast data can be directly sent from the source to the receivers. PIM-SM builds SPTs through RPT-to-SPT switchover more economically than PIM-DM does through the "flood and prune" mechanism.

Assert

PIM-SM uses exactly the same assert mechanism as PIM-DM does. Refer to “Assert” on page 566 described above.

Introduction to BSR Admin-scope Regions 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.

To implement refined management and provide group-specific services, a PIM-SM domain can be divided into one global scope zone and multiple BSR administratively scoped regions (BSR admin-scope regions), like the division of subnets.

Specific to particular multicast groups, the BSR administrative scoping mechanism effectively lessens the management workload of a single-BSR domain and provides group-specific services.

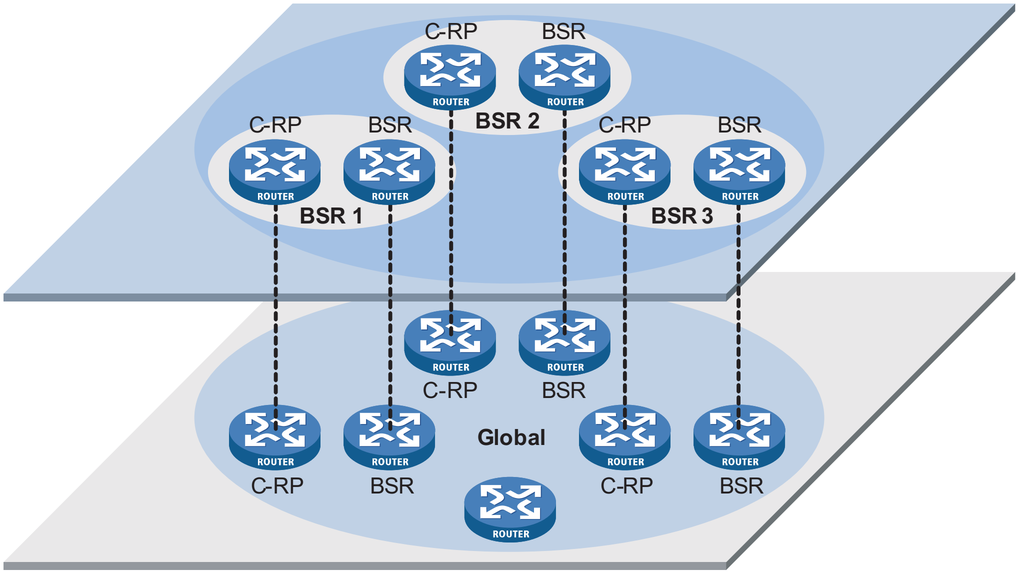
Relationship between BSR admin-scope regions and the global scope zone

A better understanding of the global scope zone and BSR admin-scope regions should be based on two aspects: geographical space and group address range.

1 Geographical space

BSR admin-scope regions are logical regions specific to particular multicast groups, and each BSR admin-scope region must be geographically independent of another, as shown in Figure 180.

Figure 180 Relationship between BSR admin-scope regions and the global scope zone in geographic space

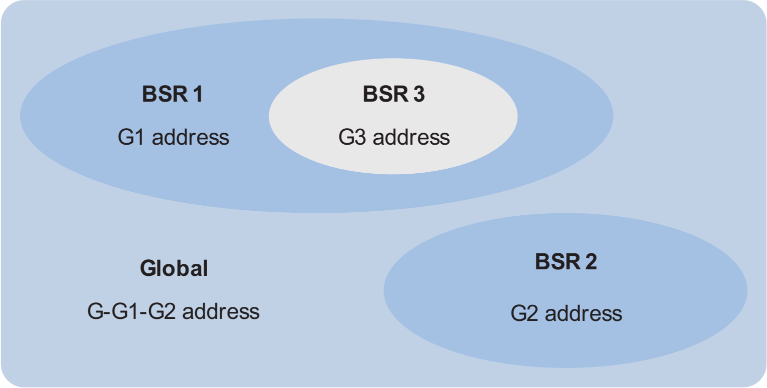


BSR admin-scope regions are geographically segregated from one another. Namely, a router must not serve different BSR admin-scope regions. In other words, different BSR admin-scope regions contain different routers, whereas the global scope zone covers all routers in the PIM-SM domain.

2 In terms of multicast group address ranges

Each BSR admin-scope region serves specific multicast groups. Usually, these addresses have no intersections; however, they may overlap one another.

Figure 181 Relationship between BSR admin-scope regions and the global scope zone in group address ranges



In Figure 181, the group address ranges of admin-scope-scope regions BSR1 and BSR2 have no intersection, whereas the group address range of BSR3 is a subset of the address range of BSR1. The group address range of the global scope zone covers all the group addresses other than those of all the BSR admin-scope regions. 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 BSR admin-scope regions in terms of group address ranges.

Relationships between BSR admin-scope regions and the global scope zone are as follows:

- The global scope zone and each BSR admin-scope region have their own C-RPs and BSR. These devices are effective only in their respective admin-scope regions. Namely, the BSR election and RP election are implemented independently within each admin-scope region.
- Each BSR admin-scope region has its own boundary. The multicast information (such as C-RP-Adv messages and BSR bootstrap messages) can be transmitted only within the domain.
- Likewise, the multicast information in the global scope zone cannot enter any BSR admin-cope region.
- In terms of multicast information propagation, BSR admin-scope regions are independent of one another and each BSR admin-scope region is independent of the global scope zone, and no overlapping is allowed between any two BSR admin-scope regions.

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" on page 564 described above.

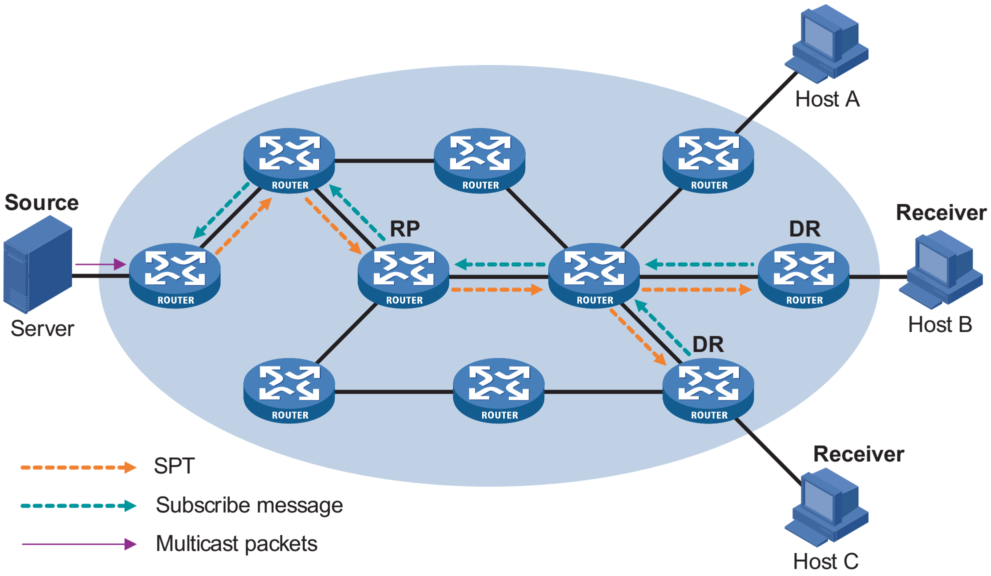
DR election

PIM-SSM uses the same DR election mechanism as in PIM-SM. Refer to "DR election" on page 567 described above.

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 address range (the default SSM group address range is 232.0.0.0/8).

Figure 182 SPT establishment in PIM-SSM



As shown in Figure 182, Host B and Host C are multicast information receivers. They send IGMPv3 report messages denoted as (Include S, G) to the respective DRs to express their interest in the information of the specific multicast source S. However described, the position of multicast source S is explicitly specified for receivers.

The DR that has received the report first checks whether the group address in this message falls in the SSM group address range:

- If so, the DR sends a subscribe message for channel subscription hop by hop toward the multicast source S. An (Include S, G) 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.

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 2362: Protocol Independent Multicast-sparse Mode (PIM-SM): Protocol Specification
- RFC 3973: Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification(Revised)
- draft-ietf-pim-sm-v2-new-06: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- draft-ietf-pim-dm-new-v2-02: Protocol Independent Multicast-Dense Mode (PIM-DM)
- draft-ietf-pim-v2-dm-03: Protocol Independent Multicast Version 2 Dense Mode Specification
- draft-ietf-pim-sm-bsr-03: Bootstrap Router (BSR) Mechanism for PIM Sparse Mode
- draft-ietf-ssm-arch-02: Source-Specific Multicast for IP
- draft-ietf-ssm-overview-04: 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" on page 576	Required
"Enabling State Refresh" on page 576	Optional
"Configuring State Refresh Parameters" on page 576	Optional
"Configuring PIM-DM Graft Retry Period" on page 577	Optional
"Configuring PIM Common Information" on page 587	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 device sends hello messages periodically to discover PIM neighbors and processes messages from PIM neighbors. When deploying a PIM-DM domain, you are recommended to enable PIM-DM on all interfaces of non-border devices (border devices are PIM-enabled routers or PIM-enabled switches located on the boundary of BSR admin-scope regions).

Follow these steps to enable PIM-DM:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable IP multicast routing	**multicast routing-enable**	Required Disable by default
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
Enable PIM-DM	**pim dm**	Required Disabled by default

CAUTION:

- *All the interfaces of the same router must work in the same PIM mode.*
- *After PIM-DM is enabled on a VLAN interface, IGMP snooping cannot be enabled in the VLAN corresponding to the VLAN interface, and vice versa.*

Enabling State Refresh

An interface without the state refresh capability cannot forward state refresh messages.

Follow these steps to enable the state refresh capability:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN/POS 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 device 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 devices on the path.

A device may receive multiple state refresh messages within a short time, of which some may be duplicated messages. To keep a device from receiving such duplicated messages, you can configure the time the device must wait before receiving the next state refresh message. If a new state refresh message is received within the waiting time, the device will discard it; if this timer times out, the device will accept a new state refresh message, refresh its own PIM state, and reset the waiting timer.

The TTL value of a state refresh message decrements by 1 whenever it passes a device 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.

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 device does not receive a graft-ack message from the upstream device within the specified time after it sends a graft message, the device 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 VLAN/POS interface view	**interface** *interface-type interface-number*	-
Configure graft retry period	**pim timer graft-retry** *interval*	Optional 3 seconds by default

For the configuration of other timers in PIM-DM, refer to “Configuring PIM Common Timers” on page 590.

Configuring PIM-SM

A device can serve as a C-RP and a C-BSR at the same time.

PIM-SM Configuration Task List

Complete these tasks to configure PIM-SM:

Task		Remarks
"Enabling PIM-SM" on page 579		Required
"Configuring a BSR" on page 579	"Performing basic C-BSR configuration" on page 579	Optional
	"Configuring a global-scope C-BSR" on page 580	Optional
	"Configuring an admin-scope C-BSR" on page 581	Optional
	"Configuring a BSR admin-scope region boundary" on page 581	Optional
	"Configuring global C-BSR parameters" on page 582	Optional
"PIM Configuration" on page 563	"Configuring a C-RP" on page 583	Optional
	"Enabling auto-RP" on page 583	Optional
	"Configuring C-RP timers" on page 584	Optional
	"Configuring a static RP" on page 582	Optional
"PIM Configuration" on page 563		Optional
"Disabling RPT-to-SPT Switchover" on page 585		Required
"Configuring PIM Common Information" on page 587		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:

- An ACL rule defining a legal BSR address range
- Hash mask length for RP selection calculation
- C-BSR priority
- Bootstrap interval
- Bootstrap timeout time
- An ACL rule defining a legal C-RP address range and the range of multicast groups to be served
- C-RP-Adv interval
- C-RP timeout time
- The IP address of a static RP
- An ACL rule for register message filtering
- Register suppression timeout time
- Probe time
- Whether to disable RPT-to-SPT switchover

Enabling PIM-SM

With PIM-SM enabled, a device sends hello messages periodically to discover PIM neighbors and processes messages from PIM neighbors. When deploying a PIM-SM domain, you are recommended to enable PIM-SM on all interfaces of non-border devices (border devices are PIM-enabled routers or PIM-enabled switches located on the boundary of BSR admin-scope regions).

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 VLAN/POS interface view	**interface** *interface-type interface-number*	-
Enable PIM-SM	**pim sm**	Required Disabled by default

CAUTION:

- *All the interfaces of the same device must work in the same PIM mode.*
- *After PIM-SM is enabled on a VLAN interface, IGMP snooping cannot be enabled in the VLAN corresponding to the VLAN interface, and vice versa.*

Configuring a BSR

The BSR is dynamically elected from a number of C-BSRs. Because it is unpredictable which device will finally win a BSR election, the commands introduced in this section must be configured on all C-BSRs.

About the Hash mask length and C-BSR priority for RP selection calculation:

- *You can configure these parameters at three levels: global configuration level, global scope level, and BSR admin-scope level.*
- *By default, the global scope parameters and BSR admin-scope parameters are those configured at the global configuration level.*
- *Parameters configured at the global scope level or BSR admin-scope level have higher priority than those configured at the global configuration level.*

Performing basic C-BSR configuration

A PIM-SM domain can have only one BSR, but must have at least one C-BSR. Any device can be configured as C-BSR. Elected from C-BSRs, a BSR is responsible for collecting and advertising RP information in the PIM-SM.

C-BSRs should be configured on devices in the backbone network. When configuring a router as a C-BSR, be sure to specify a PIM-SM-enabled interface. The BSR election process is 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 the

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 keeps its own BSR address and continues assuming itself to be the BSR.

Configuring a legal range of BSR addresses enables filtering of BSR messages based on the address range, thus to prevent malicious hosts from initiating attacks by disguising themselves as legitimate BSRs. To protect legitimate BSRs from being maliciously replaced, preventive measures are taken specific to the following two situations:

1. **Some malicious hosts intend to fool routers by forging BSR messages and change the RP mapping relationship. Such attacks often occur on border devices. 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 border devices to perform neighbor check and RPF check on BSR messages and discard unwanted messages.**
2. **When a device in the network is controlled by an attacker or when an illegal device is present in the network, the attacker can configure such a device to be a C-BSR and make it win BSR election so as to gain the right of advertising RP information in the network. After being configured as a C-BSR, a device automatically floods the network with BSR messages. As a BSR message has a TTL value of 1, the whole network will not be affected as long as the neighbor device discards these BSR messages. Therefore, if a legal BSR address range is configured on all devices in the entire network, all devices will discard BSR messages from out of the legal address range, and thus this kind of attacks can be prevented.**

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 C-BSR configuration:

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-BSR is configured by default
Configure a legal BSR address range	**bsr-policy** *acl-number*	Optional No restrictions on BSR address range by default

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-BSR and the other devices in the PIM-SM domain.

Configuring a global-scope C-BSR

Follow these steps to configure a global-scope C-BSR:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter PIM view	**pim**	-
Configure a global-scope C-BSR	**c-bsr global** [**hash-length** *hash-length* | **priority** *priority*] *	Required No global-scope C-BSRs by default

Configuring an admin-scope C-BSR

By default, a PIM-SM domain has only one BSR. The entire network should be managed by this one BSR. To manage your network more effectively and specially, you can divide a PIM-SM domain into multiple BSR admin-scope regions, with each BSR admin-scope region having one BSR, which services specific multicast groups.

Specific to particular multicast groups, the BSR administrative scoping mechanism effectively lessens the management workload of a single-BSR domain and provides group-specific services.

In a network divided into BSR admin-scope regions, BSRs are elected from multitudinous C-BSRs to service different multicast groups. The C-RPs in a BSR admin-scope region send C-RP-Adv messages to only the corresponding BSR. The BSR summarizes the advertisement messages into an RP-set and advertises it to all the devices in the BSR admin-scope region. All the devices use the same algorithm to get the RP addresses corresponding to specific multicast groups.

Follow these steps to configure an admin-scope C-BSR:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter PIM view	**pim**	-
Enable BSR administrative scoping	**c-bsr admin-scope**	Required Disabled by default
Configure an admin-scope C-BSR	**c-bsr group** *group-address* { *mask* | *mask-length* } [**hash-length** *hash-length* | **priority** *priority*] *	Optional No admin-scope BSRs by default

A BSR admin-scope region is effective only for the multicast groups whose addresses fall in the range of 239.0.0.0 to 239.255.255.255.

Configuring a BSR admin-scope region boundary

A BSR has its specific service scope. A number of BSR boundary interfaces divide a network into different BSR admin-scope regions. Bootstrap messages cannot cross the admin-scope region boundary, while other types of PIM messages can.

Follow these steps to configure a BSR admin-scope region boundary:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter VLAN/POS interface view	**interface** *interface-type interface-number*	-
Configure a BSR admin-scope region boundary	**pim bsr-boundary**	Required No BSR admin-scope region boundary by default

Configuring global C-BSR parameters

The BSR election winner advertises its own IP address and RP-Set information throughout the region it serves through bootstrap messages. The BSR floods bootstrap messages throughout the network periodically. Any C-BSR that receives a bootstrap message maintains the BSR state for a configurable period of time (BSR state timeout), during which no BSR election takes place. When the BSR state times out, a new BSR election process will be triggered among the C-BSRs.

Follow these steps to global C-BSR parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter PIM view	**pim**	-
Configure the Hash mask length for RP selection calculation	**c-bsr hash-length** *hash-length*	Optional 30 by default
Configure the C-BSR priority	**c-bsr priority** *priority*	Optional 0 by default
Configure the bootstrap interval	**c-bsr interval** *interval*	Optional 60 seconds by default
Configure the bootstrap timeout time	**c-bsr holdtime** *interval*	Optional 130 seconds by default

CAUTION: *In configuration, make sure that the bootstrap interval is smaller than the bootstrap timeout time.*

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. To enable a static RP to work normally, you must perform this configuration on all the devices in the PIM-SM domain and specify the same RP address.

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**]	Optional No static RP by default

Configuring a C-RP

In a PIM-SM domain, you can configure devices that intend to become the RP into 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 devices and organizes the information into to an RP-Set, which is flooded throughout the entire network. Then, the other devices 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 devices.

To guard again 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.

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*] *	Optional 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

- *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 respectively addressed to the multicast group addresses 224.0.1.39 and 224.0.1.40. 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**	Optional Disabled by default

Configuring C-RP timers

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 devices (224.0.0.13) in the network.

Each C-RP encapsulates a timeout value in its C-RP-Adv message. Upon receiving this 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.

Follow these steps to configure C-RP timers:

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

- *The commands introduced in this section are to be configured on C-RPs.*
- *For the configuration of other timers in PIM-SM, see "Configuring PIM Common Timers" on page 590.*

Configuring PIM-SM Register Messages

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 a specific 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 enters the register suppression state.

In a probe suppression cycle, the DR can send a null register message (a register message without multicast data encapsulated), a certain length of time defined by the probe time before the register suppression timer expires, to the RP to indicate that the multicast source is active. When the register suppression expires, the DR starts sending register messages again. A smaller register suppression timeout setting will cause the RP to receive bursting multicast data more frequently, while a larger timeout setting will result in a larger delay for new receivers to join the multicast group they are interested in.

Follow these steps to configure PIM-SM 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-header-checksum**	Optional By default, the checksum is calculated based on the header of register messages
Configure the register suppression timeout time	**register-suppression-timeout** *interval*	Optional 60 seconds by default
Configure the probe time	**probe-interval** *interval*	Optional 5 seconds by default

Typically, you need to configure the above-mentioned parameters on the receiver-side DR and the RP only. Since both the DR and RP are elected, however, you should carry out these configurations on the devices that may win the DR election and on the C-RPs that may win RP elections.

Disabling RPT-to-SPT Switchover

When a Switch 8800 serves as the receiver-side DR, by default, it initiates an RPT-to-SPT switchover process immediately after receiving the first multicast packet along the RPT. You can disable the RPT-to-SPT switchover function with the following command.

Follow these steps to disable RPT-to-SPT switchover:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter PIM view	**pim**	-

To do...	Use the command...	Remarks
Disable RPT-to-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 along the RPT.

To avoid forwarding failure, do not disable RPT-to-SPT switchover on a switch that may become an RP (namely, a static RP or a C-RP).

Configuring PIM-SSM

The SSM module needs the support of IGMPv3. Therefore, be sure to enable IGMPv3 on PIM devices with multicast receivers.

PIM-SSM Configuration Task List

Complete these tasks to configure PIM-SSM:

Task	Remarks
"Enabling PIM-SM" on page 586	Required
"Configuring the SSM Group Range" on page 587	Optional
"Configuring PIM Common Information" on page 587	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 device 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 all interfaces of non-border devices.

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 VLAN/POS interface view	**interface** *interface-type interface-number*	-
Enable PIM-SM	**pim sm**	Required Disabled by default

CAUTION:

- *All the interfaces of the same router must work in the same PIM mode.*
- *After PIM-SM is enabled on a VLAN interface, IGMP snooping cannot be enabled in the VLAN corresponding to the VLAN interface, and vice versa.*

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 working in the SSM model.

Follow these steps to configure the SSM 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*	Required 232.0.0.0/8 by default

The commands introduced in this section are to be configured on all devices in the PIM domain.

CAUTION:

- *Make sure that the same SSM group range is configured on all devices in the entire PIM-SM domain. Otherwise, multicast information cannot be delivered through the SSM model.*
- *If a multicast group falls in the PIM-SSM range and members of this group send IGMPv1 or IGMPv2 joins, the device that receives these join messages will not trigger (*, G) joins.*

Configuring PIM Common Information

For the configuration tasks 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 performed in interface view is given priority, regardless of the configuration sequence.*

PIM Common Information Configuration Task List

Complete these tasks to configure PIM common information:

Task	Remarks
“Configuring a PIM Filter” on page 588	Optional
“Configuring PIM Hello Options” on page 589	Optional
“Configuring PIM Common Timers” on page 590	Optional

Task	Remarks
"Configuring Join/Prune Message Limits" on page 592	Optional

Configuration Prerequisites

Before configuring PIM common information, 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 PIM common information, prepare the following data:

- An ACL rule as multicast data filter
- 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 PIM Filter

No matter in a PIM-DM domain or a PIM-SM domain, devices 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 devices 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 PIM 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

Generally, a smaller distance from the filter to the multicast source results in a more remarkable filtering effect.

Configuring PIM Hello Options

No matter in a PIM-DM domain or a PIM-SM domain, the hello messages sent among devices 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 devices 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 device has received no hello message from a neighbor, it assumes that this neighbor has expired or become unreachable. You can configure this parameter on all devices in the PIM domain. If you configure different values for this timer on different neighboring devices, the largest value will take effect.
- 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 bit. You can configure this parameter on all devices in the PIM domain. If different LAN-delay or override-interval values result from the negotiation among all the PIM devices, the largest value will take effect.

The LAN-delay setting will cause the upstream devices to delay processing received prune messages. If the LAN-delay setting is too small, it may cause the upstream device to stop forwarding multicast packets before a downstream device sends a prune override message. Therefore, be cautious when configuring this parameter.

The override-interval sets the length of time a downstream device is allowed to wait before sending a prune override message. When a device receives a prune message from a downstream device, it does not perform the prune action immediately; instead, it maintains the current forwarding state for a period of time defined by LAN-delay. If the downstream device 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 LAN-delay timer times out.

A hello message sent from a PIM device 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 device does not change unless the status of the device changes (for example, when PIM is just enabled on the interface or the device is restarted). When the device starts or restarts sending hello messages, it generates a new generation ID. If a PIM device finds that the generation ID in a hello message from the upstream device 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 upstream device will explicitly track which downstream devices are joined to it. The join suppression feature should be enabled or disable on all PIM devices on the same subnet.

Configuring hello options globally

Follow these steps to configure hello options globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
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**	Optional Enabled by default

Configuring hello options on an interface

Follow these steps to configure hello options for an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN/POS 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**	Optional Enabled by default
Configure the interface to reject hello messages without a generation ID	**pim require-genid**	Optional By default, hello messages without Generation_ID are accepted.

Configuring PIM Common Timers

PIM devices discover PIM neighbors and maintain PIM neighboring relationships with other devices by periodically sending out hello messages.

Upon receiving a hello message, a PIM device waits a random period, which is equal to or smaller than the maximum delay between hello messages, before sending out a hello message. This avoids collisions that occur when multiple PIM devices send hello messages simultaneously.

Any device 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.

A PIM device periodically sends join/prune messages to its upstream for state update. A join/prune message contains the join/prune timeout time. The upstream device sets a join/prune timeout timer for each pruned downstream interface, and resumes the forwarding state of the pruned interface when this timer times out.

When a device fails to receive subsequent multicast data from the multicast source S, the device will not immediately deletes the corresponding (S, G) entries; instead, it maintains (S, G) entries for a period of time, namely the multicast source lifetime, before deleting the (S, G) entries.

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
Configure assert timeout time	**holdtime assert** *interval*	Optional 180 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 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 VLAN/POS 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 assert timeout time	**pim holdtime assert** *interval*	Optional 180 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

If there are no special networking requirements, we recommend that you use the default settings.

Configuring Join/Prune Message Limits

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 limits:

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 sent and received PIM control messages	**display pim control-message counters** [**message-type** { **probe** | **register** | **register-stop** } | [**interface** *interface-type interface-number* | **message-type** { **assert** | **bsr** | **crp** | **graf**t | **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 the specified interface or all interfaces	**display pim interface** [*interface-type interface-number*] [**verbose**]	Available in any view
View the information of joint/prune messages to be sent	**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

To do...	Use the command...	Remarks
View the content of the PIM multicast 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
Enable PIM debugging	**debugging pim** { **all** | **event** [*acl-number*] | **routing-table** [*acl-number*] | **neighbor** [*acl-number*] [**receive** | **send**] | **assert** [*acl-number*] [**receive** | **send**] | **rp** [**receive** | **send**] | **join-prune** [*acl-number*] [**receive** | **send**] | **register** [*acl-number*] | **msdp** [*acl-number*] | **state-refresh** [*acl-number*] [**receive** | **send**] }	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

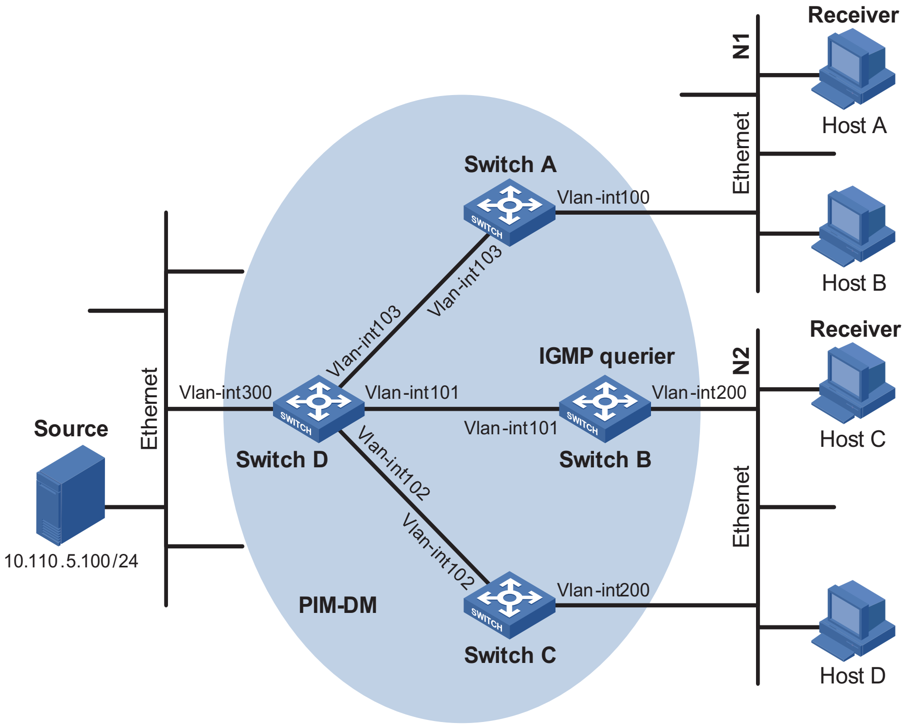
PIM-DM Configuration Example

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.
- As shown in Figure 183, 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.
- IGMPv3 is required on Switch A, Switch B, Switch C, and hosts in N1 and N2. Switch B is the IGMP querier on the multi-access subnet.

Network diagram

Figure 183 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

Only the commands related to the PIM-DM configuration are listed below.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the OSPF protocol for interoperation among the switches in the PIM-DM domain. Ensure the network-layer interoperation among Switch A, Switch B, Switch C and Switch D 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.

Configure the IP address and subnet mask for each interface as per Figure 183. Detailed configuration steps are omitted here.

2 Enable IP multicast routing, and enable PIM-DM on each interface

# Enable IP multicast routing on Switch A, enable PIM-DM on each interface, and enable 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 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 the configuration 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
Vpn-instance: public net
Interface                    NbrCnt HelloInt   DR-Pri     DR-Address
Vlan300                      0      30         1          10.110.5.1
Vlan103                      1      30         1          192.168.1.2
Vlan101                      1      30         1          192.168.2.2
Vlan102                      1      30         1          192.168.3.2
```

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
 Vpn-instance: public net
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 a multicast group G (225.1.1.1/24). 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 registers with Switch A, 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
 Vpn-instance: public net
 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
 Vpn-instance: public net
 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
```

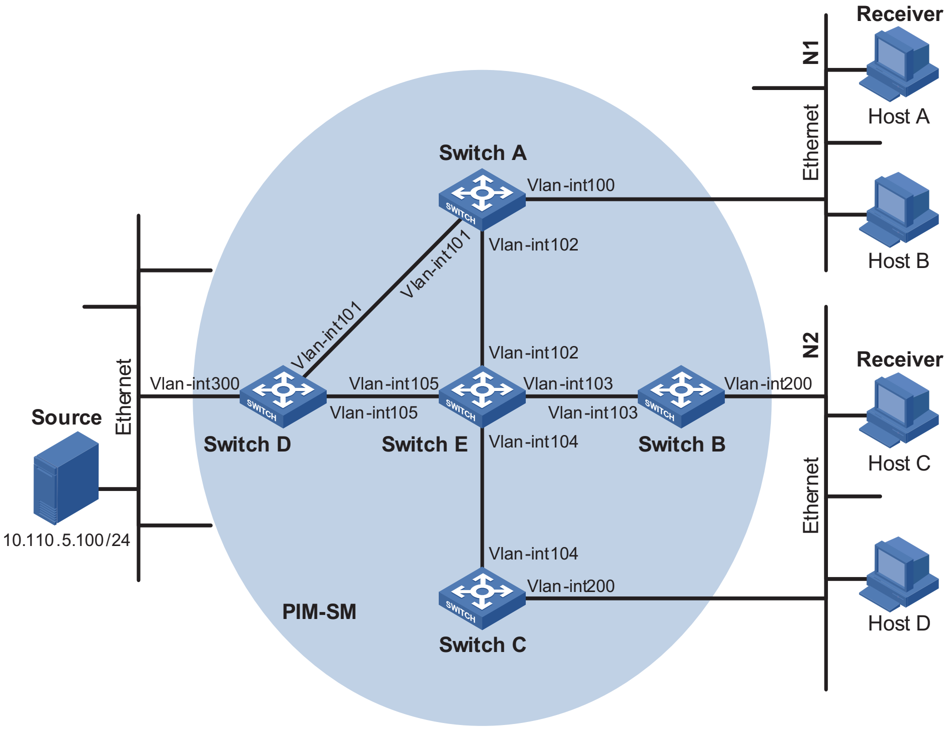
PIM-SM Configuration Example

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 (not divided into different BSR admin-scope regions).
- 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, and its VLAN-interface 102 interface acts a C-BSR and a C-RP, with the range of multicast groups served by the C-RP being 225.1.1.0/24.

Network diagram

Figure 184 Network diagram for PIM-SM domain configuration (on switches)



Device	Interface	IP address	Device	Interface	IP address

Switch A	Vlanint100	10.110.1.1/24	Switch D	Vlanint300	10.110.5.1/24
	Vlanint101	192.168.1.1/24		Vlanint101	192.168.1.2/24
	Vlanint102	192.168.9.1/24		Vlanint105	192.168.4.2/24
Switch B	Vlanint200	10.110.2.1/24	Switch E	Vlanint104	192.168.3.2/24
	Vlanint103	192.168.2.1/24		Vlanint103	192.168.2.2/24
Switch C	Vlanint200	10.110.2.2/24		Vlanint102	192.168.9.2/24
	Vlanint104	192.168.3.1/24		Vlanint105	192.168.4.1/24

Configuration procedure

Only the commands related to the PIM-SM configuration are listed below

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the OSPF protocol for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation among Switch A, Switch B, Switch C, Switch D and Switch E 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.

Configure the IP address and subnet mask for each interface as per Figure 184. Detailed configuration steps are omitted here.

2 Enable IP multicast routing, and enable PIM-SM on each interface

# Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable 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 a C-BSR and a C-RP

# Configure the service scope of RP advertisements and the positions of the C-BSR and C-RP on Switch E.

```
<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
[SwitchE-pim] c-rp vlan-interface 102 group-policy 2005
[SwitchE-pim] return
```

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
Vpn-instance: public net
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
 Vpn-instance: public net
 Elected BSR Address: 192.168.9.2
     Priority: 0
     Hash mask length: 30
     State: Accept Preferred
     Scope: Not scoped
     Uptime: 01:40:40
     Next BSR message scheduled at: 00:01:42
```

# View the BSR information and the locally configured C-RP information in effect on Switch E.

```
<SwitchE> display pim bsr-info
 Vpn-instance: public net
Elected BSR Address: 192.168.9.2
     Priority: 0
     Hash mask length: 30
     State: Elected
     Scope: Not scoped
     Uptime: 00:00:18
     Next BSR message scheduled at: 00:01:52
 Candidate BSR Address: 192.168.9.2
     Priority: 0
     Hash mask length: 30
     State: Pending
     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
 Vpn-instance: public net
 PIM-SM BSR RP information:
 Group/MaskLen: 225.1.1.0/24
     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 (225.1.1.1/24). An RPT will be built between Switch A and Switch E. When the multicast source S (10.110.5.100/24) registers with 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 for example) contain (*, G) and (S, G) entries, while routers on the SPT path (Switch A and Switch D for example) contain 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
 Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

(*, 225.1.1.1), 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: pim-sm, UpTime: 00:13:46, Expires: -
(10.110.5.100, 225.1.1.1), RP: 192.168.9.2
     Protocol: pim-sm, Flag: SPT LOC
     UpTime: 00:00:42
     Upstream interface: Vlan-interface101,
         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: pim-sm, UpTime: 00:00:42, Expires:-
```

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
 Vpn-instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.1), RP: 192.168.9.2
     Protocol: pim-sm, Flag: SPT LOC
     UpTime: 00:00:42
     Upstream interface: Vlan-interface300
         Upstream neighbor: 10.110.5.100,
 RPF prime neighbor: 10.110.5.100
     Downstream interface(s) information:
Total number of downstreams: 1
1: Vlan-interface105
Protocol: pim-sm, UpTime: 00:00:42, Expires:-
```

# View the PIM routing table information on Switch E.

```
<SwitchE> display pim routing-table
 Vpn-instance: public net
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, 225.1.1.1), 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
```

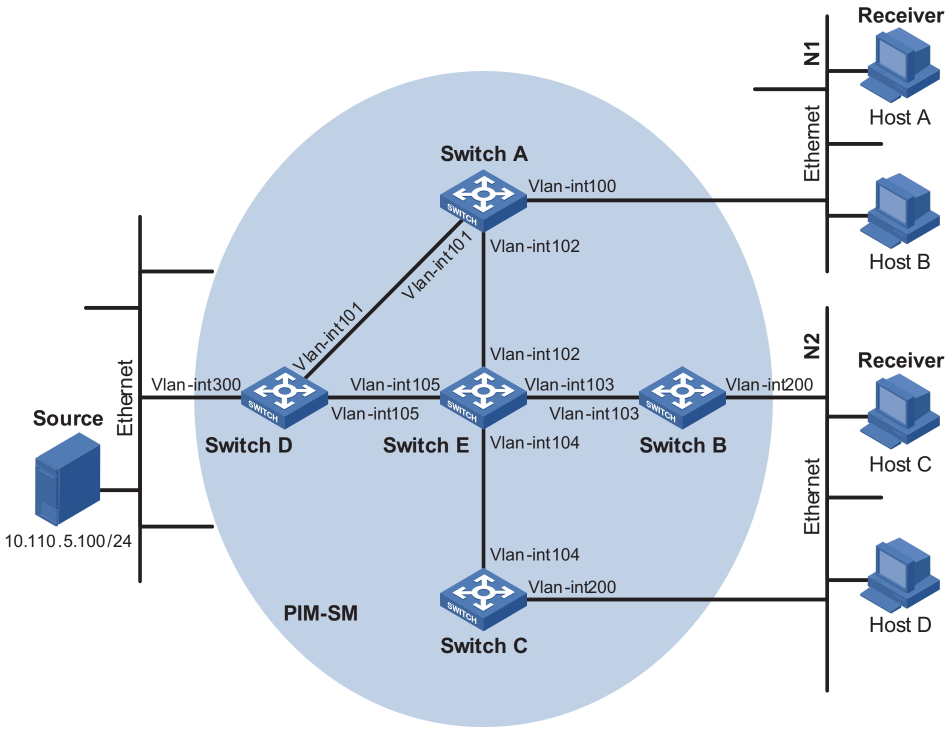
PIM-SSM Configuration Example

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 range of SSM multicast group addresses is 232.1.1.0/24.
- IGMPv3 is required on Switch A, Switch B, Switch C, and hosts in N1 and N2.

Network diagram

Figure 185 Network diagram for PIM-SSM configuration (on switches)



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlanint100	10.110.1.1/24	Switch D	Vlanint300	10.110.5.1/24
	Vlanint101	192.168.1.1/24		Vlanint101	192.168.1.2/24
	Vlanint102	192.168.9.1/24		Vlanint105	192.168.4.2/24
Switch B	Vlanint200	10.110.2.1/24	Switch E	Vlanint104	192.168.3.2/24
	Vlanint103	192.168.2.1/24		Vlanint103	192.168.2.2/24
Switch C	Vlanint200	10.110.2.2/24		Vlanint102	192.168.9.2/24
	Vlanint104	192.168.3.1/24		Vlanint105	192.168.4.1/24

Configuration procedure

Only the commands related to the PIM-SMM configuration are listed below.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the OSPF protocol for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation among Switch A, Switch B, Switch C, Switch D and Switch E 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.

Configure the IP address and subnet mask for each interface as per Figure 185. Detailed configuration steps are omitted here.

2 Enable IP multicast routing, and enabling PIM-SM on each interface

# Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable 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 range of PIM-SSM multicast group addresses

# Configure the range of PIM-SSM multicast group addresses to be 232.1.1.0/24 one Switch A.

```
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule permit ip source 232.1.1.0 0.0.0.255
[SwitchA-acl-basic-2000] quit
[SwitchA] pim
[SwitchA-pim] ssm-policy 2000
[SwitchA-pim] return
```

The configuration on Switch B, Switch C, Switch D and Switch E is similar to the configuration 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
Vpn-instance: public net
Interface             NbrCnt HelloInt   DR-Pri     DR-Address
Vlan100               0      30         1          10.110.1.1      (l
ocal)
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 (232.1.1.1/24). Switch A builds an SPT towards the multicast source. Switches on the SPT path (Switch A and Switch

D for example) generates (S, G) entries, 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
 Vpn-instance: public net
 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:  pim-ssm, UpTime:  00: 13: 25, Expires:  -
```

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
 Vpn-instance: public net
Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 232.1.1.1)
     Protocol:  pim-ssm, Flag:
     UpTime:  00: 12: 05
     Upstream interface:  Vlan-interface300
         Upstream neighbor:  10.110.5.100,
         RPF prime neighbor:  10.110.5.100
     Downstream interface(s) information:
Total number of downstreams:  1
         1:  Vlan-interface105
Protocol:  pim, UpTime:  00: 12: 05, Expires:  00: 03: 25
```

Troubleshooting PIM Configuration

Failure of Building a Multicast Distribution Tree Correctly

Symptom

None of the devices in the network (including devices 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 device connected with the multicast source to the last hop device connected with the clients along the SPT. When the multicast data is flooded to a device, no matter which device is, it creates (S, G) entries only if it has a route to the multicast source. If the device does not have a route to the multicast source, or if PIM-DM is not enabled on the device's RPF interface to the multicast source, the device cannot create (S, G) entries.

- When PIM-SM runs on the entire network, and when a device is to join the SPT, the device creates (S, G) entries only if it has a route to the multicast source. If the device does not have a route to the multicast source, or if PIM-DM is not enabled on the device's RPF interface to the multicast source, the device cannot create (S, G) entries.
- When a multicast device 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 device 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 device 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 device, 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 exist 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 device.**
6. **Check that the same PIM mode is enabled on all the devices in the entire network. Make sure that the same PIM mode is enabled on all the devices: PIM-SM on all devices, or PIM-DM on all devices. In the case of PIM-SM, also check that the BSR and RP configurations are correct.**

Multicast Data Abnormally Terminated on an Intermediate Device

Symptom

An intermediate device can receive multicast data successfully, but the data cannot reach the last hop device. An interface on the intermediate device 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 devices 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 devices 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 device 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 devices.**

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 devices in the entire network.**

No Unicast Route Between BSR and C-RPs 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 device 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 devices 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, the RP and the BSR are available. Carry out the display ip routing-table command to check whether routes are available on each device 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 devices 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 device, 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 devices.**

43 MSDP CONFIGURATION

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

- "MSTP Overview" on page 609
- "Configuring Basic Functions of MSDP" on page 616
- "Configuring an MSDP Peer Connection" on page 618
- "Configuring SA Messages Related Parameters" on page 619
- "Displaying and Maintaining MSDP" on page 622
- "MSDP Configuration Examples" on page 623
- "Troubleshooting MSDP" on page 635

- *The term "router" in this document refers to a router in a generic sense or a Switch 8800 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" on page 563.*

MSTP 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 objective. By establishing MSDP peer relationships among RPs of different PIM-SM domains, source active (SA) messages can be forwarded among domains and the multicast source information can be shared.

- *MSDP is applicable only if the intra-domain multicast protocol is PIM-SM.*
- *MSDP is meaningful only for the any-source multicast (ASM) model.*

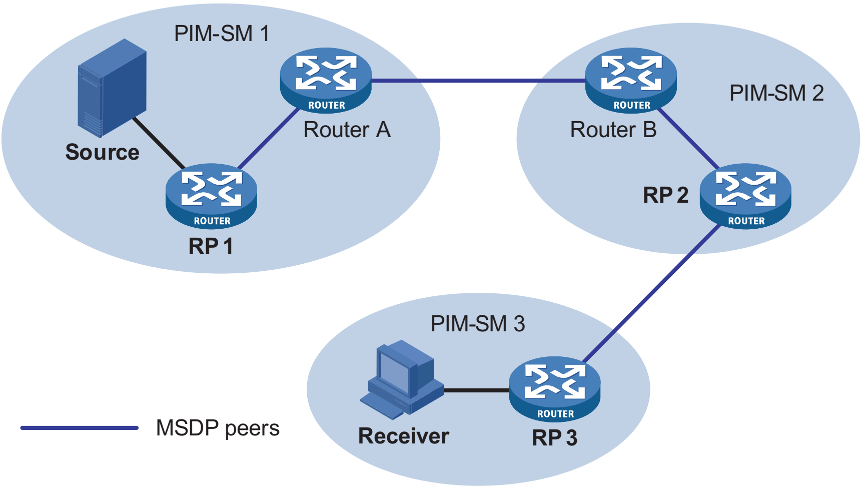
How MSDP Works

MSDP peers

As shown in Figure 186, an active multicast source (Source) exists in the domain PIM-SM 1, and RP 1 has learned the existence of Source through multicast source registration. If receiver hosts in PIM-SM 2 and PIM-SM 3 also wish to receive multicast data from Source, the RPs in these domains can learn the location of Source by establishing MSDP peering relationships. MSDP peering relationships can be established among RPs of different or the same PIM-SM domain, between RPs and multicast routers, or even between multicast routers.

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 RPs in other domains.

Figure 186 Where MSDP peers are in the network



The blue lines in Figure 186 interconnect MSDP peers. 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 in Figure 186. 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 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 source-side RP, like RP 3 in Figure 186. 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 in Figure 186. 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 multicast routers (other than RPs)

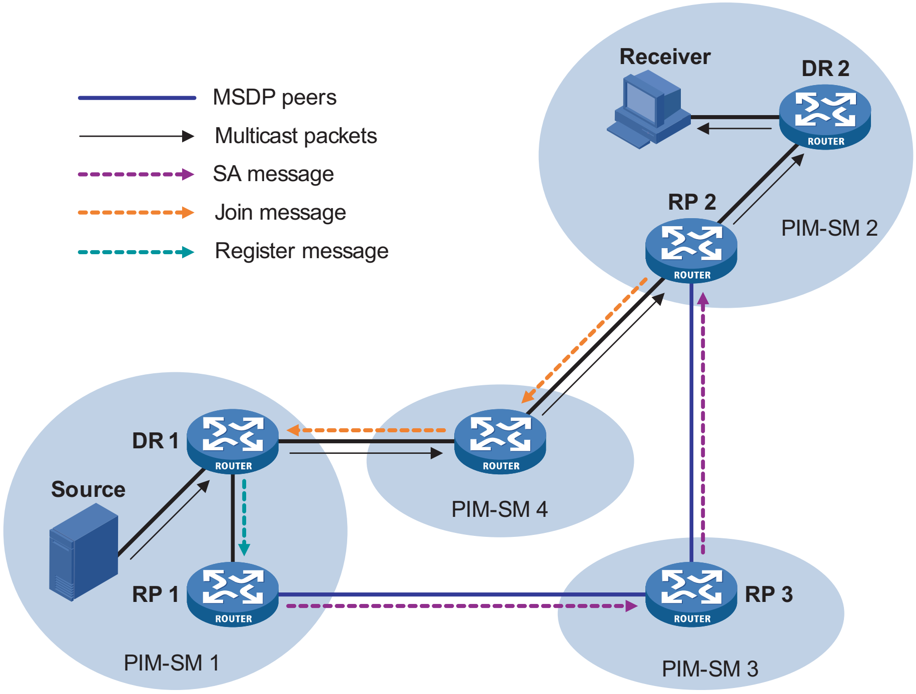
Router A and Router B are MSDP peers on common multicast routers. Such MSDP peers just forward received SA messages.

An 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 looser C-RPs will assume the role of command PIM-SM routers on the "MSDP interconnection map".

Implementing inter-domain multicast delivery by leveraging MSDP peers

As shown in Figure 187, 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 187 MSDP peering relationships



The process if 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 dot create an (S, G) entry and does join the SPT rooted at the source.

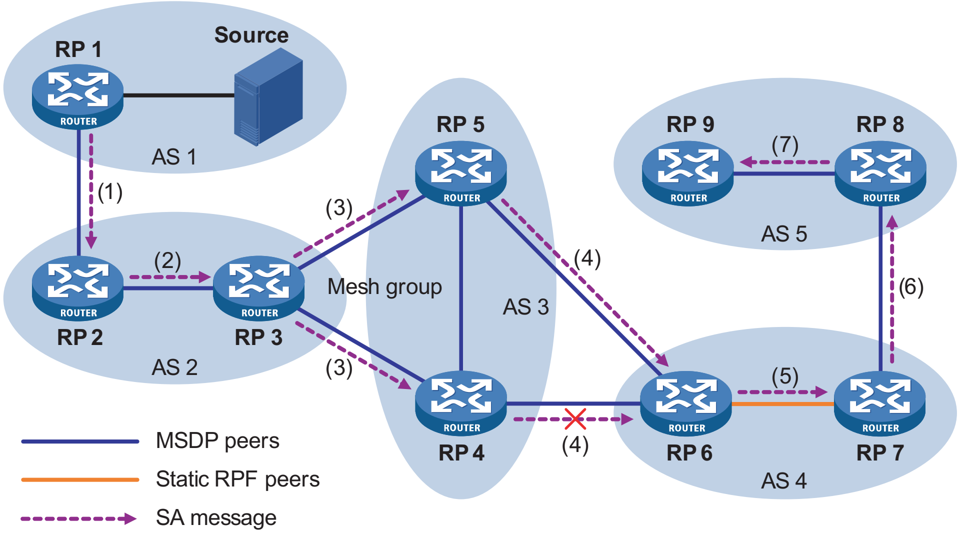
- *An MSDP mesh group refers to a group of MSDP peers that have an MSDP peering relationship among one another and share the same group name.*
- *When using MSDP for inter-domain multicasting, once an RP receives information form a multicast source, it no longer relies on RPs in other PIM-SM domains. 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 188, there are five autonomous systems in the network, AS 1 through AS 5, with IGP enabled on routers within each AS and EBGP 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.

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 188 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 188 Diagram for RPF check for SA messages



As illustrated in Figure 188, 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 SA (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

An EBGP 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 EBGP 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.

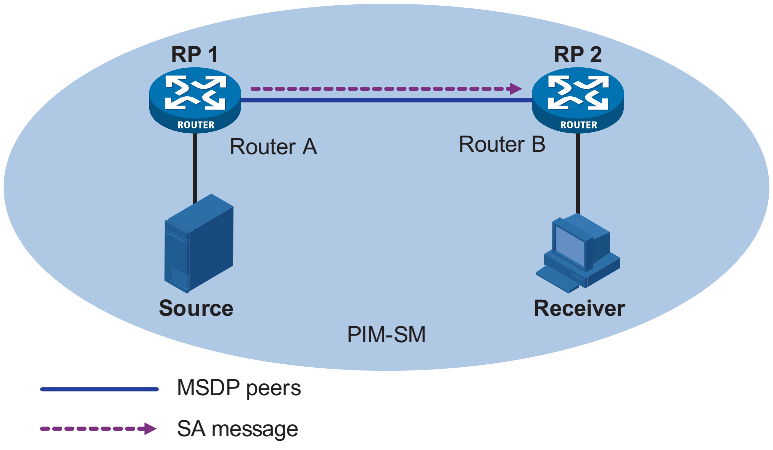
Implementing intra-domain Anycast RP by leveraging MSDP peers

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 189, within the same 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.

Usually an Anycast RP address is configured on a logic interface, like a loopback interface.

Figure 189 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 to 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.

- *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.*

MSDP Related 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" on page 616	"Enabling MSDP" on page 616	Required
	"Creating an MSDP Peer Connection" on page 616	Required
	"Configuring a Static RPF Peer" on page 617	Optional
"Configuring an MSDP Peer Connection" on page 618	"Configuring MSDP Peer Description" on page 618	Optional
	"Configuring an MSDP Mesh Group" on page 618	Optional
	"Configuring MSDP Peer Connection Control" on page 619	Optional

Task		Remarks
"Configuring SA Messages Related Parameters" on page 619	"Configuring SA Message Content" on page 620	Optional
	"Configuring SA Request Messages" on page 620	Optional
	"Configuring an SA Message Filtering Rule" on page 621	Optional
	"Configuring SA Message Cache" on page 622	Optional

Configuring Basic Functions of MSDP

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 the basic functions of 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:

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

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 pairs 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**	-

To do...	Use the command...	Remarks
Create an MSDP peer connection	**peer** *peer-address* **connect-interface** *interface-type interface-number*	Required No MSDP peer connection created by default

If an interface of the device is shared by an MSDP peer and a BGP peer at the same time, you are recommended to configuration the same IP address for the MSDP peer and BGP 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

When configuring multiple static RPF peers on the same device, observe the following rules:

- If you use the **rp-policy** keyword for all the static RPF peers, all the static RPF peers will be activated concurrently. SA messages will be filtered as per the configured prefix list and only those SA messages whose RP addresses pass the filtering will be accepted. If multiple static RPF peers use the same filtering policy at the same time, when a peer receives an SA message, it will forward the SA message to the other peers.
- If you use the **rp-policy** keyword for none of the static RPF peers, according to the configuration sequence, only the first static RPF peer whose connection is in the UP state will be activated, and all SA messages from this peer will be accepted while the SA messages from other static RPF peers will be discarded. When this active static RPF peer fails (for example, when the configuration is removed or when the connection is torn down), based on the configuration sequence, the next RPF peer with its connection in the UP state will be selected as the activated RPF peer.

CAUTION:

- *An MSDP peering connection must be created before static RPF peers can be configured.*
- *If only one MSDP peer is configured on a device, this MSDP peer will act 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 forwarded the message within the mesh group either. This mechanism not only avoids SA flooding but also simplifies the RPF check mechanism, because BGP 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 with these MSDP peers.

Follow these steps to create an MSDP mesh group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MSDP view	**msdp**	-

To do...	Use the command...	Remarks
Create an MSDP peer as a mesh group member	**peer** *peer-address* **mesh-group** *name*	Required An MSDP peer does not belong to any mesh group by default

CAUTION:

- *Before grouping multiple devices into an MSDP mesh group, make sure that these devices are interconnected with one another.*
- *Make sure to configure the same mesh group name on each peer.*
- *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. 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 related parameters, 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 related parameters, prepare the following data:

- ACL as a filtering rule for SA request messages

- ACL as an SA message creation rule
- ACL as a filtering rule for receiving or forwarding SA messages
- Minimum TTL value of multicast packets encapsulated in SA messages
- Maximum SA message cache size

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 device performs RPF check on the message. If the device 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 devices in the same PIM-SM domain, and configure these devices 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 registration 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 device does not send an SA request message to its designated 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 devices 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 sending 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

- *With the function of sending SA request messages enabled, when receiving a Join message from a new multicast receiver, the device sends an SA request message to the remote MSDP peer specified by this command, and the remote peer responds with cached SA information. Upon sending an SA request message, the device receives the information about all the active multicast sources.*
- *If you do not specify an ACL when configuring an SA message filtering rule, all the SA requests sent by the device's MSDP peers will be ignored. If you specify an ACL, SA request messages complying with the filtering rule will be accepted, while all other SA request messages will be ignored.*

CAUTION: *Before you can enable the device to send SA requests, be sure to disable the SA message cache mechanism.*

Configuring an SA Message Filtering Rule

By configuring an SA message creation rule, you can enable the device 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.

In addition to controlling SA message creation, you can also configure filtering rules for forwarding and receiving SA messages, so as to control the propagation of multicast source information in the SA messages.

- By configuring a filtering rule for receiving or forwarding SA messages, you can enable the device 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.
- An SA message with encapsulated multicast data can be forwarded to a designated MSDP peer only if the TTL value in its IP header exceeds the threshold. Therefore, you can control the forwarding of such an SA message by configuring the TTL threshold of the encapsulated data packet.

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**	-
Configuring an SA message creation rule	**import-source** [**acl** *acl-number*]	Required No restrictions on (S, G) entries by default

To do...	Use the command...	Remarks
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 minimum TTL value of multicast packets to be encapsulated in SA messages	**peer** *peer-address* **minimum-ttl** *ttl-value*	Optional 0 by default

Configuring SA Message Cache

To reduce the time spent in obtaining the multicast source information, you can have SA messaged cached on the device. However, the more SA messages are cached, the larger memory space of the device is used.

With the SA cache mechanism enabled, when receiving a new Join message, the device will not send an SA request message to its MSDP peer; instead, it acts as follows:

- If there is no SA message in the cache, the device will wait for the SA message sent by its MSDP peer in the next cycle;
- If there is an SA message in the cache, the device will obtain the information of all active sources directly from the SA message and join the corresponding SPT.

To protect the device against denial of service (DoS) attacks, you can configure the maximum number of SA messages the device 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 message cache mechanism	**cache-sa-enable**	Required Enabled by default
Configure the maximum number of SA messages the router can cache	**peer** *peer-address* **sa-cache-maximum** *sa-limit*	Required 8192 by default

Displaying and Maintaining MSDP

To do...	Use the command...	Remarks
View the brief information of MSDP peers status	**display msdp brief**	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 MSDP cache	**display msdp sa-cache** [*group-address* | *source-address* | *as-number*] *	Available in any view
View the number of SA messages in the MSDP cache	**display msdp sa-count** [*as-number*]	Available in any view
Reset the TCP connection with an MSDP peer	**reset msdp pee**r [*peer-address*]	Available in user view

To do...	Use the command...	Remarks
Clear (S, G) entries in the MSDP 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

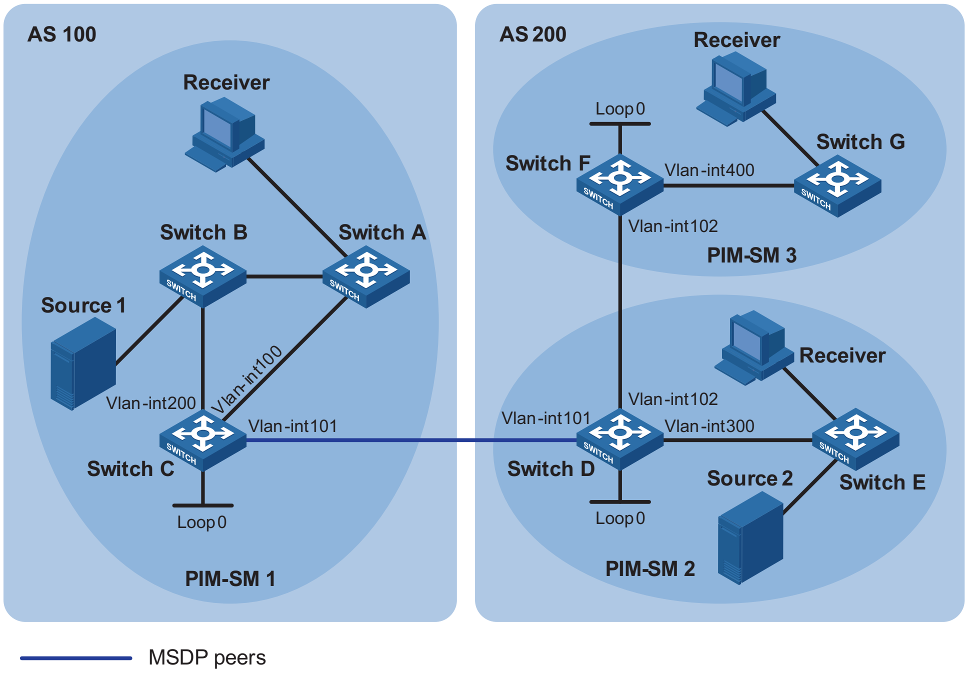
Example of Configuration Leveraging BGP Routes

Network requirements

- Two ISPs maintains their ASs, 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 one or more receivers. OSPF runs within each domain to provide unicast routes.
- MSDP peering relationships need to be established among RPs of different PIM-SM domains leveraging BGP routes.
- The respective Loopback 0 interfaces of Switch C, Switch D and Switch F are configured as the C-BSR and C-RP of the respective PIM-SM domains.
- An MSDP peering relationship needs to be established between Switch C and Switch D through EBGP, and an MSDP peering relationship needs to be established between Switch D and Switch F through IBGP.

Network diagram

Figure 190 Network diagram for configuration leveraging a BGP route



Device	Interface	IP address	Device	Interface	IP address

Switch C	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.4.1/24
	Vlan-int200	10.110.2.1/24		Vlan-int102	192.168.3.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int101	192.168.1.2/24
	Loop0	1.1.1.1/32		Loop0	2.2.2.2/32
Switch F	Vlan-int400	10.110.3.1/24			
	Vlan-int102	192.168.3.2/24			
	Loop0	3.3.3.3/32			

Configuration procedure

Only the commands related to the MSDP configuration leveraging a BGP route are listed in this example.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure OSPF for interconnection between devices in each PIM-SM domain. Ensure the network-layer interoperation among Switch A, Switch B and Switch C in PIM-SM 1, the network-layer interoperation between Switch D and Switch E in PIM-SM 2, and the network-layer interoperation between Switch F and Switch G in PIM-SM 3, and ensure the dynamic update of routing information between the devices in each PIM-SM domain through a unicast routing protocol. Detailed configuration steps are omitted.

Configure the IP address and subnet mask for each interface as per Figure 190. Detailed configuration steps are omitted.

2 Enable IP multicast routing, and enable PIM-SM on each interface

# Enable IP multicast routing on Switch C, and enable PIM-SM on each interface.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] pim sm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] pim sm
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim sm
```

The configuration on Switch A, Switch B, Switch D, Switch E, Switch F and Switch G is similar to the configuration on Switch C.

# Configure BSR boundary on Switch C.

```
[SwitchC-Vlan-interface101] pim bsr-boundary
[SwitchC-Vlan-interface101] quit
```

The configuration on Switch D and Switch F is similar to the configuration on Switch C.

3 Configure Loopback 0 and the position of C-BSR, and C-RP

# Configure the position of Loopback0, C-BSR, and C-RP on Switch C.

```
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 1.1.1.1 255.255.255.255
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
[SwitchC] pim
[SwitchC-pim] c-bsr loopback 0
[SwitchC-pim] c-rp loopback 0
[SwitchC-pim] quit
```

The configuration on Switch D and Switch F is similar to the configuration on Switch C.

4 Configure inter-AS BGP and configure mutual route redistribution between BGP and OSPF

# Configure EBGP on Switch C, and import OSPF routes.

```
[SwitchC] bgp 100
[SwitchC-bgp] router-id 1.1.1.1
[SwitchC-bgp] peer 192.168.1.2 as-number 200
[SwitchC-bgp] import-route ospf 1
[SwitchC-bgp] quit
```

# Configure IBGP and EBGP on Switch D, and import OSPF routes.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp] router-id 2.2.2.2
[SwitchD-bgp] peer 192.168.1.1 as-number 100
[SwitchD-bgp] peer 192.168.3.2 as-number 200
[SwitchD-bgp] import-route ospf 1
[SwitchD-bgp] quit
```

# Configure IBGP on Switch F, and import OSPF routes.

```
<SwitchF> system-view
[SwitchF] bgp 200
[SwitchF-bgp] router-id 3.3.3.3
[SwitchF-bgp] peer 192.168.3.1 as-number 200
[SwitchF-bgp] import-route ospf 1
[SwitchF-bgp] quit
```

# Import BGP routing information into OSPF on Switch C.

```
[SwitchC] ospf 1
[SwitchC-ospf-1] import-route bgp
[SwitchC-ospf-1] quit
```

The configuration on Switch D and Switch F is similar to the configuration on Switch C.

Use the **display bgp peer** command to view the BGP peering relationships between the switches. For example:

# View the information about BGP peering relationship on Switch C.

```
[SwitchC] 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 relationship on Switch D.

```
[SwitchD] 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 F.

```
[SwitchF] 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
```

5 Configure MSDP peers

# Configure an MSDP peer on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchC-msdp] quit
```

# Configure an MSDP peer on Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchD-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchD-msdp] quit
```

# Configure MSDP peers on Switch F.

```
[SwitchF] msdp
[SwitchF-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchF-msdp] quit
```

When the multicast source (Source 1) sends multicast information, receivers in PIM-SM 2 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 relationship 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
  192.168.1.2        Up        00:12:27       200     13         0
```

# View the brief information about MSDP peering relationship on Switch D.

```
[SwitchD] display msdp brief
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2            0            0            0            0
MSDP Peer Brief Information
  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 F.

```
[SwitchF] 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 C.

```
[SwitchC] 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
```

Example of Anycast RP Application Configuration

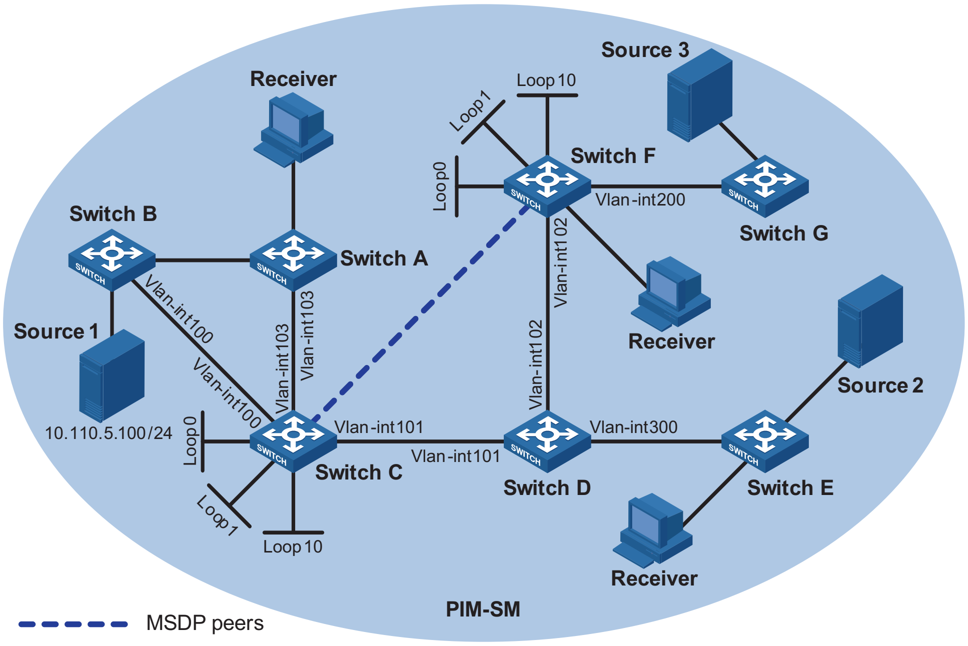
Network requirements

- The PIM-SM domain in this example has multiple multicast sources and receivers. OSPF runs within the domain to provide unicast routes.
- The anycast RP application is configured in the PIM-SM domain. When a new member joins the multicast group, the switch directly connected to receivers can initiate a Join message to the topologically nearest RP.
- An MSDP peering relationship is set up between Switch C and Switch F.

- On Switch C and Switch F, the interface Loopback 1 is configured as a C-BSR, and Loopback 10 is configured as a C-RP.
- The router ID of Switch C is 1.1.1.1, while the router ID of Switch F is 2.2.2.2.

Network diagram

Figure 191 Network diagram for anycast RP configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int103	10.110.1.2/24	Switch D	Vlan-int300	10.110.4.1/24
Switch B	Vlan-int100	10.110.2.2/24		Vlan-int102	192.168.3.1/24
Switch C	Vlan-int103	10.110.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int100	10.110.2.1/24	Switch F	Vlan-int200	10.110.3.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int102	192.168.3.2/24
	Loop0	1.1.1.1/32		Loop0	2.2.2.2/32
	Loop1	3.3.3.3/32		Loop1	4.4.4.4/32
	Loop10	10.1.1.1/32		Loop10	10.1.1.1/32

Configuration procedure

Only the commands related to the configuration of Anycast RP application are listed in this example.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the IP address and subnet mask for each interface as per Figure 191. Detailed configuration steps are omitted.

Configure OSPF for interconnection between the switches. Detailed configuration steps are omitted.

2 Enable IP multicast routing, and enable PIM-SM on each interface

# Enable IP multicast routing on Switch C, and enable PIM-SM on each interface.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 103
[SwitchC-Vlan-interface103] pim sm
[SwitchC-Vlan-interface103] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] pim sm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface Vlan-interface 101
[SwitchC-Vlan-interface101] pim sm
[SwitchC-Vlan-interface101] quit
```

The configuration on Switch A, Switch B, Switch D, Switch E, Switch F and Switch G is similar to the configuration on Switch C.

3 Configure the position of interface Loopback 1, Loopback 10, C-BSR, and C-RP.

# Configure different Loopback 1 addresses and identical Loopback 10 address on Switch C and Switch F, configure C-BSR on each Loopback 1 and configure C-RP on each Loopback 10.

```
[SwitchC] interface loopback 1
[SwitchC-LoopBack1] ip address 3.3.3.3 255.255.255.255
[SwitchC-LoopBack1] pim sm
[SwitchC-LoopBack1] quit
[SwitchC] interface loopback 10
[SwitchC-LoopBack10] ip address 10.1.1.1 255.255.255.255
[SwitchC-LoopBack10] pim sm
[SwitchC-LoopBack10] quit
[SwitchC] pim
[SwitchC-pim] c-bsr loopback 1
[SwitchC-pim] c-rp loopback 10
[SwitchC-pim] quit
```

The configuration on Switch F is similar to the configuration on Switch C.

To view the PIM routing information on the switches, use the **display pim routing-table** command. When the multicast source (Source 1, with the address of 10.110.5.100/24) in the PIM-SM domain sends multicast data to the multicast group G (225.1.1.1/24), the receivers attached to Switch F can receive the multicast data. By comparing the PIM routing information displayed on Switch C with that displayed on Switch F, you can see that Switch C acts now as the RP.

# View the PIM routing information on Switch C.

```
[SwitchC] display pim routing-table
Vpn-instance: public net
Total 0 (*, G) entry; 1 (S, G) entry

 (10.110.5.100, 225.1.1.1),
 RP: 10.1.1.1 (local)
     Protocol: pim-sm, Flag: SPT LOC ACT
     UpTime: 00:10:20
     Upstream interface: Vlan-interface100
         Upstream neighbor: 10.110.2.2
         RPF prime neighbor: 10.110.1.2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface101
             Protocol: pim-sm, UpTime: 00:10:20, Expires: 00:03:10
```

# View the PIM routing information on Switch F.

```
[SwitchF] display pim routing-table
Vpn-instance: public net
Total 0 (*, G) entry; 1 (S, G) entry

(10.110.5.100, 225.1.1.1)
 RP: 10.1.1.1
     Protocol: pim-sm, Flag: SPT ACT
     UpTime: 00:03:32
     Upstream interface: Vlan-interface102
         Upstream neighbor: 192.168.3.1
         RPF prime neighbor: 192.168.3.1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface200
             Protocol: pim-sm, UpTime: 00:03:32, Expires: -
```

4 Configure Loopback0 and MSDP peers

# Configure an MSDP peer on Loopback0 of Switch C.

```
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 1.1.1.1 255.255.255.255
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
[SwitchC] msdp
[SwitchC-msdp] originating-rp loopback 0
[SwitchC-msdp] peer 2.2.2.2 connect-interface loopback 0
[SwitchC-msdp] quit
```

# Configure an MSDP peer on Loopback0 of Switch F.

```
<SwitchF> system-view
[SwitchF] interface loopback 0
[SwitchF-LoopBack0] ip address 2.2.2.2 255.255.255.255
[SwitchF-LoopBack0] pim sm
[SwitchF-LoopBack0] quit
[SwitchF] msdp
[SwitchF-msdp] originating-rp loopback 0
[SwitchF-msdp] peer 1.1.1.1 connect-interface loopback 0
[SwitchF-msdp] quit
```

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 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
  2.2.2.2            Up        00:10:17        ?      0          0
```

# View the brief MSDP peer information on Switch F.

```
[SwitchF] 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
```

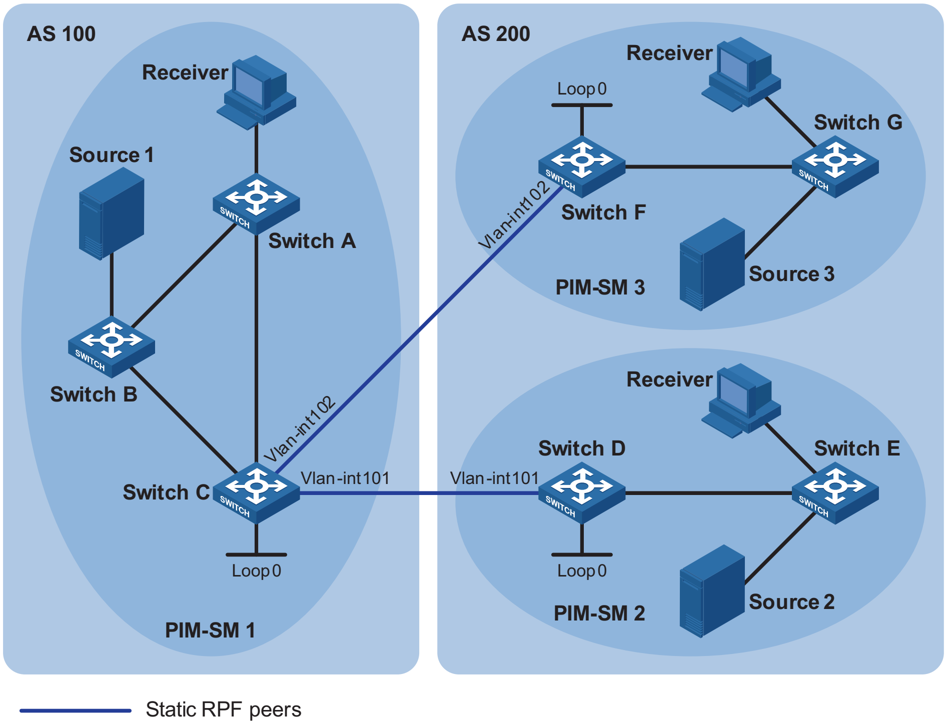
Static RPF Peer Configuration Example

Network requirements

- Two ISPs maintains their ASs, 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 one or more receivers. OSPF runs within each domain to provide unicast routes.
- PIM-SM 2 and PIM-SM 3 are both PIM stub domains, and BGP is not required between these two domains and PIM-SM 1. Instead, static RPF peers are configured to avoid RPF check on SA messages.
- The respective loopback interfaces of Switch C, Switch D and Switch F are configured as the C-BSR and C-RP of the respective PIM-SM domains.
- The static RPF peers of Switch C are Switch D and Switch F, while Switch C is the only RPF peer of Switch D and Switch F. Any switch can receive SA messages sent by its static RPF peer(s) and permitted by the corresponding filtering policy.

Network diagram

Figure 192 Network diagram for static RPF peer configuration



Device	Interface	IP address	Device	Interface	IP address
Switch D	Vlan-int101	192.168.1.2/24	Switch C	Vlan-int101	192.168.1.1/24
	Loop0	2.2.2.2/32		Vlan-int102	192.168.3.1/24
Switch F	Vlan-int102	192.168.3.2/24		Loop0	1.1.1.1/32
	Loop0	3.3.3.3/32			

Configuration procedure

Only the commands related to the MSDP configuration for static RPF peering connections in PIM stub domains are listed in this example.

1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the IP address and subnet mask for each interface of each switch as per Figure 192.

Configure OSPF for interconnection between the switches. Ensure the network-layer interoperation among Switch A, Switch B and Switch C in PIM-SM 1, the network-layer interoperation between Switch D and Switch E in PIM-SM 2, and the network-layer interoperation between Switch F and Switch G in PIM-SM 3, and ensure the dynamic update of routing information between the switches in each PIM-SM domain through a unicast routing protocol. Detailed configuration steps are omitted.

Configure EBGP among Switch C, Switch D, Switch C and Switch F, and configure mutual route redistribution between BGP and OSPF. Detailed configuration steps are omitted.

Configure the IP address and subnet mask for each interface as per Figure 192. Detailed configuration steps are omitted.

2 Enable IP multicast routing, and enable PIM-SM on each interface

# Enable IP multicast routing on Switch C, and enable PIM-SM on each interface.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim sm
[SwitchC-Vlan-interface101] quit
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim sm
```

The configuration on Switch A, Switch B, Switch D, Switch E, Switch F and Switch G is similar to the configuration on Switch C.

# Configure BSR boundary on Switch C.

```
[SwitchC-Vlan-interface102] pim bsr-boundary
[SwitchC-Vlan-interface102] quit
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim bsr-boundary
[SwitchC-Vlan-interface101] quit
```

The configuration on Switch D and Switch F is similar to the configuration on Switch C.

3 Configure the position of interface Loopback0, C-BSR, and C-RP.

# Configure the position of Loopback0, C-BSR, and C-RP on Switch C.

```
[SwitchC] router-id 1.1.1.1
[SwitchC] interface loopback 0
[SwitchC-LoopBack0] ip address 1.1.1.1 255.255.255.255
[SwitchC-LoopBack0] pim sm
[SwitchC-LoopBack0] quit
[SwitchC] pim
[SwitchC-pim] c-bsr loopback 0
[SwitchC-pim] c-rp loopback 0
[SwitchC-pim] quit
```

The configuration on Switch D and Switch F is similar to the configuration on Switch C.

4 Configure static RPF peers

# Configure Switch D and Switch F as MSDP peers and static RPF peers of Switch C.

```
[SwitchC] ip ip-prefix list-df permit 192.168.0.0 16 greater-equal 1
6 less-equal 32
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchC-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchC-msdp] static-rpf-peer 192.168.3.1 rp-policy list-df
```

```
[SwitchC-msdp] static-rpf-peer 192.168.1.2 rp-policy list-df
[SwitchC-msdp] quit
```

# Configure Switch C as MSDP peer and static RPF peer of Switch D.

```
<SwitchD> system-view
[SwitchD] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16
 less-equal 32
[SwitchD] msdp
[SwitchD-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchD-msdp] static-rpf-peer 192.168.3.1 rp-policy list-c
[SwitchD-msdp] quit
```

# Configure Switch C as MSDP peer and static RPF peer of Switch F.

```
<SwitchF> system-view
[SwitchF] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16
 less-equal 32
[SwitchF] msdp
[SwitchF-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchF-msdp] static-rpf-peer 192.168.3.2 rp-policy list-c
[SwitchF-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 (Source 1) in PIM-SM 1 sends multicast information, receivers in PIM-SM 2 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 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        01:07:08       ?        8           0
  192.168.1.2        UP        00:16:39       ?        13          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
  192.168.1.1        UP        01:07:09       ?        8           0
```

# View the brief MSDP peer information on Switch F.

```
[SwitchF] 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
```

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 switch.
- 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 devices. Carry out the **display ip routing-table** command to check whether the unicast route between the switches is correct.
2. Check that a unicast route is available between the two devices 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 switch are the same.

No SA Entries in the Device's SA Cache

Symptom

MSDP fails to send (S, G) entries through SA messages.

Analysis

- The **import-source** command is used 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 correct configured.

Solution

1. Check that a route is available between the devices. Carry out the **display ip routing-table** command to check whether the unicast route between the devices is correct.
2. Check that a unicast route is available between the two devices 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 devices. Carry out the **display ip routing-table** command to check whether the unicast route between the devices is correct.
2. Check that a unicast route is available between the two devices 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.

44 MLD CONFIGURATION

The term "router" in this document refers to a router in a generic sense or a Switch 8800 running the MLD protocol.

When configuring MLD, go to the following sections for information you are interested in:

MLD Overview

Introduction to MLD

Multicast listener discovery protocol (MLD) is used by an IPv6 router or a routing switch to discover the presence of multicast listeners on directly-attached subnets. Multicast listeners are nodes wishing to receive multicast packets.

Through MLD, the router can learn whether there are any IPv6 multicast listeners on 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 Version

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.

At present, the Switch 8800s support only MLDv1.

How MLDv1 Works

MLDv1 implements IPv6 multicast listener management based on the query/response mechanism.

MLDv1 uses two types of query messages:

- General query: an IPv6 multicast router or routing switch sends periodical general queries to determine what IPv6 multicast addresses have active listeners on the local subnet.
- Multicast-address-specific query: an IPv6 multicast router or routing switch sends multicast-address-specific queries to determine whether any listeners for particular IPv6 multicast addresses exist on the local subnet.

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.

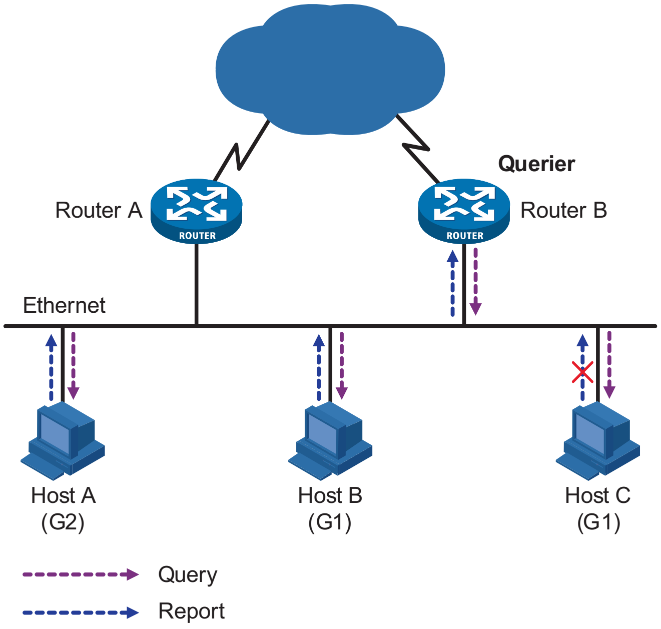
1 Initially, every MLD router assumes itself as the querier and sends MLD general query messages (often referred to as general queries) to all hosts and routers on the local subnet (the destination address is FF02::1).

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 193 How hosts use MLD to join IPv6 multicast groups



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 193. The basic process that the hosts join the IPv6 multicast groups is as follows:

1 The MLD querier (Router B in the figure) periodically multicasts MLD queries (with the destination address of FF02::1) to all hosts and routers on the local subnet.

2 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 interest in G1. Assume it is Host B that sends the report message.

3 Host C, which is on the same subnet, hears the report from Host B for joining G1. Upon hearing the report, Host C will suppress itself from sending a report message for the same IPv6 multicast group, because the MLD routers (Router A and Router B) already know that at least one host on the local subnet is interested in G1.

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/response process, the MLD routers learn that members of G1 and G2 are attached to the local subnet, and generate (*, G1) and (*, G2) multicast forwarding entries through an IPv6 multicast routing protocol (such as IPv6 PIM), which will be the basis for subsequent IPv6 multicast forwarding, where * represents any 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 a done message to all IPv6 multicast routers (the destination address is FF02::2) on the local subnet.

2 Upon receiving the leave 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 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.

MLD Message Types

The MLD querier learns the multicast listening states of neighbor interfaces by sending MLD query messages. Figure 194 shows the format of an MLDv1 query message.

Figure 194 Format of MLDv1 query message

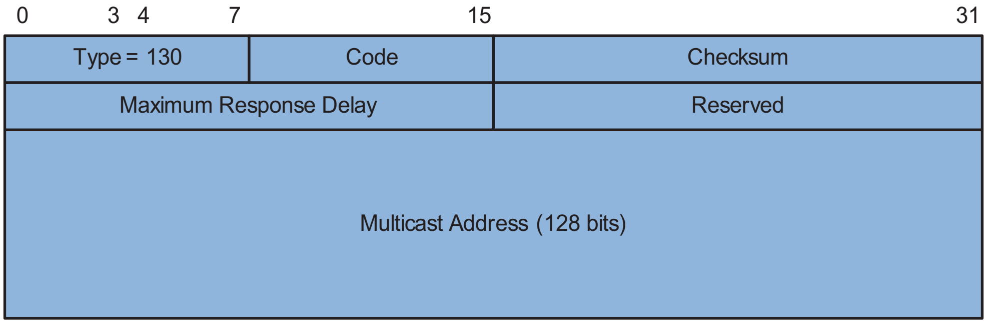


Table 29 describes the fields in Figure 194.

Table 29 Description on fields in an MLD query message

Field	Description
Type	Message type. 130 stands for query message; 131 stands for report message; 132 for leave group message.
Code	Code: initialized to 0 by the sender and ignored by receivers
Checksum	Standard IPv6 checksum
Maximum Response Delay	Maximum response delay allowed before a host sends a report message
Reserved	Reserved field: initialized to 0 by the sender and ignored by receivers
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. ■ It is sent to the IPv6 multicast address that the message sender joins in or leaves

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

Configuration Task List

Task		Remarks
"Configuring Basic Functions of MLD" on page 641	"Enabling MLD" on page 641	Required
	"Configuring the MLD Version" on page 641	Option
"Adjusting MLD Performance" on page 642	"Configuring MLD Message Options" on page 642	Optional
	"Configuring MLD Timers" on page 643	Optional

- *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.

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 VLAN interface view	**interface** *interface-type interface-number*	-
Enable MLD	**mld enable**	Required Disabled by default

CAUTION:

- *MLD must be enabled on the receiver-side DR before hosts can join IPv6 multicast groups. For details of a DR, refer to "IPv6 PIM Configuration" on page 671.*
- *After MLD is enabled on a VLAN interface, it is not allowed to enable MLD Snooping in the corresponding VLAN, and vice versa.*

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. At present, the Switch 8800s support only MLDv1.

Configure an MLD version globally

Follow these steps to configure the MLD version globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MLD view	**mld**	-

To do...	Use the command...	Remarks
Configure the MLD version globally	**version** *version-number*	Optional MLDv1 by default

Configure an MLD version on an interface

Follow these steps to configure the MLD version on an interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure the MLD version on the VLAN interface	**mld version** *version-number*	Optional MLDv1 by default

Adjusting MLD Performance

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:

- Query interval
- Maximum response delay of MLD general query messages
- Other querier present interval
- Interval for and last listener query count

Configuring MLD Message Options

MLD involves multicast-address-specific queries, which are specific to particular IPv6 multicast groups, yet IPv6 multicast groups change dynamically, and a device cannot join all IPv6 multicast 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.

Depending on whether an MLD message carries the Router-Alert option in the IP header, the device processes the message differently. For details about the Router-Alert option, refer to RFC 2113.

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 contain the Router-Alert option.

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. In this case, when the device receives an MLD message, it checks it for the Router-Alert option, and discards it if it does not carry the Router-Alert option.

Configure 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.

Configure 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 VLAN 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 Timers

The MLD querier periodically sends MLD general query messages to decide 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 query (general query or multicast-address-specific query) message, a host starts a timers 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 message from the querier before the other querier present timer expires, it will assume that the querier has failed and will initiate a new querier election process. Otherwise, the non-querier will reset its timeout time.

Configure MLD timers globally

Follow these steps to configure MLD timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MLD view	**mld**	-
Configure the query interval	**timer query** *interval*	Optional 125 seconds by default.
Configure the maximum response delay for MLD general query messages	**max-response-time** *interval*	Optional 10 seconds by default
Configure the last listener query interval	**lastlistener-queryinterval** *interval*	Optional 1 second by default
Configure the last listener query count	**robust-count** *robust-value*	Optional 2 times by default
Configure the other querier present interval	**timer other-querier-present** *interval*	Optional For the default, see Note below.

Configure MLD timers on an interface

Follow these steps to configure MLD timers on an interface

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure the query interval	**mld timer query** *interval*	Optional 125 seconds by default.

To do...	Use the command...	Remarks
Configure the maximum response delay for MLD general query messages	**mld max-response-time** *interval*	Optional 10 seconds by default
Configure the last listener query interval	**mld lastlistener-queryinterval** *interval*	Optional 1 second by default
Configure the last listener query count	**mld robust-count** *robust-value*	Optional 2 times by default
Configure the other querier present interval	**mld timer other-querier-present** *interval*	Optional For the default, see Note below.

About other querier present interval:

- *If not configured manually, the other querier present interval is determined by the formula: [Other querier present interval (in seconds)] = [MLD query interval] times [robustness variable] plus [maximum response delay] divided by two. By default, the values of these three parameters are 125, 2 and 10, respectively, so the default other querier present interval = 125 x 2 + 10 / 2 = 255 (seconds).*
- *If manually configured, the other querier present interval takes the configured value.*

CAUTION:

- *If the configured other querier present interval is less than the query interval, the state of the MLD querier may frequently change.*
- *Make sure that the maximum response delay of MLD general query messages is less than the last listener query interval (namely, the maximum response delay of MLD multicast-address-specific query messages). Otherwise, multicast group members may be wrongly removed.*

Displaying and Maintaining MLD Configuration

To do...	Use the command...
Display information of IPv6 multicast groups	**display mld group** [*ipv6-group-address* | **interface** *interface-type interface-number*] [**static** | **verbose**]
Display MLD layer 2 port information	**display mld group port-info** [**vlan** *vlan-id*] [**slot** *slot-id*] [**verbose**]
Display MLD configuration and running information on the specified interface or all MLD-enabled interfaces	**display mld interface** [*interface-type interface-number*] [**verbose**]
Display the information of the MLD routing table	**display mld routing-table** [*ipv6-source-address* [*prefix-length*] | *ipv6-group-address* [*prefix-length*]] *
Clear MLD forwarding entries	**reset mld group** { **all** | **interface** *interface-type interface-number* { **all** | *ipv6-group-address* [*prefix-length*] [*ipv6-source-address* [*prefix-length*]] } }

CAUTION: *The* ***reset mld group*** *command cause an interruption of receivers' reception of multicast data.*

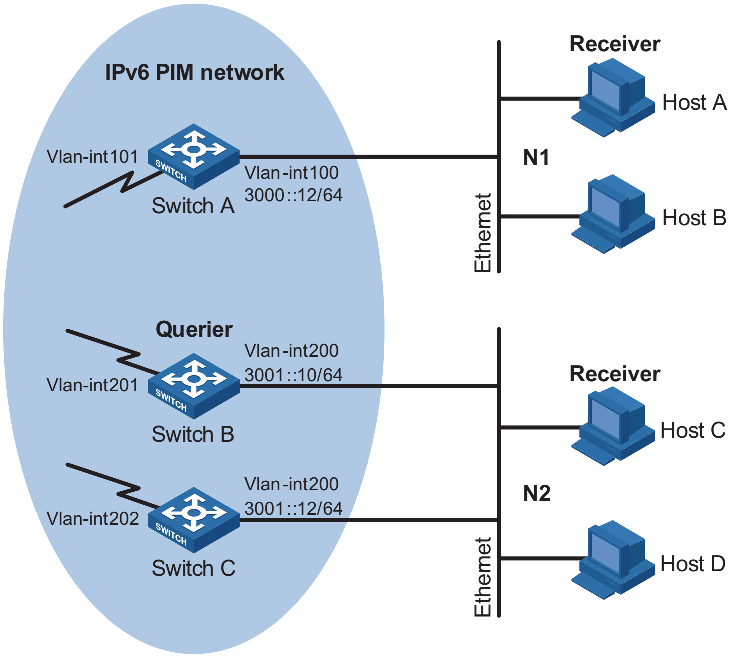
MLD Configuration Example

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-DM 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 IPv6 PIM network through VLAN-interface 201 and VLAN-interface 202.
- MLDv1 is required between Switch A and N1, and between the other two Switches (Switch B and Switch C) and N2, with Switch B as the MLD querier.

Network diagram

Figure 195 Network diagram for MLD configuration



Configuration procedure

In the configuration procedure, only the commands related to the MLD configuration are listed.

1 Configure IPv6 addresses for the router interfaces and an IPv6 unicast protocol.

Configure an IPv6 address and prefix length for each interface as shown in Figure 195.

Configure OSPFv3 for interoperation between the switches. Ensure the network-layer interoperation between Switch A, Switch B, and Switch C on the IPv6 PIM-DM 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 MLD on the host interfaces.

# Enable IPv6 multicast routing on Switch A, enable MLD and IPv6 PIM-DM on VLAN-interface 100, and set the MLD version number to 1.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] mld enable
[SwitchA-Vlan-interface100] mld version 1
[SwitchA-Vlan-interface100] pim ipv6 dm
[SwitchA-Vlan-interface100] quit
```

# Enable IPv6 multicast routing on Switch B, enable MLD and IPv6 PIM-DM on VLAN-interface 200, and set the MLD version number to 1.

```
<SwitchB> system-view
[SwitchB] multicast ipv6 routing-enable
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] mld enable
[SwitchB-Vlan-interface200] mld version 1
[SwitchB-Vlan-interface200] pim ipv6 dm
[SwitchB-Vlan-interface200] quit
```

# Enable IPv6 multicast routing on Switch C, enable MLD and IPv6 PIM-DM on VLAN-interface 200, and set the MLD version number to 1.

```
<SwitchC> system-view
[SwitchC] multicast ipv6 routing-enable
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] mld enable
[SwitchC-Vlan-interface200] mld version 1
[SwitchC-Vlan-interface200] pim ipv6 dm
[SwitchC-Vlan-interface200] quit
```

3 Verify the configuration.

Carry out the **display mld interface** command to display the MLD configuration and running information on each router interface. Example:

# Display MLD information on Ethernet 1/0 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)
```

Troubleshooting MLD

No Member Information on the Receiver-Side DR

Symptom

When a host sends a message for joining IPv6 multicast group G, there is no member information of multicast group G on the receiver-side DR.

Analysis

- The correctness of networking and interface connections directly affects the generation of IPv6 group member information.
- IPv6 multicast routing must be enabled on the device.

Solution

1. Check that the networking is correct and that interface connections are correct.
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 that the interface is normal and that a correct IPv6 address has been configured. Carry out the **display mld interface** command to display the interface information. If no interface information is output, the interface is abnormal. Typically this is because the **shutdown** command has been executed on the interface, or the interface connection is incorrect, or no correct IPv6 address has been configured on the interface.

Inconsistent Memberships on Routers on the Same Subnet

Symptom

Different memberships are maintained on different MLD routers or route switches on the same subnet.

Analysis

- A router or routing switch 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 or routing switches on the same subnet will surely result in inconsistent MLD memberships.
- If the Switch 8800s are working with other devices (third-party routers or networking devices, for example) on the same network, there can be an issue of inconsistent MLD versions. Two MLD versions are currently available. The Switch 8800s only support MLDv1. Although routers or routing switches running different MLD versions are compatible with hosts, all devices on the same subnet must run the same MLD version. Inconsistent MLD versions running on routers routing switches on the same subnet will also lead to chaos of 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 or routing switches on the same subnet to check the MLD timers for consistent configurations.
3. Use the **display mld interface** command to check that the routers or routing switches are running the same MLD version.

45 MLD SNOOPING CONFIGURATION

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

For details about MLD and IPv6 PIM, refer to "MLD Configuration" on page 637 and "IPv6 PIM Configuration" on page 671.

MLD Snooping Overview

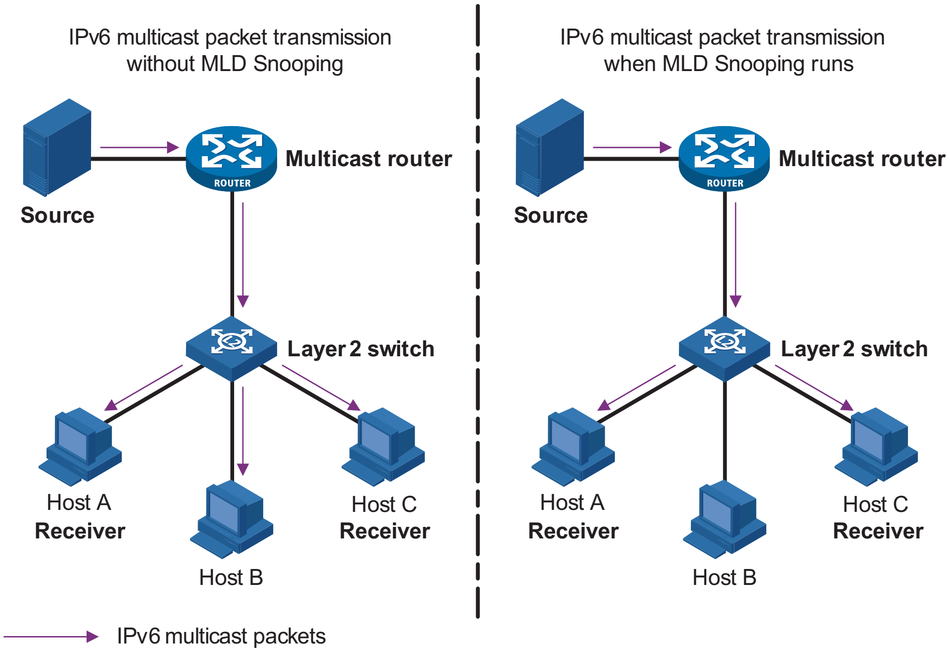
Multicast Listener Discovery Snooping (MLD Snooping) is an IPv6 multicast constraining mechanism that runs on Layer 2 devices to manage and control IPv6 multicast groups.

How MLD Snooping Works

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 196, when MLD Snooping is not running on the switch (Layer 2 device), IPv6 multicast packets are broadcast to all Layer 2 ports. When MLD Snooping runs, multicast packets for known IPv6 multicast groups are forwarded to only those Layer 2 ports with receivers attached to them.

Figure 196 Before and after MLD Snooping is enabled on a Layer 2 device

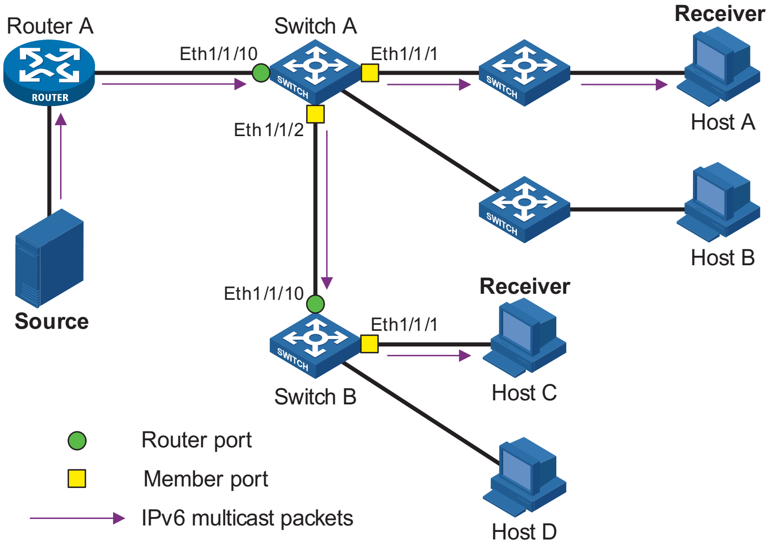


Basic Concepts in MLD Snooping

MLD Snooping related ports

As shown in Figure 197, 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 197 MLD Snooping related ports



Ports involved in MLD Snooping, as shown in Figure 197, are described as follows:

- Router port: A router port is a port on a Layer 2 switch that leads the switch to a multicast router (Layer-3 multicast device) or the MLD querier on the subnet. In the figure, Ethernet 1/1/10 of Switch A and Ethernet 1/1/10 of Switch B are router ports. A 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 or Listener Port) is a port on a Layer 2 switch that leads the switch to an IPv6 multicast group member. In the figure, Ethernet 1/1/1 and Ethernet 1/1/2 of Switch A and Ethernet1/1/1 of Switch B are member ports. The switch records all member ports on the local device in the MLD Snooping forwarding table.

- *Whenever mentioned in this document, a router port is a router-side port on a switch, rather than a port on a router.*
- *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 router ports. For details about IPv6 PIM hello messages, see "Configuring IPv6 PIM Hello Options" on page 693.*

Port aging timers in MLD Snooping

Table 30 Port aging timers in MLD Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Router port aging timer	For each router port, the switch sets a timer initialized to the aging time of the route port	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
Member port aging timer	When a port joins an IPv6 multicast group, the switch sets a timer for the port, which is initialized to the member port aging time	MLD report message	The switch removes this port from the IPv6 multicast group forwarding table

Work Mechanism of MLD Snooping

A switch running MLD Snooping performs different actions when it receives different MLD messages, as follows:

General queries

The MLD querier periodically sends MLD general queries to all hosts and routers 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 receiving port and performs the following to the receiving port:

- If the receiving port is a router port existing in its router port list, the switch resets the aging timer of this router port.
- If the receiving port is not a router port existing in its router port list, the switch adds it into its router port list and sets an aging timer for this router port.

Membership reports

A host sends an MLD report to the multicast router 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 multicast router 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 IPv6 multicast group the host is interested in, and performs the following to the receiving port:

- If the port is already in the IPv6 forwarding table, the switch resets the member port aging timer of the port.
- If the port is not in the IPv6 forwarding table, the switch installs an entry for this port in the IPv6 forwarding table and starts the member port aging timer of this port.

A switch will not forward an MLD report through a non-router port for the following reason: With MLD report suppression enabled, if member hosts of that IPv6 multicast group still exist under other non-router ports, these hosts will stop sending MLD reports when they receive the message. This prevents the switch from knowing if members of that IPv6 multicast group are still attached to these ports.

At present, the Switch 8800 supports only MLDv1 messages.

Done messages

When a host leaves an IPv6 multicast group, the host sends an MLD done message to the multicast router to announce that it is to leave the IPv6 multicast group.

Upon receiving an MLD done message, a switch forwards it through all router ports in the VLAN. Because the switch does not know whether any other member hosts of that IPv6 multicast group still exists under the port to which the MLD done message arrived, the switch does not immediately delete the forwarding entry corresponding to that port from the forwarding table; instead, it resets the aging timer of the member port.

Upon receiving an MLD done message from a host, the MLD querier resolves from the message the address of the IPv6 multicast group that the host just left and sends an MLD group-specific query to that IPv6 multicast group through the port that received the done message. Upon receiving the MLD group-specific query, a switch forwards it through all the router ports in the VLAN and all member ports of that IPv6 multicast group, and performs the following to the receiving port:

- If an MLD report from that IPv6 multicast group arrives to this member port before its aging timer expires as a response to the MLD group-specific query, this means that some other members of that IPv6 multicast group still exist under the port: the switch resets the aging timer of the member port.
- If no MLD report from that IPv6 multicast group arrives to this member port before its aging timer expires as a response to the MLD group-specific query, this means that no members of that IPv6 multicast group still exist under the member port: the switch deletes the forwarding entry for the member port from the forwarding table when its aging timer expires.

Processing of IPv6 Multicast Protocol Messages

Under different conditions, an MLD Snooping-capable switch processes IPv6 multicast protocol messages differently, 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 router port.
3. When MLD is disabled on the switch, or when MLD forwarding entries are cleared (by using the **reset mld group** command):
   - If IPv6 PIM is disabled, the switch clears all its Layer 2 multicast entries and router ports.
   - If IPv6 PIM is enabled, the switch clears only its Layer 2 multicast entries without deleting its router ports.
4. When IPv6 PIM is disabled on the switch:
   - If MLD is disabled, the switch clears all its router ports.
   - If MLD is enabled, the switch maintains all its Layer 2 multicast entries and router ports.

MLD Snooping Configuration Task List

Complete these tasks to configure MLD Snooping:

Task		Remarks
"Configuring Basic Functions of MLD Snooping" on page 654	"Enabling MLD Snooping" on page 654	Required
	"Configuring Port Aging Timers" on page 655	Optional
"Configuring MLD Snooping Port Functions" on page 656	"Configuring Static Ports" on page 656	Optional
	"Configuring Simulated Joining" on page 656	Optional
	"Configuring the Fast Leave Feature" on page 657	Optional
	"Configuring MLD Report Suppression" on page 658	Optional
"Configuring MLD-Related Functions" on page 659	"Enabling MLD Snooping Querier" on page 659	Optional
	"Configuring MLD Timers" on page 659	Optional
	"Configuring a Source IPv6 Address for MLD Queries" on page 661	Optional
	"Configuring the Function of Dropping Unknown IPv6 Multicast Data" on page 661	Optional

Task		Remarks
"Configuring an IPv6 Multicast Group Policy" on page 661	"Configuring an IPv6 Multicast Group Filter" on page 662	Optional
	"Configuring Maximum Multicast Groups that Can Pass Ports" on page 663	Optional
	"Configuring IPv6 Multicast Group Replacement" on page 663	Optional

- *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 interface view are effective only for the current interface; 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 interface view or port group view.*

Configuring Basic Functions of MLD Snooping

Configuration Prerequisites

Before configuring the basic functions of MLD Snooping, complete the following tasks:

- Configure the corresponding VLANs
- Configure the corresponding port groups

Before configuring the basic functions of MLD Snooping, prepare the following data:

- The version of MLD Snooping
- Aging time of router ports
- Aging timer of member ports

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*	-

To do...	Use the command...	Remarks
Enable MLD Snooping in the VLAN	**mld-snooping enable**	Required Disabled by default

- *MLD Snooping must be enabled globally before it can be enabled in a VLAN.*
- *When you enable MLD Snooping in a specified VLAN, this function takes effect for Ethernet ports in this VLAN only.*
- *After enabling MLD Snooping in a VLAN, you cannot enable MLD and/or IPv6 PIM on the corresponding VLAN interface, and vice versa.*

Configuring Port Aging Timers

If the switch does not receive an MLD general query or an IPv6 PIM hello message before the aging timer of a router port expires, the switch deletes this port from the router port list when the aging timer times out.

If the switch does not receive an MLD report for an IPv6 multicast group before the aging timer of a member port expires, the switch deletes this port from the forwarding table for that IPv6 multicast group when the aging timers times out.

If IPv6 multicast group memberships change frequently, you can set a relatively small value for the member port aging timer, and vice versa.

Configuring port aging timers globally

Follow these steps to configure port aging timers globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MLD Snooping view	**mld-snooping**	-
Configure router port aging time	**router-aging-time** *interval*	Optional 260 seconds by default
Configure member port aging time	**host-aging-time** *interval*	Optional 260 seconds by default

Configuring port aging timers in a VLAN

Follow these steps to configure port aging timers in a VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure router port aging time	**mld-snooping router-aging-time** *interval*	Optional 260 seconds by default
Configure member port aging time	**mld-snooping host-aging-time** *interval*	Optional 260 seconds by default

Configuring MLD Snooping Port Functions

Configuration Prerequisites

Before configuring MLD Snooping port functions, complete the following task:

- Enable MLD Snooping in the VLAN or enable MLD on the desired VLAN interface

Before configuring MLD Snooping port functions, prepare the following data:

- IPv6 multicast source addresses
- Whether to enable the fast leave function or not
- Whether to enable the MLD membership report suppression function

Configuring Static Ports

If the host attached to a port is interested in the IPv6 multicast data addressed to a particular IPv6 multicast group, you can configure this port to be a static member port for that IPv6 multicast group.

In a network where the topology structure is unlikely to change, you can configure one or more static router ports on a switch. When receiving IPv6 multicast data, it forwards the data to both multicast group member ports and static router ports.

Follow these steps to configure static ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure a static member port		**mld-snooping static-group** *ipv6-group-address* **vlan** *vlan-id*	Required Disabled by default
Configure a static router port		**mld-snooping static-router-port vlan** *vlan-id*	Required Disabled by default

- *When you configure a port as a static IPv6 multicast group member port, the port does not send an MLD report; when you remove a port as a static IPv6 multicast group member port, it does not send an MLD done message.*
- *Static member ports and static router ports never age out. To delete such a port, you need to use the corresponding command.*

Configuring Simulated Joining

Generally, a host running MLD responds to MLD queries from a multicast router. 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, namely configure a port of the switch as a simulated member of the IPv6 multicast group. When an MLD query arrives, that member port will give a response. As a result, the switch can continue receiving IPv6 multicast data.

Through this configuration, the following functions can be achieved:

- When an Ethernet port is configured as a simulated member host, it sends an MLD report.
- When receiving an MLD general query, the simulated host responds with an MLD report just like a real host.
- When the simulated joining function is disabled on an Ethernet port, the simulated host sends an MLD done message.

Follow these steps to configure simulated joining:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the port(s) to join an IPv6 multicast group as simulated host(s)		**mld-snooping host-join** *ipv6-group-address* **vlan** *vlan-id*	Required Simulated joining is disabled by default

- *Each simulated host is equivalent to an independent host. For example, when receiving an MLD query, the simulated host corresponds to each configuration responds respectively.*

Configuring the Fast Leave Feature

By default, when receiving a group-specific MLD done message on a port, the switch sends a MLD group-specific query message to that port rather than directly deleting the port from the multicast forwarding table. If the switch receives no MLD reports within a certain period of time, it deletes the port from the forwarding table.

With the fast leave feature enabled, when the switch receives a group-specific MLD done message on a port, the switch directly deletes this port from the forwarding table without first sending an MLD group-specific query to the port. If only one host is attached to the port, enable the fast leave feature to improve bandwidth and resource usage.

Configuring the fast leave feature globally

Follow these steps to configure the fast leave feature globally:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MLD Snooping view	**mld-snooping**	-
Enable the fast leave feature	**fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

Configuring the fast leave feature on a port or a group ports

Follow these steps to configure the fast leave feature on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable the fast leave feature		**mld-snooping fast-leave** [**vlan** *vlan-list*]	Required Disabled by default

CAUTION: *If the fast leave feature 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 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 MLD-Related Functions

Configuration Prerequisites

Before configuring MLD-related functions, complete the following task:

- Enable MLD Snooping in the VLAN

Before configuring MLD-related functions, 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
- Source IPv6 address of MLD group-specific queries
- Whether to enable the function of dropping the unknown IPv6 multicast data

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 general queries, thus allowing multicast forwarding entries to be established and maintained at the data link layer.

Follow these steps to enable MLD Snooping querier:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Enable the MLD Snooping querier	**mld-snooping querier**	Required Disabled by default

CAUTION:

- *An MLD Snooping querier does not take part in MLD querier election.*
- *It is meaningless to configure an MLD Snooping querier in an IPv6 multicast network running MLD; in fact, this may affect MLD querier elections because an MLD Snooping querier sends MLD general queries with a low source IPv6 address.*

Configuring MLD Timers

You can tune the MLD general query interval based on actual condition of the network.

Upon receiving an MLD query (general query or group-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 corresponding timers expires simultaneously.

- For MLD general queries, you can configure the maximum response time to fill their Max Response time field.
- For MLD group-specific queries, you can configure the MLD last-member query interval to fill their Max Response time field. Namely, for MLD group-specific queries, the maximum response time equals to the MLD last-member query interval.

Configuring MLD timers globally

Follow these steps to configure MLD timers 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 timers in a VLAN

Follow these steps to configure MLD timers in a VLAN

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN view	**vlan** *vlan-id*	-
Configure MLD general query interval	**mld-snooping query-interval** *interval*	Optional 60 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: *In the configuration, make sure that the MLD general query interval is larger than the maximum response time for MLD general queries.*

Configuring a Source IPv6 Address for MLD Queries

This configuration allows you to change the source IPv6 address of MLD queries. When a port receives an MLD general query with an all-zero IPv6 address, the switch does not put it in its router port list. In a multicast network with only Layer 2 devices, therefore, it is recommended to configure a normal link-local IPv6 address as the source address of MLD query messages.

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 group-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 the Function of Dropping Unknown IPv6 Multicast Data

Unknown IPv6 multicast data refers to IPv6 multicast data whose forwarding entries do not exist in the corresponding multicast forwarding table:

- With the function of dropping unknown IPv6 multicast data enabled, the switch drops the 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 configure globally the function of dropping unknown IPv6 multicast data:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter MLD Snooping view	**mld-snooping**	-
Enable the function of dropping unknown IPv6 multicast data	**drop-unknown**	Required Disabled by default

When enabled to drop unknown IPv6 multicast data, a Switch 8800 still forwards unknown IPv6 multicast data to other router ports in the VLAN.

Configuring an IPv6 Multicast Group Policy

Configuration Prerequisites

Before configuring an IPv6 multicast group filtering policy, complete the following tasks:

- Enable MLD Snooping in the VLAN or enable MLD on the desired VLAN interface

Before configuring an IPv6 multicast group filtering 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
- Whether enable the IPv6 multicast group replacement function.

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 different users' access.

In an actual application of Video on Demand (VoD), 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 ACL rule configured on the receiving port. If this receiving port can join this IPv6 multicast group, the switch adds this port to the MLD Snooping multicast group list; 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, namely hosts can join any IPv6 multicast group

Configuring an IPv6 multicast group filter on a port or a group ports

Follow these steps to configure an IPv6 multicast group filer on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure an IPv6 multicast group filter		**mld-snooping group-policy** *acl6-number* [**vlan** *vlan-list*]	Required No IPv6 filter configured by default, namely hosts can join any IPv6 multicast group

Configuring Maximum Multicast Groups that Can Pass Ports

By configuring the maximum number of IPv6 multicast groups that can pass a port or a group of ports, you can limit the maximum number of multicast programs available to users, thus to control the traffic on the port.

Follow these steps configure the maximum number of IPv6 multicast groups that can pass a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure the maximum number of IPv6 multicast groups that can pass the port(s)		**mld-snooping group-limit** *limit* [**vlan** *vlan-list*]	Optional 1,024 by default.

- *When the number of IPv6 multicast groups a port has joined reaches the maximum number configured, the system deletes this port from all the related MLD Snooping forwarding entries, and hosts on this port need to join IPv6 multicast groups again.*
- *If you have configured a port to be as static member port or enabled the function of simulating a member host on a port, the system deletes this port from all the related MLD Snooping forwarding entries and applies the new configurations, until the number of IPv6 multicast groups the has joined reaches the maximum number configured.*

Configuring IPv6 Multicast Group Replacement

For some special reasons, the number of IPv6 multicast groups passing through a switch or Ethernet 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. To address this situation, you can enable the IPv6 multicast group replacement function on the switch or certain Ethernet ports. When the number of IPv6 multicast groups a switch or an Ethernet 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**	-
Configure IPv6 multicast group replacement	**overflow-replace** [**vlan** *vlan-list*]	Required Disabled by default

Configuring IPv6 multicast group replacement on a port or a group ports

Follow these steps to configure IPv6 multicast group replacement on a port or a group ports:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter the corresponding view	Enter Ethernet interface view	**interface** *interface-type interface-number*	Use either command
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure 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 within the range of 1 to 1023 before configuring 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 the information of IPv6 multicast groups learned by MLD Snooping	**display mld-snooping group** [**vlan** *vlan-id*] [**slot** *slot-id*] [**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 entries	**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

- *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 be used to clear MLD Snooping entries of static joins.*

MLD Snooping Configuration Examples

Examples 1 (Simulated Joining)

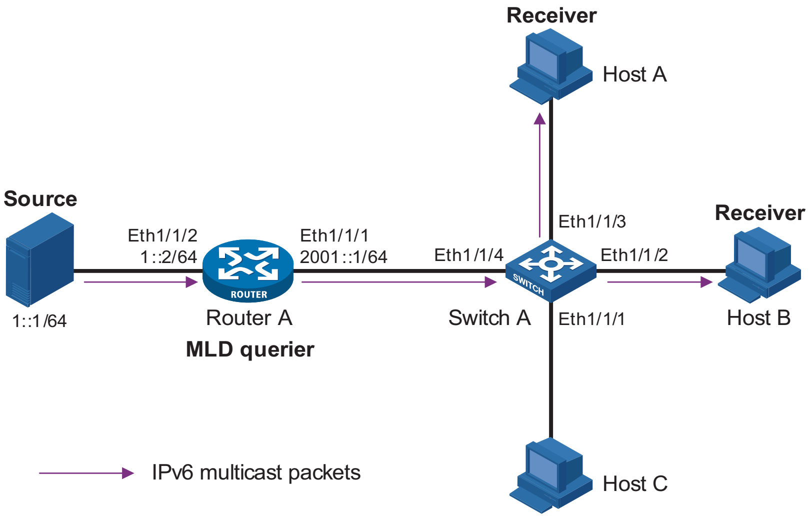
Network requirements

- As shown in Figure 198, Router A (an IPv6 multicast router) connects to the IPv6 multicast source through Ethernet1/1/2 and to Switch A (a Switch 8800) through Ethernet1/1/1.

- MLD runs between Router A and Switch A, MLD Snooping runs on Switch A, and Router A acts as the MLD querier.
- Router A runs IPv6 PIM-SM, and the Ethernet1/1/2 serves as C-BSR and C-RP.
- Perform the following configuration so that multicast data (1::1, FE1E::101:101) can be forwarded through Ethernet1/2 and Ethernet1/3 even if the receivers Host A and Host B attached to on Switch A temporarily stop receiving IPv6 multicast data for some unexpected reasons.

Network diagram

Figure 198 Network diagram for simulated joining configuration



Configuration procedure

1 Configure the IPv6 address of each interface

Configure an IPv6 address and address prefix for each interface. The detailed configuration steps are omitted.

2 Configure Router A

# Enable IPv6 multicast routing, enable IPv6 PIM-SM on each interface, and enable MLDv1 on Ethernet1/1/1. The Ethernet1/1/2 serves as C-BSR and C-RP.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface ethernet 1/1/1
[RouterA-Ethernet1/1/1] mld enable
[RouterA-Ethernet1/1/1] mld version 1
[RouterA-Ethernet1/1/1] pim ipv6 sm
[RouterA-Ethernet1/1/1] quit
[RouterA] interface ethernet 1/1/2
[RouterA-Ethernet1/1/2] pim ipv6 sm
[RouterA-Ethernet1/1/2] quit
[RouterA] pim ipv6
[RouterA-pim6] c-bsr 1::2
[RouterA-pim6] c-rp 1::2
[RouterA-pim6] quit
```

The above configuration on Router A is for reference only. Refer to the specific situation of your device when performing the configuration.

3 Configure Switch A

# Create VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
```

# Add ports Ethernet1/1/1 through Ethernet1/1/4 to VLAN 100.

```
[SwitchA-vlan100] port ethernet 1/1/1 to ethernet 1/1/4
[SwitchA-vlan100] quit
```

# Enable MLD Snooping.

```
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
[SwitchA] vlan 100
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] quit
```

# Enable the function of simulated hosts to join the IPv6 multicast group on the ports: Ethernet1/1/2 and Ethernet1/1/3.

```
[SwitchA] interface ethernet 1/1/2
[SwitchA-Ethernet1/1/2] mld-snooping host-join ff1e::101:101 vlan 100
[SwitchA-Ethernet1/1/2] quit
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/1/3] mld-snooping host-join ff1e::101:101 vlan 100
[SwitchA-Ethernet1/1/3] quit
```

4 Verify the configuration

# View the detailed information of the IPv6 multicast group in VLAN 100.

```
[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, A-Aggregation 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.
            Ethernet1/1/4                                     (D) ( 00:01:30 )
    IP group(s):the following ip group(s) match to one mac group.
      IP group address:FF1E::101:101
        (::, FF1E::101:101):
          Attribute:    Host Port
          Host port(s):total 2 port.
            Ethernet1/1/2                                     (D) ( 00:03:23 )
            Ethernet1/1/3                                     (D) ( 00:03:23 )
    MAC group(s):
      MAC group address:3333-0101-0101
          Host port(s):total 2 port.
            Ethernet1/1/2
            Ethernet1/1/3
```

As shown above, Ethernet 1/1/2 and Ethernet 1/1/3 of Switch A have joined the IPv6 multicast group FF1E::101:101.

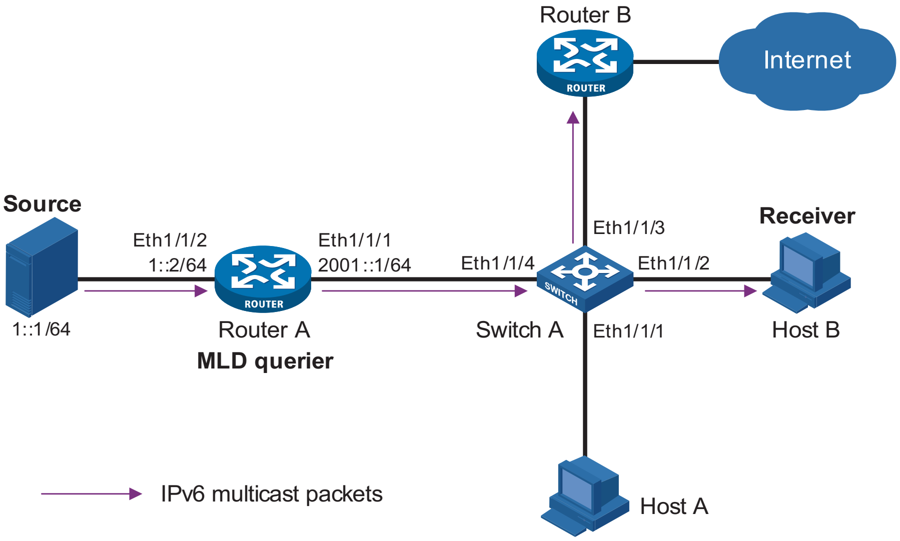
Examples 2 (Static Router Port Configuration)

Network requirements

As shown in Figure 199, Router A, which acts as the MLD querier on the subnet, connects to the IPv6 multicast source through Ethernet1/1/2 and to Switch A (a Switch 8800) through Ethernet1/1/1. While no IPv6 multicast protocol is running on Router B, perform the following configuration so that Switch A can forward all the received IPv6 multicast data to Router B.

Network diagram

Figure 199 Network diagram for static router port configuration



Configuration procedure

1. Configure the IPv6 address of each interface

   Configure an IP address and address prefix for each interface. The specific configuration procedure is omitted here.

2. Configure Router A

   # Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLDv1 on Ethernet1/1/1.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface ethernet 1/1/1
[RouterA-Ethernet1/1/1] mld enable
[RouterA-Ethernet1/1/1] mld version 1
[RouterA-Ethernet1/1/1] pim ipv6 dm
[RouterA-Ethernet1/1/1] quit
[RouterA] interface ethernet 1/1/2
[RouterA-Ethernet1/1/2] pim ipv6 dm
[RouterA-Ethernet1/1/2] quit
```

The above configuration on Router A is for reference only. Refer to the specific situation of your device when performing the configuration.

3. Configure Switch A

   # Create VLAN 100.

```
<SwitchA> system-view
[SwitchA] vlan 100
```

# Add the ports, Ethernet1/1/1 to Ethernet1/1/4, to VLAN 100.

```
[SwitchA-vlan100] port ethernet 1/1/1 to ethernet 1/1/4
[SwitchA-vlan100] quit
```

# Enable MLD Snooping.

```
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
[SwitchA] vlan 100
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] quit
```

# Configure Ethernet1/1/3 to be a static router port.

```
[SwitchA] interface ethernet 1/1/3
[SwitchA-Ethernet1/3] mld-snooping static-router-port vlan 100
[SwitchA-Ethernet1/3] quit
```

Switch A can forward all the received IPv6 multicast data in VLAN 100 to Router B through Ethernet1/1/3, which has been configured as a static router port.

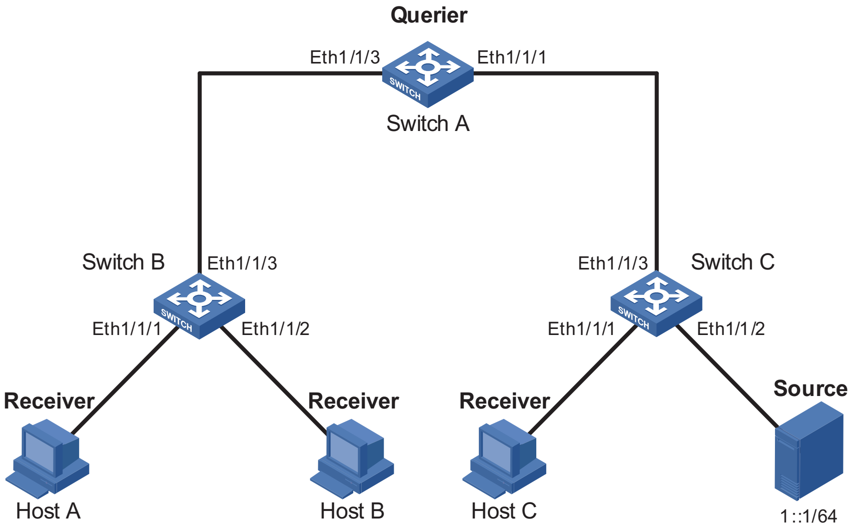
Examples 3 (MLD Snooping Querier Configuration)

Network requirements

- As shown in Figure 200, in a Layer-2 network environment without Layer-3 device, Switch C is attached to the multicast source (Source) through Ethernet1/2. At least one receiver is connected to Switch B and Switch C respectively.
- MLDv1 is enabled on all the receivers. Switch A, Switch B, and Switch C run MLD Snooping. Switch A acts as the MLD Snooping querier.

Network diagram

Figure 200 Network diagram for MLD Snooping querier configuration



Configuration procedure

1 Configure switch A

# Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

# Create VLAN 100 and add Ethernet1/1/3 and Ethernet1/1/1 to VLAN 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port ethernet 1/1/3 ethernet 1/1/1
```

# Enable MLD Snooping in VLAN 100 and enable the MLD-Snooping querier function.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] mld-snooping querier
```

2 Configure Switch B

# Enable MLD Snooping globally.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

# Create VLAN 100, add Ethernet1/1/1 through Ethernet1/1/3 into the VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port ethernet 1/1/1 to ethernet 1/1/3
[SwitchB-vlan100] mld-snooping enable
```

3 Configuration on Switch C

# Enable MLD Snooping globally.

```
<SwitchC> system-view
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

# Create VLAN 100, add Ethernet1/1/1 through Ethernet1/1/3 to VLAN 100, and enable MLD Snooping in this VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port ethernet 1/1/1 to ethernet 1/1/3
[SwitchC-vlan100] mld-snooping enable
```

Troubleshooting MLD Snooping

Switch Fails in Layer 2 Multicast Forwarding

Symptom

A switch fails to implement MLD Snooping.

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 broadcast
- Certain ports have been configured as static member ports of IPv6 multicasts groups, and this configuration conflicts with the configured IPv6 multicast group policy.

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 interface 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 whether the function of dropping unknown IPv6 multicast data is enabled. If not, use the **drop-unknown** command to enable the function of dropping unknown IPv6 multicast data.

4 Use the **display mld-snooping group** command to check whether any port has been configured as a static member port of any IPv6 multicast group. If so, check whether this configuration conflicts with the configured IPv6 multicast group policy. If any conflict exists, remove the port as a static member of the IPv6 multicast group.

46 IPv6 PIM Configuration

When configuring IPv6 PIM, go to these sections for information you are interested in:

- *The term "router" in this document refers to a router in a generic sense or a Switch 8800 running IPv6 PIM.*
- *Currently, for Switch 8800s , only modules with the suffix DA, DB, or DC support IPv6 multicast. That is, a module that provides incoming ports for IPv6 multicast data should be such a module suffixed with DA, DB or DC, namely a IPv6-capable module. A module that provides outgoing interfaces can be any type of module.*
- *In the case that IPv6 multicast data can be delivered to the switch only through ports on a non-IPv6-capable module, you can configure IPv6 multicast centralized mode to enable normal forwarding of IPv6 multicast data. See "Configuring IPv6 Multicast and IPv6 Unicast Centralized Mode Example" on page 710 .*
- *Currently, a POS interface on a Switch 8800 does not support IPv6 multicast, namely those commands used in interface view cannot be executed in POS interface view.*

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 *"Implementation of the RPF mechanism" on page 516*.

Based on the forwarding 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).

To facilitate description, a network comprising IPv6 PIM routers or IPv6 PIM routing switches 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 building
- Graft
- Assert

Neighbor discovery

In a 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.

Every activated interface on a router sends hello messages periodically, and thus learns the IPv6 PIM neighboring information pertinent to the interface.

SPT building

1 In an IPv6 PIM-DM domain, an IPv6 multicast source first floods IPv6 multicast packets when it sends IPv6 multicast data to an IPv6 multicast group G: The packet is subject to an RPF check. If the packet passes the RPF check, the router creates an

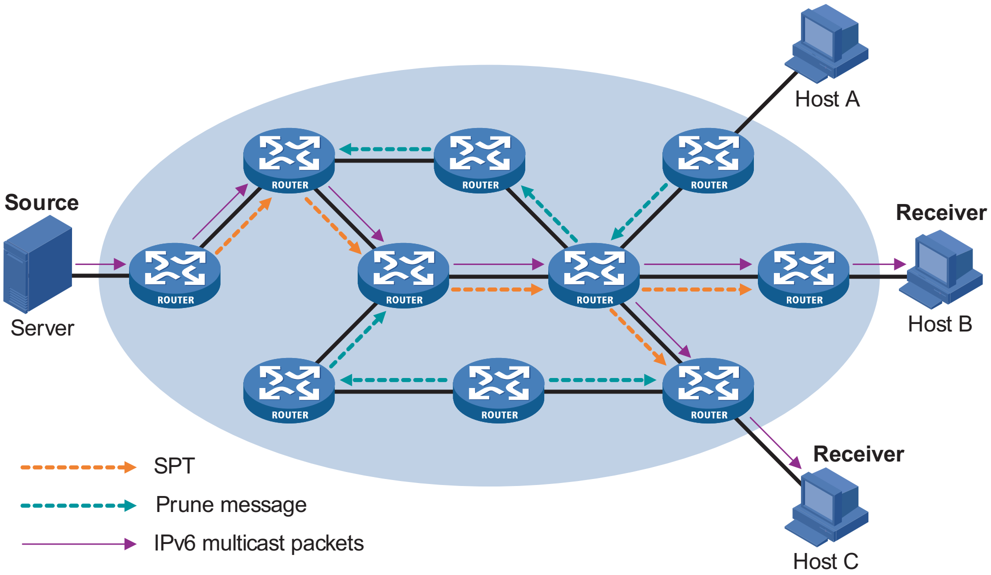
(S, G) entry and forwards the packet to all downstream nodes in the network. In the flooding process, an (S, G) entry is created on all the routers in the IPv6 PIM-DM domain.

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.

- *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 201, 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 201 Building an SPT in an IPv6 PIM-DM domain



The "flood and prune" process takes place periodically. The device sets a countdown timer for each pruned interface. When the countdown timer for a pruned interface expires, multicast traffic starts to flow to the interface again, and then the interface is pruned again when it no longer has any multicast receiver attached to it.

Graft

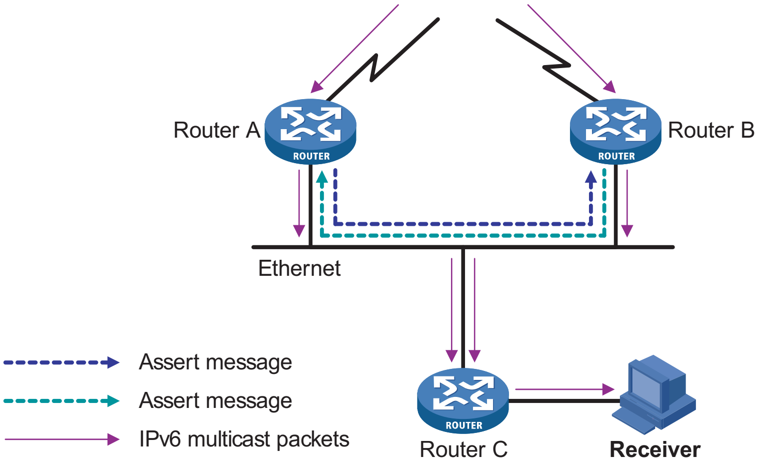
When a host attached to a pruned node joins an IPv6 multicast group, to reduce the join latency, IPv6 PIM-DM uses the graft mechanism to resume IPv6 multicast data forwarding to that branch. The process is as follows:

1. The node that needs to receive IPv6 multicast data sends a graft message hop by hop toward the source, 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

If multiple multicast routers exist on a multi-access subnet, duplicate IPv6 multicast packets may flow to the same subnet. To shutoff duplicate flows, the assert mechanism is used for election of a single IPv6 multicast forwarder on a multi-access network.

Figure 202 Assert mechanism



As shown in Figure 202, 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 IP address of the local interface 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 mod 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 a multicast source sends an IPv6 multicast packet to an IPv6 multicast group, the router directly connected with the multicast source first registers the multicast source with the RP by sending a register message to the RP by unicast. The arrival of this message at the RP triggers the establishment of an SPT. Then, the 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.

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
- Extracting an embedded RP address
- RPT building
- IPv6 Multicast source registration
- Switchover from RPT to SPT
- Assert

Neighbor discovery

IPv6 PIM-SM uses exactly the same neighbor discovery mechanism as IPv6 PIM-DM does. Refer to "Neighbor discovery" on page 672.

DR election

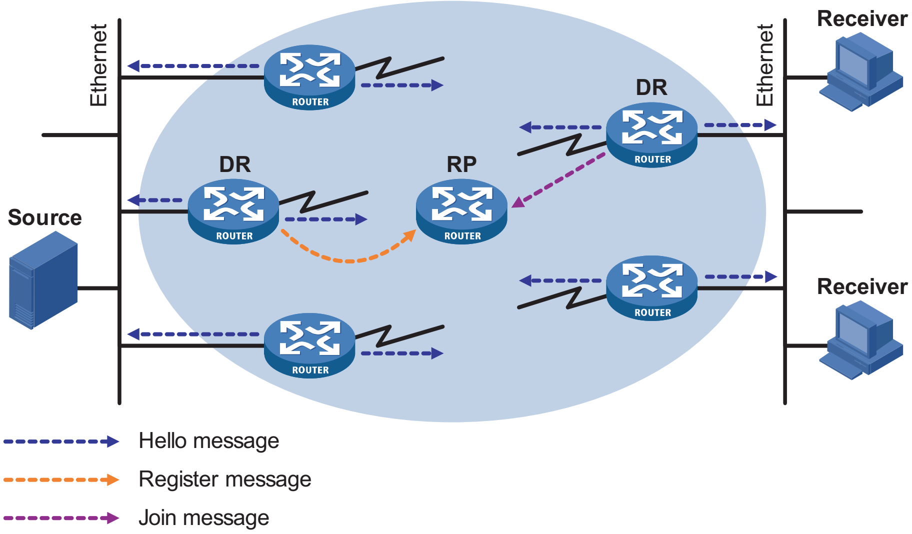
IPv6 PIM-SM also uses hello messages to elect a designated router (DR) for a multi-access network. The elected DR will be the only multicast forwarder on this multi-access network.

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.

- *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 IPv6 PIM-SM. IPv6 PIM-DM itself does not require a DR. However, if MLDv1 runs on any multi-access network in an IPv6 PIM-DM domain, a DR must be elected to act as the MLDv1 querier on that multi-access network.*
- *MLD must be enabled on a device that acts as a DR before receivers attached to this device can join IPv6 multicast groups through this DR.*

For details about MLD, refer to *"MLD Overview" on page 637*.

Figure 203 DR election



As shown in Figure 203, 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

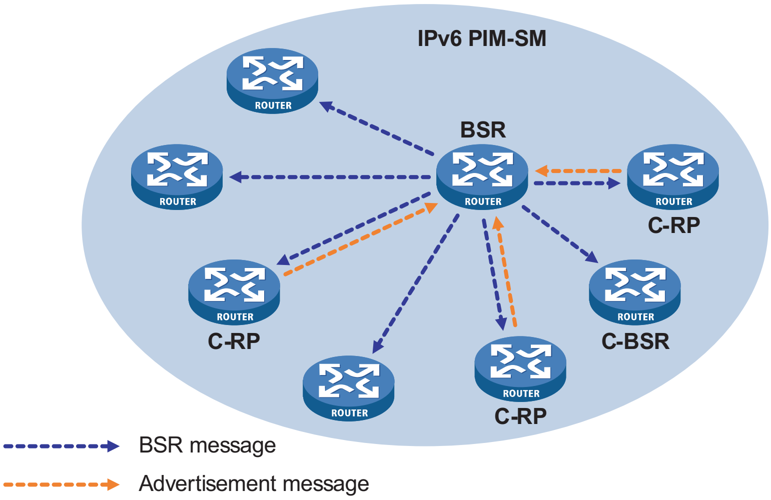
The RP is the core of an IPv6 PIM-SM domain. For a small-sized, simple network, one RP is enough for forwarding IPv6 multicast information throughout the network, and the position of the RP can be statically specified on each router in the IPv6 PIM-SM domain. In most cases, however, an IPv6 PIM-SM network 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, each IPv6 multicast group should have its own RP. Therefore, a bootstrap mechanism is needed for dynamic RP election. For this purpose, a bootstrap router (BSR) should be configured.

As the administrative core of an IPv6 PIM-SM domain, the BSR collects advertisement messages (C-RP-Adv messages) from candidate-RPs (C-RPs) and chooses the appropriate C-RP information for each IPv6 multicast group to form an RP-Set, which is a database of mappings between IPv6 multicast groups and RPs. The BSR then floods the RP-Set to the entire IPv6 PIM-SM domain. Based on the information in these RP-Sets, all routers (including the DRs) in the network can calculate the location of the corresponding RPs.

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 through the bootstrap mechanism to avoid service interruption. Similarly, multiple C-RPs can be configured in an IPv6 PIM-SM domain, and the position of the RP corresponding to each IPv6 multicast group is calculated through the BSR mechanism.

Figure 204 shows the positions of C-RPs and the BSR in the network.

Figure 204 Communication between the BSR and C-RPs



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 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 Upon getting the IPv6 multicast address, the IPv6 multicast source starts sending 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 by unicast.

Currently, the IPv6 multicast address range for embedded RP used by the Switch 8800s is FF7x::/12 or FFFx::/12, where "x" refers to the Scope field in the IPv6 multicast address. Different values define different scopes. For details about the Scope field, refer to "IPv6 Multicast Addresses" on page 498.

Extracting an embedded RP address

Figure 205 shows the structure of an IPv6 multicast address with an embedded RP address. Fields of such an IPv6 address are described as follows:

- FF7x: The first 16 bits of an IPv6 multicast address with an embedded RP address must be FF7x or FFFx.
- rsv: Reserved bits.
- RIID: Interface ID of the RP.
- plen: Obtained prefix length, in the hexadecimal notation. The maximum value is 40, indicating to copy the first 4*16 (64) bits of the prefix to the IPv6 unicast address structure of the RP.
- prefix: Prefix of the RP's IPv6 address. The maximum length is 64 bits.
- group ID: IPv6 multicast group ID.

Figure 205 Structure of an IPv6 multicast address with an embedded RP address

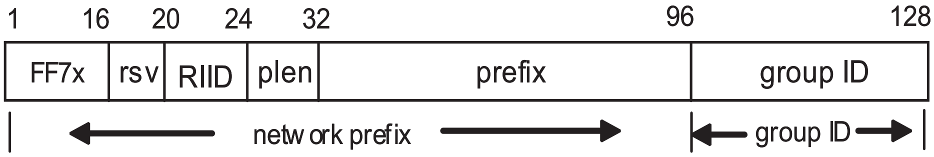
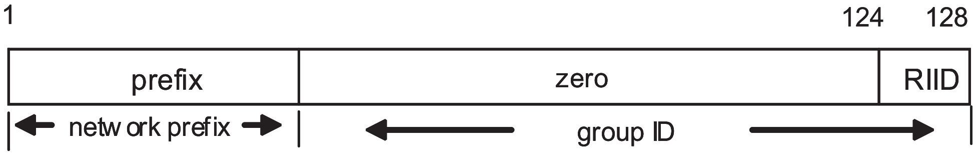


Figure 206 shows the structure of an RP address extracted an IPv6 address.

Figure 206 Extracted IPv6 unicast address of the RP



prefix: The prefix of the embedded RP unicast address extracted from the IPv6 multicast address. The number of bits extracted is determined by the plen field of the IPv6 multicast address.

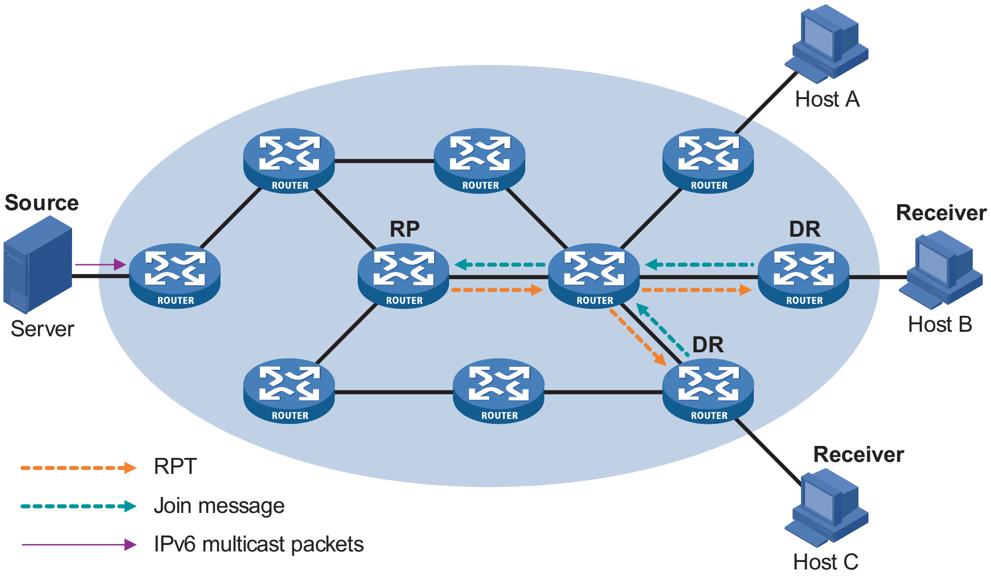
zero: These bits are zeroed.

RIID: The RIID field of the IPv6 multicast address is extracted as the interface ID of the IPv6 unicast address of the RP.

For example, if an IPv6 multicast address with an embedded RP address is FF77:0630:2001:DB8:BEEF::/80, the IPv6 unicast address of the RP is 2001:DB8:BEEF::6/48.

RPT building

Figure 207 Building an RPT in IPv6 PIM-SM



As shown in Figure 207, the process of building an RPT is as follows:

1. When a receiver joins an 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.

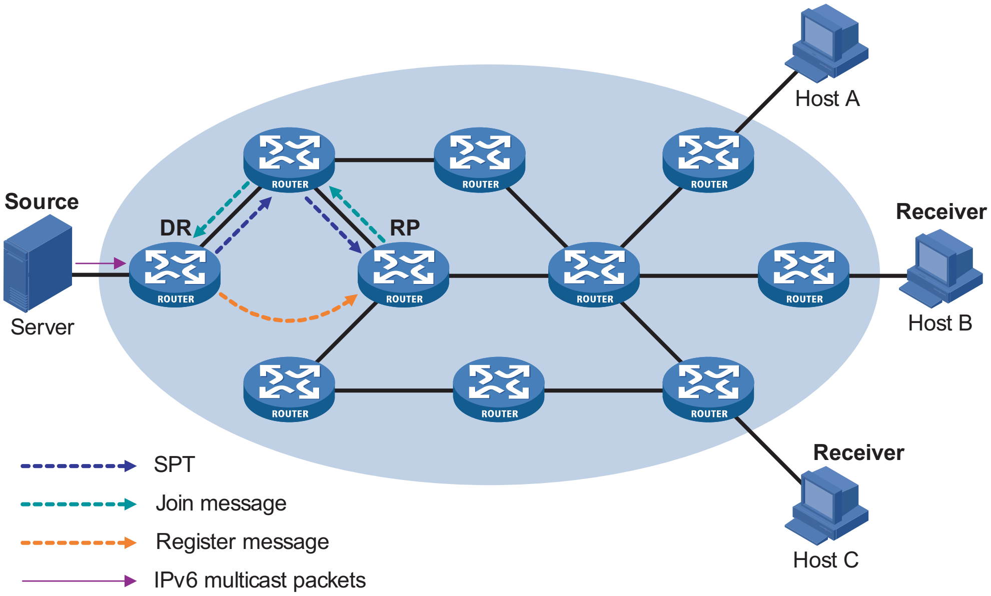
   The IPv6 multicast data addressed to the IPv6 multicast group G flows through the RP, reaches the corresponding DR along the established RPT, and finally is delivered to the receiver.

   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 its link with this downstream node from the outgoing interface list and checks whether it itself 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 208 IPv6 multicast source registration



As shown in Figure 208, the IPv6 multicast source registers with the RP as follows:

1 When the IPv6 multicast source S sends the first IPv6 multicast packet to an 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, on one hand, it extracts the multicast packet from the register message and forwards the multicast IPv6 multicast packet down the RPT, and, on the other hand, 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 a (S, G) entry in its forwarding table. The IPv6 multicast source 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.

Switchover from RPT to SPT

Initially, multicast traffic flows along an RPT from the RP to the receivers. Because the RPT is not necessarily the tree that has the shortest path, the receiver-side DR initiates an RPT-to-SPT switchover process upon receiving the first multicast packet along the RPT by default. The RPT-to-SPT switchover process is as follows:

1. First, the receiver-side DR sends an (S, G) join message hop by hop to 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.
2. Subsequently, the receiver-side DR sends a prune message hop by hop to the RP. Upon receiving this prune message, the RP forwards it towards the IPv6 multicast source, thus to implement RPT-to-SPT switchover.

After the RPT-to-SPT switchover, IPv6 multicast data can be directly sent from the source to the receivers. IPv6 PIM-SM builds SPTs through RPT-to-SPT switchover more economically than IPv6 PIM-DM does through the "flood and prune" mechanism.

Assert

IPv6 PIM-SM uses exactly the same assert mechanism as IPv6 PIM-DM does. Refer to “Assert” on page 674.

Protocols and Standards

IPv6 PIM-related specifications are as follows:

- RFC 2362: Protocol Independent Multicast-sparse Mode(PIM-SM):Protocol Specification
- 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
- draft-ietf-pim-sm-v2-new-06: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- draft-ietf-pim-dm-new-v2-02: Protocol Independent Multicast-Dense Mode (PIM-DM)
- draft-ietf-pim-v2-dm-03: Protocol Independent Multicast Version 2 Dense Mode Specification
- draft-ietf-pim-sm-bsr-02: Bootstrap Router (BSR) Mechanism for PIM Sparse Mode
- draft-ietf-ssm-arch-01: Source-Specific Multicast for IP
- draft-ietf-ssm-overview-04: 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” on page 682	Required
“Enabling State Refresh” on page 682	Optional
“Configuring State Refresh Parameters” on page 683	Optional

Task	Remarks
"Configuring IPv6 PIM-DM Graft Retry Period" on page 683	Optional
"Configuring IPv6 PIM Common Information" on page 691	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
- TTL value of state refresh messages
- Graft retry period

Enabling IPv6 PIM-DM

An IPv6 PIM-DM enabled device sends hello messages periodically to discover IPv6 PIM neighbors and processes messages from IPv6 PIM neighbors. When deploying an IPv6 PIM-DM domain, you are recommended to enable IPv6 PIM-DM on all interfaces of non-border devices (border devices are IPv6 PIM routers or IPv6 routing switches located on the boundary of BSR admin-scope regions).

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 VLAN 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.*
- *After IPv6 PIM-DM is enabled on a VLAN interface, MLD snooping cannot be enabled in the corresponding VLAN, and vice versa.*

Enabling State Refresh

An interface without the state refresh capability cannot forward state refresh messages.

Follow these steps to enable state refresh:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-

To do...	Use the command...	Remarks
Enable state refresh	**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 device directly connected with the IPv6 multicast source periodically sends (S, G) state refresh messages, which are forwarded hop by hop along the initial flooding path of the IPv6 PIM-DM domain, to refresh the prune timer state of all the devices on the path.

A device may receive multiple state refresh messages within a short time, of which some may be duplicated messages. To keep device from receiving such duplicated messages, you can configure the time the device must wait before receiving the next state refresh message. If a new state refresh message is received within the waiting time, the device will discard it; if this timer times out, the device will accept a new state refresh message, refresh its own IPv6 PIM state, and reset the waiting timer.

The TTL value of a state refresh message decrements by 1 whenever it passes a device 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.

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 TTL value of state refresh messages	**state-refresh-ttl** *ttl-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 device does not receive a graft-ack message from the upstream device within the specified time after it sends a graft message, the device keeps sending new graft messages at a configurable interval, namely graft retry period, until it receives a graft-ack from the upstream device.

Follow these steps to configure IPv6 PIM-DM graft retry period:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-

To do...	Use the command...	Remarks
Configure graft retry period	**pim ipv6 timer graft-retry** *interval*	Optional 3 seconds by default

For the configuration of other timers in IPv6 PIM-DM, refer to "Configuring IPv6 PIM Common Timers" on page 694.

Configuring IPv6 PIM-SM

A device can serve as a C-RP and a C-BSR at the same time.

IPv6 PIM-SM Configuration Task List

Complete these tasks to configure IPv6 PIM-SM:

Task		Remarks
"Enabling IPv6 PIM-SM" on page 685		Required
"Configuring a BSR" on page 685	"Performing basic C-BSR configuration" on page 685	Optional
	"Configuring a BSR admin-scope region boundary" on page 687	Optional
	"Configuring global C-BSR parameters" on page 687	Optional
"Configuring an RP" on page 688	"Configuring a static RP" on page 688	Optional
	"Configuring a C-RP" on page 688	Optional
	"Enabling embedded RP" on page 689	Optional
	"Configuring C-RP timers" on page 689	Optional
"Configuring IPv6 PIM-SM Register Messages" on page 690		Optional
"Disabling RPT-to-SPT Switchover" on page 691		Optional
"Configuring IPv6 PIM Common Information" on page 691		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:

- An IPv6 ACL rule defining a legal BSR address range
- Hash mask length for RP selection calculation
- C-BSR priority
- Bootstrap interval
- Bootstrap timeout time

- An IPv6 ACL rule defining a legal C-RP address range and the range of IPv6 multicast groups to be served
- C-RP-Adv interval
- C-RP timeout time
- The IPv6 address of a static RP
- An IPv6 ACL rule for register message filtering
- Register suppression timeout time
- Probe time
- Whether to disable RPT-to-SPT switchover

Enabling IPv6 PIM-SM

With IPv6 PIM-SM enabled, a device sends hello messages periodically to discover IPv6 PIM neighbors and processes messages from IPv6 PIM neighbors. When deploying an IPv6 PIM-SM domain, you are recommended to enable IPv6 PIM-SM on all interfaces of non-border devices (border devices are IPv6 PIM devices located on the boundary of BSR admin-scope regions).

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
Enter VLAN 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.*
- *After IPv6 PIM-SM is enabled on a VLAN interface, MLD snooping cannot be enabled in the corresponding VLAN, and vice versa*

Configuring a BSR

The BSR is dynamically elected from a number of C-BSRs. Because it is unpredictable which device will finally win a BSR election, the commands introduced in this section must be configured on all C-BSRs.

Performing basic C-BSR configuration

An IPv6 PIM-SM domain can have only one BSR, but must have at least one C-BSR. Any device can be configured as C-BSR. Elected from C-BSRs, a BSR is responsible for collecting and advertising RP information in the IPv6 PIM-SM.

C-BSRs should be configured on devices in the backbone network. When configuring a device as a C-BSR, be sure to specify an IPv6 PIM-SM-enabled. The BSR election process is 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 BSR messages based on the address range, thus to prevent malicious hosts from initiating attacks by disguising themselves as legitimate BSRs. To protect legitimate BSRs from being maliciously replaced, preventive measures are taken specific to the following two situations:

1. Some malicious hosts intend to fool devices by forging BSR messages and change the RP mapping relationship. Such attacks often occur on border devices. 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 border devices to perform neighbor check and RPF check on BSR messages and discard unwanted messages.
2. When a device in the network is controlled by an attacker or when an illegal device is present in the network, the attacker can configure such a device to be a C-BSR and make it win BSR election so as to gain the right of advertising RP information in the network. After being configured as a C-BSR, a device automatically floods the network with BSR messages. As a BSR message has a TTL value of 1, the whole network will not be affected as long as the neighbor device discards these BSR messages. Therefore, if a legal BSR address range is configured on all devices in the entire network, all devices will discard BSR messages from out of the legal address range, and thus this kind of attacks can be prevented.

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 aforesaid problem may 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*]]	Optional No C-BSRs are configured by default.
Configure a legal BSR address range	**bsr-policy** *acl6-number*	Optional No restrictions by default

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 a BSR admin-scope region boundary

A BSR has its specific service scope. A number of BSR boundary interfaces divide a network into different BSR admin-scope regions. Bootstrap messages cannot cross the admin-scope region boundary, while other types of IPv6 PIM messages can.

Follow these steps to configure a BSR admin-scope region boundary:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configuring a BSR Admin-scope region boundary	**pim ipv6 bsr-boundary**	Required No BSR admin-scope region boundary by default

Configuring global C-BSR parameters

The BSR election winner advertises its own IPv6 address and RP-Set information throughout the region it serves through bootstrap messages. The BSR floods bootstrap messages throughout the network periodically. Any C-BSR that receives a bootstrap message maintains the BSR state for a configurable period of time (BSR state timeout), during which no BSR election takes place. When the BSR state times out, a new BSR election process will be triggered among the C-BSRs.

Follow these steps to configure global C-BSR parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IPv6 PIM view	**Pim ipv6**	-
Configure the Hash mask length for RP selection calculation	**c-bsr hash-length** *hash-length*	Optional 126 by default
Configure the C-BSR priority	**c-bsr priority** *priority*	Optional 0 by default
Configure the bootstrap interval	**c-bsr interval** *interval*	Optional For the system default, see "Note" below.
Configure the bootstrap timeout time	**c-bsr holdtime** *interval*	Optional For the system default, see "Note" below.

About the bootstrap timeout time:

- *By default, the bootstrap timeout time is determined by this formula: Bootstrap timeout = Bootstrap interval × 2 + 10. The default bootstrap interval is 60 seconds, so the default bootstrap timeout = 60 × 2 + 10 = 130 (seconds).*
- *If this parameter is manually configured, the system will use the configured value.*

About the bootstrap interval:

- *By default, the bootstrap timeout time is determined by this formula: Bootstrap interval = (Bootstrap timeout - 10) ÷ 2. The default bootstrap timeout is 130 seconds, so the default bootstrap interval = (130 - 10) ÷ 2 = 60 (seconds).*
- *If this parameter is manually configured, the system will use the configured value.*

***CAUTION:** In configuration, make sure that the bootstrap interval is smaller than the bootstrap timeout time.*

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. To enable a static RP to work normally, you must perform this configuration on all the devices in the IPv6 PIM-SM domain and specify the same RP address.

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**]	Optional No static RP by default

Configuring a C-RP

In an IPv6 PIM-SM domain, you can configure devices that intend to become the RP into 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 devices and organizes the information into to an RP-Set, which is flooded throughout the entire network. Then, the other devices 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 devices.

To guard again 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.

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**	-

To do...	Use the command...	Remarks
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*] *	Optional 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

- *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 mechanism enabled, the device can resolve the RP address from an IPv6 multicast addresses to replace the statically configured RP or the RP dynamically calculated based on the BSR mechanism. The DR can get the RP address just by analyzing the multicast data.

Follow these steps to enable embedded RP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IPv6 PIM view	**pim ipv6**	-
Enable embedded RP	**embedded-rp** [*acl6-number*]	Optional Enabled by default

Configuring C-RP timers

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 devices in the network.

Each C-RP encapsulates a timeout value in its C-RP-Adv message. Upon receiving this 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.

Follow these steps to configure C-RP timers:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IPv6 PIM view	**pim ipv6**	-

To do...	Use the command...	Remarks
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

- *The commands introduced in this section are to be configured on C-RPs.*
- *For the configuration of other timers in IPv6 PIM-SM, refer to "Configuring IPv6 PIM Common Timers" on page 694.*

Configuring IPv6 PIM-SM Register Messages

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 a specific 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 enters the register suppression state.

In a probe suppression cycle, the DR can send a null register message (a register message without multicast data encapsulated), a certain length of time defined by the probe time before the register suppression timer expires, to the RP to indicate that the multicast source is active. When the register suppression timer expires, the DR starts sending register messages again. A smaller register suppression timeout setting will cause the RP to receive bursting IPv6 multicast data more frequently, while a larger timeout setting will result in a larger delay for new receivers to join the IPv6 multicast group they are interested in.

Follow these steps to configure IPv6 PIM-SM register messages:

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

To do...	Use the command...	Remarks
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 timeout time	**register-suppression-timeout** *interval*	Optional 60 seconds by default
Configure the probe time	**probe-interval** *interval*	Optional 5 seconds by default

Typically, you need to configure the above-mentioned parameters on the receiver-side DR and the RP only. Since both the DR and RP are elected, however, you should carry out these configurations on the devices that may win the DR election and on the C-RPs that may win RP elections.

Disabling RPT-to-SPT Switchover

When a Switch 8800 serves as the receiver-side DR, by default, it initiates an RPT-to-SPT switchover process immediately after receiving the first multicast packet along the RPT. You can disable RPT-to-SPT switchover with the following command.

Follow these steps to disable RPT-to-SPT switchover:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter IPv6 PIM view	**pim ipv6**	-
Disable RPT-to-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 from the RPT.

To avoid forwarding failure, do not disable RPT-to-SPT switchover on a device that may become an RP (namely, a static RP or a C-RP).

Configuring IPv6 PIM Common Information

For the configuration tasks 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 PIM IPv6 view and interface view, the configuration performed in interface view has given priority, regardless of the configuration sequence.*

IPv6 PIM Common Information Configuration Task List

Complete these tasks to configure IPv6 PIM common information:

Task	Remarks
“Configuring an IPv6 PIM filter” on page 692	Optional
“Configuring IPv6 PIM Hello Options” on page 693	Optional
“Configuring IPv6 PIM Common Timers” on page 694	Optional
“Configuring Join/Prune Message Limits” on page 696	Optional

Configuration Prerequisites

Before configuring IPv6 PIM common information, 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)

Before configuring IPv6 PIM common information, prepare the following data:

- An IPv6 ACL rule as IPv6 multicast data filter
- 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 PIM filter

No matter in an IPv6 PIM-DM domain or an IPv6 PIM-SM domain, devices 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 devices 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 PIM 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

- *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 IPv6 PIM Hello Options

No matter in an IPv6 PIM-DM domain or an IPv6 PIM-SM domain, the hello messages sent among devices 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 device to win DR election. You can configure this parameter on all the devices 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 device has received no hello message from an IPv6 PIM neighbor, it assumes that this neighbor has expired or become unreachable. You can configure this parameter on all devices in the IPv6 PIM domain. If you configure different values for this timer on different IPv6 PIM neighbors, the largest value will take effect.
- 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 bit. You can configure this parameter on all devices in the IPv6 PIM domain. If different LAN-delay or override-interval values result from the negotiation among all the IPv6 PIM devices, the largest value will take effect.

The LAN-delay setting will cause the upstream devices to delay processing received prune messages. If the LAN-delay setting is too small, it may cause the upstream device to stop forwarding IPv6 multicast packets before a downstream device sends a prune override message. Therefore, be cautious when configuring this parameter.

The override-interval sets the length of time a downstream device is allowed to wait before sending a prune override message. When a device receives a prune message from a downstream device, it does not perform the prune action immediately; instead, it maintains the current forwarding state for a period of time defined by LAN-delay. If the downstream device 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 LAN-delay timer times out.

A hello message sent from an IPv6 PIM device 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 device does not change unless the status of the device changes (for example, when IPv6 PIM is just enabled on the interface or the device is restarted). When the device starts or restarts sending hello messages, it generates a new generation ID. If an IPv6 PIM device finds that the generation ID in a hello message from the upstream device 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 upstream device will explicitly track which downstream devices are joined to it. The join suppression feature should be enabled or disable on all IPv6 PIM devices on the same subnet.

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**	Optional 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 VLAN 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**	Optional Enabled by default
Configure the interface to reject hello messages without a generation ID	**pim ipv6 require-genid**	Optional By default, hello messages without Generation_ID are accepted.

Configuring IPv6 PIM Common Timers

IPv6 PIM devices discover IPv6 PIM neighbors and maintain IPv6 PIM neighboring relationships with other devices by periodically sending out hello messages.

Upon receiving a hello message, an IPv6 PIM device waits a random period, which is equal to or smaller than the maximum delay between hello messages, before sending out a hello message. This avoids collisions that occur when multiple IPv6 PIM devices send hello messages simultaneously.

Any device 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 IPv6 multicast forwarding.

An IPv6 PIM device periodically sends join/prune messages to its upstream device for state update. A join/prune message contains the join/prune timeout time. The upstream device sets a join/prune timeout timer for each pruned downstream interface, and resumes the forwarding state of the pruned interface when this timer times out.

When a device fails to receive subsequent IPv6 multicast data from the IPv6 multicast source S, the device will not immediately deletes the corresponding (S, G) entries; instead, it maintains (S, G) entries for a period of time, namely the IPv6 multicast source lifetime, before delete the (S, G) entries.

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 assert timeout time	**holdtime assert** *interval*	Optional 180 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 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 VLAN 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 assert timeout time	**pim ipv6 holdtime assert** *interval*	Optional 180 seconds by default

To do...	Use the command...	Remarks
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

If there are no special networking requirements, we recommend that you use the default settings.

Configuring Join/Prune Message Limits

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 limits:

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** | **graf**t | **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

To do...	Use the command...	Remarks
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 joint/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 use view

IPv6 PIM Configuration Examples

IPv6 PIM-DM Configuration Example

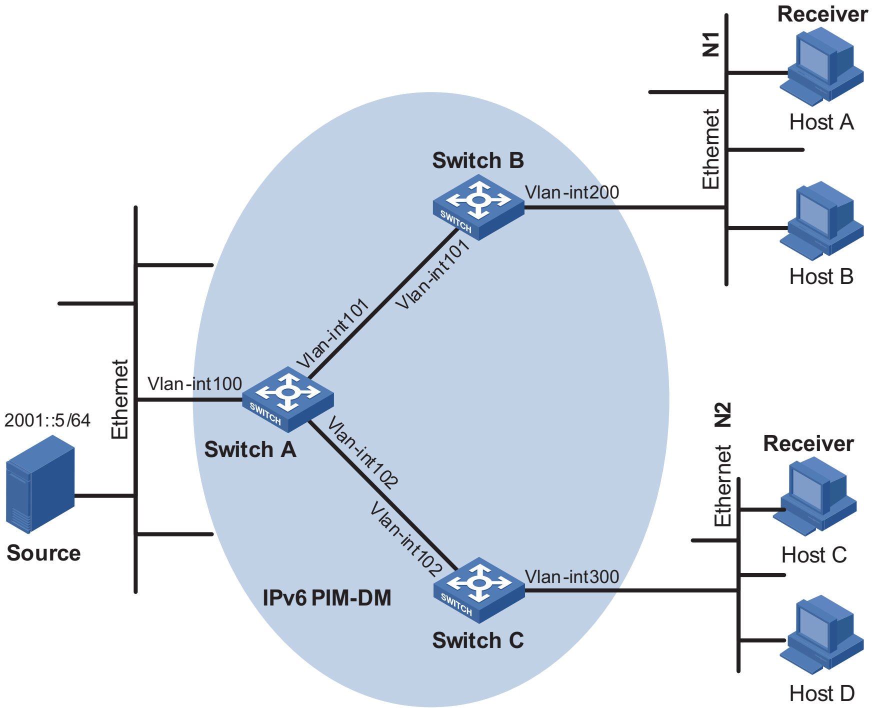
Network requirements

- 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.
- Switch A connects to the network where the source resides through VLAN-interface 100, to Switch B through VLAN-interface 101, and to Switch C through VLAN-interface 102.
- Switch B and Switch C connect to N1 and N2 through VLAN-interface 200 and VLAN-interface 300 respectively, and to Switch A through VLAN-interface 101 and VLAN-interface 102 respectively.

- Switch A is the MLD querier on the multi-access subnet.

Network diagram

Figure 209 Network diagram for IPv6 PIM-DM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	2001::1/64	Switch C	Vlan-int102	2003::2/64
	Vlan-int101	2002::1/64		Vlan-int300	4001::1/64
	Vlan-int102	2003::1/64			
Switch B	Vlan-int101	2002::1/64			
	Vlan-int200	3001::1/64			

Configuration procedure

1 Configure the interface IPv6 addresses and unicast routing protocol for each switch

Configure the IP address and prefix length for each interface and enable OSPFv3 on each VLAN interface. Detailed configuration steps are omitted here.

2 Enable IPv6 multicast routing, and enabling IPv6 PIM-DM on each interface

# Enable IPv6 multicast routing on Switch A, enable IPv6 PIM-DM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[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
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim ipv6 dm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to the configuration on Switch A.

3 Enable MLD on the host-side interfaces of Switch B and Switch C

# Enable IPv6 multicast routing on Switch B, and enable MLDv1 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] quit
```

The configuration on Switch is similar to the configuration on Switch B.

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
Vpn-instance: public net
 Interface           NbrCnt HelloInt   DR-Pri    DR-Address
 Vlan-interface100     0      30          1      FE80::200:AFF:F
                                                 E01:109 (local)
 Vlan-interface101     1      30          1       FE80::A01:109:1
                                                  (local)
 Vlan-interface102     1      30          1       FE80::A01:109:2
                                                  (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 A.

```
[SwitchA] display pim ipv6 neighbor
Vpn-instance: public net
 Total Number of Neighbors = 2

 Neighbor         Interface             Uptime     Expires  Dr-Priority
 FE80::A01:104:1  Vlan-interface101     00:04:16 00: 01:29   1
 FE80::A01:105:1  Vlan-interface102     00:03:54 00: 01:17   1
```

Assume that Host A needs to receive the information addressed to an IPv6 multicast group G (FF0E::101:101). After IPv6 multicast source S (2001::5) sends IPv6 multicast packets to the IPv6 multicast group G, an SPT is established through traffic flooding. Upon receiving the (*, G) join message from Host A, Switch B updates its (S, G) routing entry. Then, all the IPv6 PIM switches in the SPT update their respective (S,G) 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
Vpn-instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (2001::5, FF0E::101:101)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:01:20
     Upstream interface: Vlan-interface100
         Upstream neighbor: FE80::200:AFF:FE01:108
         RPF prime neighbor: FE80::200:AFF:FE01:108
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: Vlan-interface101
             Protocol: pim-dm, UpTime: 00:01:20, Expires:  -
         2: Vlan-interface102
             Protocol: pim-dm, UpTime: 00:01:20, Expires:  -
```

# View the IPv6 PIM multicast routing table information on Switch B.

```
[SwitchB] display pim ipv6 routing-table
Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::101:101)
     Protocol: pim-dm, Flag: WC
     UpTime: 01:46:23
     Upstream interface: Vlan-interface101
         Upstream neighbor: FE80::A01:109:1
         RPF prime neighbor: FE80::A01:109:1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface200
             Protocol: mld, UpTime: 01:46:23, Expires: never

 (2001::5, FF0E::101:101)
     Protocol: pim-dm, Flag: ACT
     UpTime: 00:02:19
     Upstream interface: Vlan-interface101
         Upstream neighbor: FE80::A01:109:1
         RPF prime neighbor: FE80::A01:109:1
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface200
             Protocol: pim-dm, UpTime: 00:02:19, Expires:  -
```

The information on Switch C is similar to that on Switch B.

IPv6 PIM-SM Configuration Example

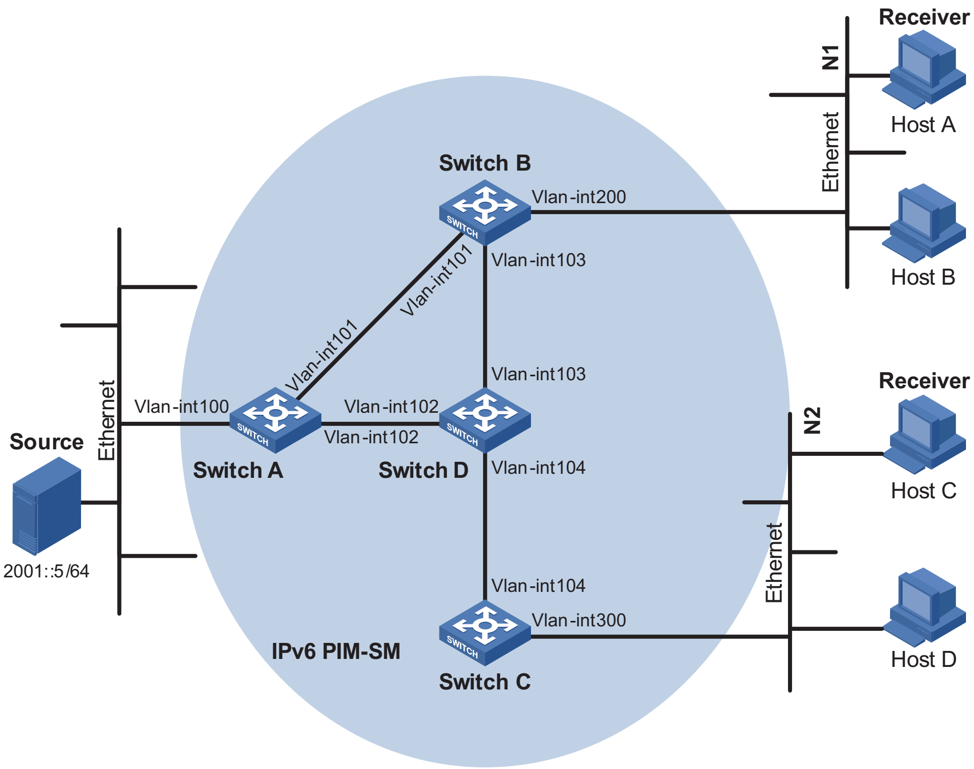
Network requirements

- 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 A connects to the network where the source resides through VLAN-interface 100.
- Switch B connect to N1 through VLAN-interface 200, and to Switch A and Switch D through VLAN-interface 101 and VLAN-interface 103 respectively.
- Switch C connects to N2 through VLAN-interface 300 and to Switch D through VLAN-interface 104.
- Switch D connects to Switch A, Switch B, and Switch C, and its VLAN-interface 103 acts a C-BSR and a C-RP, with the range of IPv6 multicast groups served by the C-RP being FF0E::1/64.

Network diagram

Figure 210 Network diagram for IPv6 PIM-SM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	2001::1/64	Switch C	Vlan-int104	2005::2/64
	Vlan-int101	2002::1/64		Vlan-int300	4001::2/64
	Vlan-int102	2003::1/64		Vlan-int102	2003::2/64
Switch B	Vlan-int101	2002::2/64	Switch D	Vlan-int103	3002::2/64
	Vlan-int103	2004::1/64		Vlan-int104	2005::1/64
	Vlan-int200	3001::1/64			

Configuration procedure

1 Configure the interface IPv6 addresses and unicast routing protocol for each switch

Configure the IP address and prefix length for each interface as per Figure 210. Configure OSPFv3 for interoperation among the switches in the IPv6 PIM-SM domain. Detailed configuration steps are omitted here.

2 Enable IPv6 multicast routing, and enabling IPv6 PIM-SM on each interface

# 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 102
[SwitchA-Vlan-interface102] pim ipv6 sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B, Switch C and Switch D is similar to that on Switch A.

3 Configure a C-BSR and a C-RP

# Configure the RP service range and the C-BSR and C-RP locations on Switch D.

```
<SwitchD> system-view
[SwitchD] acl ipv6 number 2000
[SwitchD-acl6-basic-2000] rule permit source ff0e:: 64
[SwitchD-acl6-basic-2000] quit
[SwitchD] pim ipv6
[SwitchD-pim6] c-bsr 2004::2
[SwitchD-pim6] c-rp 2004::2 group-policy 2000
[SwitchD-pim6] quit
```

4 Enable MLD on the host-side interfaces of the switches connecting to the leaf networks.

# Enable IPv6 multicast routing on Switch B and enable MLDv1 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] quit
```

The configuration on Switch C is similar to that on Switch B.

5 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 B.

```
[SwitchB] display pim ipv6 interface
Vpn-instance: public net
 Interface           NbrCnt HelloInt   DR-Pri     DR-Address
Vlan-interface101     1      30         1          FE80::A01:10E:1
                                                       (local)
Vlan-interface200     0      30         1          FE80::200:AFF:F
                                                      E01:10E (local)
Vlan-interface103     1      30         1          FE80::9D62:0:FD
                                                      C5:2
```

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 B.

```
[SwitchB] display pim ipv6 bsr-info
Vpn-instance: public net
Elected BSR Address: 2004::2
     Priority: 0
     Hash mask length: 126
     State: Accept Preferred
     Uptime: 00:04:22
     Next BSR message scheduled at: 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
Vpn-instance: public net
Elected BSR Address: 2004::2
     Priority: 0
     Hash mask length: 126
     State: Elected
     Uptime: 00:01:10
     Next BSR message scheduled at: 00:00:48
 Candidate BSR Address: 2004::2
     Priority: 0
     Hash mask length: 126
     State: Elected

Candidate RP: 2004::2(Vlan-interface 103)
     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 B.

```
[SwitchB] display pim ipv6 rp-info
Vpn-instance: public net
 PIM-SM BSR RP information:
 prefix/prefix length: FF0E::101:101/64
     RP: 2004::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::101:101). An RPT will be built between Switch D and Switch B. A (*, G) entry is created on Switch D and Switch B on the RPT path. Once the multicast source S (2001::5/64) sends an IPv6 multicast packet to the IPv6 multicast group G, an (S, G) entry is created on the switches (Switch A and Switch D) on the source tree. 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
Vpn-instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (2001::5, FF0E::101:101)
     RP: 2004::2
     Protocol: pim-sm, Flag: SPT LOC ACT
     UpTime: 00:02:15
     Upstream interface: Vlan-interface100
         Upstream neighbor: FE80::200:AFF:FE01:10D
         RPF prime neighbor: FE80::200:AFF:FE01:10D
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: Vlan-interface101
             Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:15
         2: Vlan-interface102
             Protocol: pim-sm, UpTime: 00:02:15, Expires: 00:03:15
```

# View the IPv6 PIM multicast routing table information on Switch B.

```
[SwitchB] display pim ipv6 routing-table
Vpn-instance: public net
 Total 1 (*, G) entry; 0 (S, G) entry

 (*, FF0E::101:101)
     RP: 2004::2
     Protocol: pim-sm, Flag: WC
     UpTime: 00:14:44
     Upstream interface: Vlan-interface103
         Upstream neighbor: FE80::9D62:0:FDC5:2
         RPF prime neighbor: FE80::9D62:0:FDC5:2
     Downstream interface(s) information:
     Total number of downstreams: 1
         1: Vlan-interface200
             Protocol: mld, UpTime: 00:14:44, Expires: -
```

The configuration on Switch C is similar to that on Switch B.

View the IPv6 PIM multicast routing table information on Switch D.

```
[SwitchD] display pim ipv6 routing-table
Vpn-instance: public net
 Total 1 (*, G) entry; 1 (S, G) entry

 (*, FF0E::101:101)
     RP: 2004::2 (local)
     Protocol: pim-sm, Flag: WC
     UpTime: 00:16:56
     Upstream interface: Register
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: Vlan-interface103
             Protocol: pim-sm, UpTime: 00:16:56, Expires: 00:02:34
         2: Vlan-interface104
             Protocol: pim-sm, UpTime: 00:07:56, Expires: 00:02:35

 (2001::5, FF0E::101:101)
     RP: 2004::2 (local)
     Protocol: pim-sm, Flag: SWT ACT
     UpTime: 00:02:54
     Upstream interface: Vlan-interface102
         Upstream neighbor: FE80::9D62:0:FDC4:2
         RPF prime neighbor: FE80::9D62:0:FDC4:2
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: Vlan-interface103
             Protocol: pim-sm, UpTime: 00:02:54, Expires:  -
         2: Vlan-interface104
             Protocol: pim-sm, UpTime: 00:02:54, Expires: 00:02:36
```

Troubleshooting IPv6 PIM Configuration

Failure of Building a Multicast Distribution Tree Correctly

Symptom

None of the devices in the network 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.
- The RPF interface must support IPv6 PIM. 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 exist to the IPv6 multicast source or the RP.
2. Check that the RPF interface supports IPv6 PIM. 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 device.
6. Check that the same IPv6 PIM mode is enabled on all the devices 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 devices: IPv6 PIM-SM on all devices, or IPv6 PIM-DM on all devices.

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 devices 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 devices 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. Use the **display ipv6 routing-table** command to check whether a route is available on each device 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 devices. In the case of inconsistent RP information, configure consistent RP address on all the devices in IPv6 PIM view.
3. Check the static RP configuration. Use the **display pim ipv6 rp-info** command to check whether the same RP address has been configured on all the devices throughout the network.

No Unicast Route Between BSR and C-RPs in IPv6 PIM-SM Domain

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 will not advertise bootstrap messages.
- The RP is the core of an IPv6 PIM-SM domain. Make sure that the RP information on all devices is exactly the same, a specific group G 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. Use the **display ipv6 routing-table** command to check whether routes are available on each device to the RP and the BSR, and whether a route is available between the C-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 devices 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 device, 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 devices.

47 CENTRALIZED MODE FOR IPV6

Centralization is used when an IPv4 hardware module needs to forward IPv6 traffic. This accomplished through the use of Link Aggregation service ports, which creates a loopback group, on an IPv6 hardware ready module. When configuring centralization refer to the following sections:

- "Centralized Mode for IPv6 Multicast and IPv6 Unicast" on page 709
- "Configuring Centralized Mode for IPv6 Multicast and IPv6 Unicast" on page 709
- "Configuring IPv6 Multicast and IPv6 Unicast Centralized Mode Example" on page 710

Before configuring Centralized Mode for IPv6 multicast, make sure that at least one IPv6-capable module is installed in the device.

Centralized Mode for IPv6 Multicast and IPv6 Unicast

Centralized Mode for IPv6 traffic is designed to make IPv6 multicast and IPv6 unicast possible on modules that do not have the hardware support for IPv6. Currently, for Switch 8800s, only IPv6-capable modules support IPv6 traffic. In practice, however, a Switch 8800 may comprise different types of modules. If you need to forward IPv6 traffic on a non-IPv6 module, you can configure an IPv6 module to forward that traffic by using Link Aggregation service ports.

Service loops are specialized link aggregation groups, which will automatically re-direct IPv6 traffic from a non IPv6-capable module to the loopback port of an IPv6-cpable module.

Configuring Centralized Mode for IPv6 Multicast and IPv6 Unicast

Configuration Prerequisites

Before configuring Centralized Mode for IPv6 multicast or IPv6 unicast traffic, make sure that at least one IPv6 module is installed on your Switch 8800.

- 3C17537 3Com Switch 8800 2-port 10GBASE-X (XFP) IP6
- 3C17536 3Com Switch 8800 4-port 10GBASE-X (XFP) IP6
- 3C17533 3Com Switch 8800 24-port 1000BASE-X (XFP) IP6
- 3C17538 3Com Switch 8800 48-port 1000BASE-X (XFP) IP6
- 3C17534 3Com Switch 8800 24-port 10/100/1000BASE-T (XFP) IP6
- 3C17528 3Com Switch 8800 48-port 10/100/1000BASE-T (XFP) IP6

- 3C17532 3Com Switch 8800 48-port 10/100/1000BASE-T Access (XFP) IP6

Configuring a Service Loop Group for IPv6 Multicast and IPv6 Unicast

Configure an IPv6 multicast and/or an IPv6 unicast service loop group to implement Centralized Mode IPv6 multicast and/or IPv6 unicast. For details about a loop group, refer to *"Configuring a Service Loop Group" on page 79*.

Follow these steps to configure an IPv6 multicast service loop group:

Operation	Command	Description
Enter system view	**system-view**	-
Create a manual link aggregation group	**link-aggregation group** *agg-id* **mode manual**	Required
Configure the aggregation group as the service loop group for the specified service	**link-aggregation group** *agg-id* **service-type** { { **ipv6** | **ipv6mc** } * **| tunnel** }	Required
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Assign the Ethernet port to the service loop group	**port link-aggregation group** *agg-id*	Required
Display information about the specified service loop group	**display link-aggregation service-type** [*agg-id*]	Optional Available in any view

CAUTION:

- *Before adding an Ethernet port into a service loop group for IPv6 multicast or Iv6 unicast, be sure that this port is on an IPv6-capable module.*
- *Before adding an Ethernet port into a service loop group for IPv6 multicast or IPv6 unicast, make sure that STP is disabled on port.*

- *Only one IPv6 multicast service loop group can be configured in the system.*
- *No more than 8 ports can be added to the IPv6 multicast or Iv6 unicast service service loop group.*

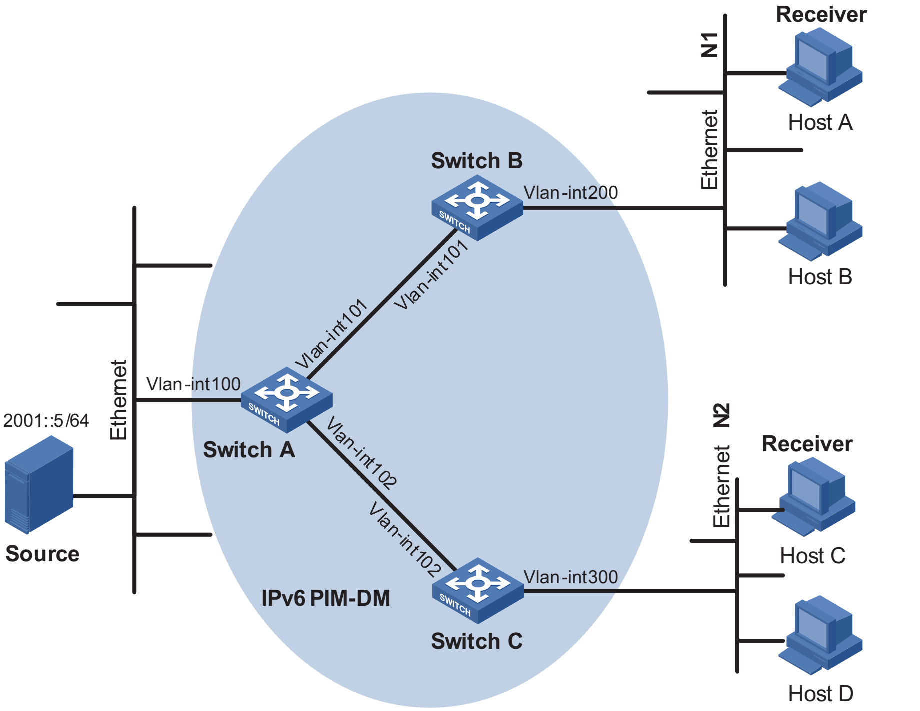
Configuring IPv6 Multicast and IPv6 Unicast Centralized Mode Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form two stub networks, and at least one receiver host exists in each stub network. The entire IPv6 PIM domain operates in the dense mode.
- Due to some restrictions, Switch B and Switch C connect to Switch A through ports on their respective IPv4 modules in slot 3. An IPv6 board is installed in slot 2 on Switch B and Switch C respectively.
- Configure Centralized Mode for IPv6 multicast and IPv6 unicast so that Switch B and Switch C can forward IPv6 multicast and IPv6 unicast data normally.

Network diagram

Figure 211 Network diagram for Centralized Mode for IPv6 multicast and IPv6 unicast



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	2001::1/64	Switch C	Vlan-int102	2003::2/64
	Vlan-int101	2002::1/64		Vlan-int300	4001::1/64
	Vlan-int102	2003::1/64			
Switch B	Vlan-int101	2002::1/64			
	Vlan-int200	3001::1/64			

Configuration procedure

1 Configure the IPv6 addresses of the VLAN interfaces of each routing switch and configure a unicast routing protocol on each VLAN interface

Configure an IPv6 address and prefix for each VLAN interface and enable OSPFv3 on each VLAN interface, as shown in Figure 211. The specific configuration is omitted.

2 Enable IPv6 multicast routing, and enable IPv6 PIM-DM on each interface

# Enable IPv6 multicast routing on Switch A, and enable IPv6 PIM-DM on each VLAN interface.

```
<SwitchA> system-view
[SwitchA] multicast ipv6 routing-enable
[SwitchA] interface vlan-interface 100
[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
[SwitchA] interface vlan-interface 102
```

```
[SwitchA-Vlan-interface102] pim ipv6 dm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to the configuration on Switch A.

3 Enable MLD on the host-side VLAN interfaces of Switch B and Switch C

# Enable IPv6 multicast routing on Switch B and enable MLDv1 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] quit
```

The configuration on Switch C is similar to the configuration on Switch B.

4 Configure a service loop group for IPv6 multicast and IPv6 unicast on Switch B and Switch C:

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type ipv6
[SwitchB] interface gigabitethernet 2/1/1
[SwitchB-GigabitEthernet2/1/1] stp disable
[SwitchB-GigabitEthernet2/1/1] port link-aggregation group 1
[SwitchB-GigabitEthernet2/1/1] quit
[SwitchB]link-aggregation group 2 mode manual
[SwitchB]link-aggregation group 2 service-type ipv6mc
[SwitchB]interface GigabitEthernet 2/1/2
[SwitchB-GigabitEthernet2/1/2]stp disable
[SwitchB-GigabitEthernet2/1/2]port link-aggregation group 2
[SwitchB-GigabitEthernet2/1/2]q
```

After the configuration mentioned above, when receiving IPv6 multicast data from Switch A, Switch B forwards the data to its IPv6-capable module, which processes and forwards the data to the downstream device.

The configuration on Switch C is similar to the configuration on Switch B.

5 Verify the configuration

Use the **display link-aggregation service-type** command to view the configuration of the service loop group.

# View the configuration of the service loop group on Switch B.

```
[SwitchB] display link-aggregation service-type
Service-Loop         Service                    Quote
  Group ID            Type                      Number
---------------------------------------------------------
     1               ipv6                          0
     2               ipv6mc                          0
[SwitchB]
```

The above information shows that a service loop group has been correctly established for IPv6 multicast and unicast.

The information on Switch C is similar to that on Switch B.

48 UDP HELPER CONFIGURATION

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

UDP Helper can be currently configured on VLAN interfaces only.

Introduction to UDP Helper

UDP Helper makes the device function as a relay that converts broadcast packets with the specified UDP destination port number into unicast packets and forwards them to a specified server.

With UDP Helper enabled, the device decides whether to forward a received UDP broadcast packet according to its UDP destination port number. If the packet needs to be forwarded, the device modifies the destination IP address in the IP header and then sends the packet to the specified destination server. Otherwise, the device sends the packet to its upper layer protocol for processing.

By default, with UDP Helper enabled, the device forwards broadcast packets with the six UDP destination port numbers listed in Table 31.

Table 31 List of default UDP ports

Protocol	UDP port number
Trivial file transfer protocol (TFTP)	69
Domain name system (DNS)	53
Time service	37
NetBIOS name service (NetBIOS-NS)	137
NetBIOS datagram service (NetBIOS-DS)	138
Terminal access controller access control system (TACACS)	49

Configuring UDP Helper

Follow these steps to configure UDP Helper:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
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** }	Optional By default, the UDP helper enabled device forwards broadcast packets with destination port numbers 69, 53, 37, 137, 138, and 49.
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure the destination server to which UDP packets are to be forwarded	**udp-helper server** *ip-address*	Required By default, no destination server is configured.

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.*
- *The* ***dns****,* ***netbios-ds****,* ***netbios-ns****,* ***tacacs****,* ***tftp****, and* ***time*** *keywords correspond to the six default UDP port numbers. You can configure these default UDP port numbers by specifying port numbers or the corresponding parameters. For example,* ***udp-helper port*** *53 and* ***udp-helper port dns*** *specify the same UDP port number.*
- *When you view the configuration information by using the* ***display current-configuration*** *command, the UDP Helper configuration of the default ports will not be displayed. UDP Helper configuration of these ports will be displayed only after UDP Helper is disabled.*
- *The configuration of all UDP ports (including the default 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.*
- *If a destination server is configured on a VLAN interface, broadcast packets with the specified UDP destination port number that are received from a VLAN port will be unicast to that destination server after UDP Helper is enabled.*

Displaying and Maintaining UDP Helper

To do...	Use the command...	Remarks
Display the information of the destination servers and the number of packets forwarded to the destination servers	**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

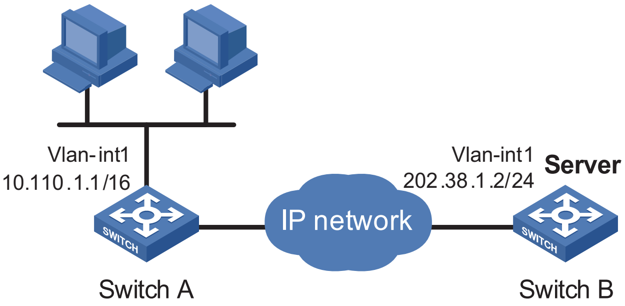
UDP Helper Configuration Examples

Network requirements

The interface VLAN-interface 1 on the UDP Helper enabled Switch A has the IP address of 10.110.1.1/16, connecting to the network segment 10.110.0.0/16. Enable the forwarding of broadcast packets with the UDP destination port number 55 to the destination server 202.38.1.2/24.

Network diagram

Figure 212 Network diagram for UDP Helper configuration



Configuration procedure

The following configuration assumes that a route from Switch A to the network segment 202.38.1.0/24 is available.

# Enable UDP Helper.

```
<SwitchA> system-view
[SwitchA] udp-helper enable
```

# Enable the forwarding of broadcast packets with the UDP destination port number 55.

```
[SwitchA] udp-helper port 55
```

# Specify the server with the IP address of 202.38.1.2 as the destination server to which UDP packets are to be forwarded.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.110.1.1 16
[SwitchA-Vlan-interface1] udp-helper server 202.38.1.2
```

49 DHCP OVERVIEW

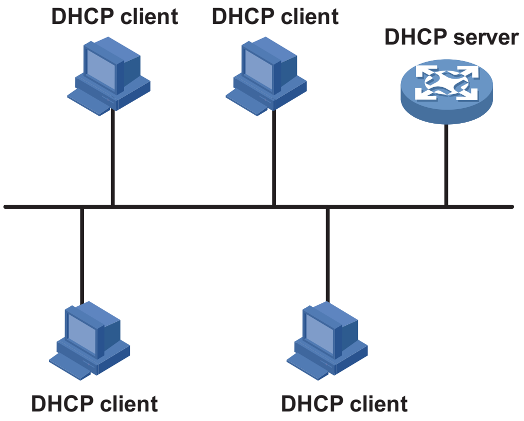
Introduction to DHCP

The fast expansion and growing complexity of networks result in scarce IP addresses assignable to hosts. Meanwhile, with the wide application of the wireless network, the frequent movement of laptops across the network requires that the IP addresses be changed accordingly. Therefore, related configurations on hosts become more complex. Dynamic host configuration protocol (DHCP) was introduced to ease network configuration by providing a framework for passing configuration information to hosts on a TCP/IP network.

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.

A typical DHCP application, as shown in Figure 213, includes a DHCP server and multiple clients (PCs and laptops).

Figure 213 A typical DHCP application



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 clients obtain their addresses in this way.

Dynamic IP Address Allocation Procedure

For dynamic allocation, 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 such as an IP address to the client in a DHCP-OFFER message.
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 to which the client sent a formal request for the offered IP address returns a DHCP-ACK message to the client, confirming that the IP address has been allocated to the client, or returns a DHCP-NAK unicast message, denying the IP address allocation.

- *After the client receives the DHCP-ACK message, it will probe whether the IP address assigned by the server is in use by broadcasting gratuitous ARP. If the client receives no response within specified time, the client can use this IP address. Otherwise, the client sends a DHCP-DECLINE message to the server to request an IP address again.*
- *If there are multiple DHCP servers, IP addresses offered by other DHCP servers are assignable to other clients.*

IP Address Lease Extension

The IP address dynamically allocated by a DHCP server to a client has a lease. After the lease duration elapses, the IP address will be reclaimed by the DHCP server. If the client wants to use the IP address again, it has to extend the lease duration.

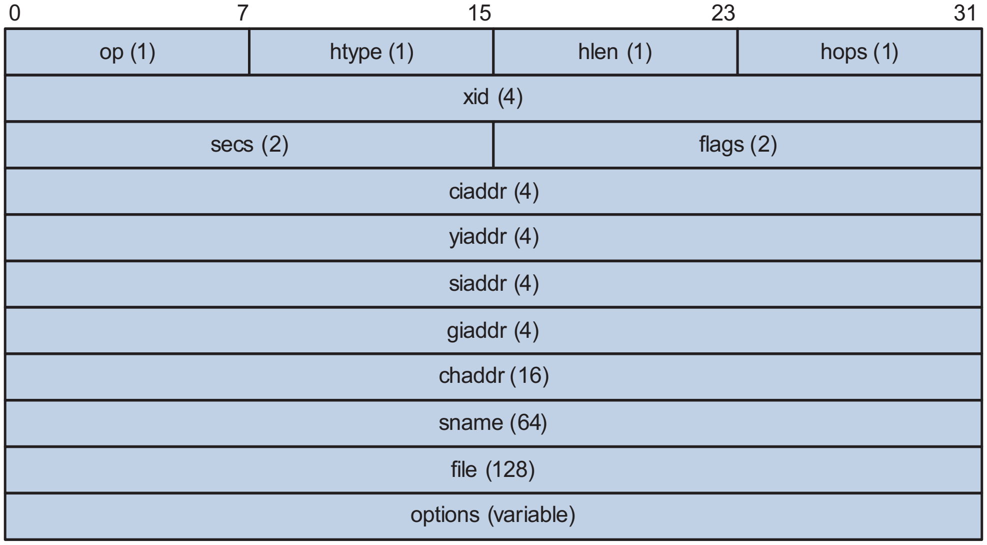
After the half lease duration elapses, the DHCP client will send the DHCP server a DHCP-REQUEST unicast message to extend the lease duration. Upon availability of the IP address, the DHCP server returns a DHCP-ACK unicast confirming that the client’s lease duration has been extended, or a DHCP-NAK unicast denying the request.

If the client receives the DHCP-NAK message, it will broadcast another DHCP-REQUEST message for lease extension after 7/8 lease duration elapses. The DHCP server will handle the request as above mentioned.

DHCP Message Format

Figure 214 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 octets.

Figure 214 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: The first relay agent IP address a request message traveled.
- chaddr: Client hardware address.
- sname: The server host name, from which the client obtained configuration parameters.
- file: Bootfile name and routing information, defined by the server to the client.
- options: Optional parameters field that is variable in length, which includes the message type, lease, DNS IP address, WINS IP address and so forth.

Protocols and Standards

- RFC2131:Dynamic Host Configuration Protocol
- RFC2132:DHCP Options and BOOTP Vendor Extensions
- RFC1542:Clarifications and Extensions for the Bootstrap Protocol

- RFC 3046: DHCP Relay Agent Information Option

50 DHCP SERVER CONFIGURATION

When configuring the DHCP server, go to these sections for information you are interested in:

The DHCP server configuration is supported only on VLAN interfaces.

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 to access 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

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 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 the domain name, 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 of the tree will be:

- Inherited if the lower level has no such configuration, or
- Overridden if the lower level has such configuration.

The IP address lease does not have any inheritance.

The DHCP server observes the following principles to select an address pool to assign IP addresses to clients:

1. If there is an address pool where IP addresses are statically bound to the MAC addresses or IDs of clients, the DHCP server will select this address pool and assign statically bound IP addresses to clients. For the configuration of this address pool, refer to "Configure manual address allocation" on page 724.
2. Otherwise, the DHCP server will select the smallest address pool that contains the IP address of the interface receiving DHCP requests, regardless of the mask. If no IP address is available in the smallest address pool, the DHCP server will fail to assign addresses to clients because it will not assign those in the father address pool to clients. For the configuration of the smallest address pool, refer to "Configure dynamic address allocation" on page 725.

   For example, two address pools are configured on the DHCP server. The ranges of IP addresses that can be dynamically assigned are 1.1.1.0/24 and 1.1.1.0/25 respectively. 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 the 1.1.1.0/25 address pool. If no IP address is available in the 1.1.1.0/25 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.

Keep the IP addresses for dynamic allocation within the subnet where the interface of the DHCP server 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 IP address manually bound to the client's MAC address or ID
2. The IP address that was ever assigned to the client
3. The IP address designated by the Option 50 field in a DHCP-DISCOVER message
4. The first IP address found in the DHCP address pool
5. The IP address that was a conflict or passed its lease duration

If no IP address is assignable, the server will not respond.

DHCP Server Configuration Task List

To configure the DHCP server feature, perform the tasks described in the following sections:

Task	Remarks
"Enabling DHCP" on page 723	Required
"Enabling the DHCP Server on an Interface" on page 723	Optional
"Configuring an Address Pool for the DHCP Server" on page 724	Optional
"Configuring the DHCP Server Security Functions" on page 730	Optional
"Enabling the DHCP Server to Support Option 82" on page 731	Optional

Enabling DHCP

Enable DHCP before performing other configurations.

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 VLAN interface view	**interface** *interface-type interface-number*	-
Enable the DHCP server on an interface	**dhcp select server global-pool** [**subaddress**]	Optional Enabled by default.

*The **subaddress** keyword is valid only when the server and client are on the same subnet. If a DHCP relay agent exists in between, regardless of **subaddress**, the DHCP server will select an IP address from the address pool of the subnet which contains the primary IP address of the DHCP relay agent's interface (connected to the client).*

When the DHCP server and client are on the same subnet, the server will:

- *With **subaddress** specified, assign an IP address from the address pool of the subnet which the secondary IP address of the server's interface connected to the client belongs to, or assign from the first secondary IP address if several secondary IP addresses exist. If no secondary IP address is configured for the interface, the server is unable to assign an IP address to the client.*

- *Without **subaddress** specified, assign an IP address from the address pool of the subnet which the primary IP address of the server's interface (connected to the client) belongs to.*

Configuring an Address Pool for the DHCP Server

Configuration Task List

To configure an address pool, perform the tasks described in the following sections:

Task		Remarks
"Creating a DHCP Address Pool" on page 724		Required
"Configuring an Address Allocation Mechanism" on page 724	"Configure manual address allocation" on page 724	Required to configure either of the two
	"Configure dynamic address allocation" on page 725	
"Configuring a Domain Name for the Client" on page 726		Optional
"Configuring DNS Servers for the Client" on page 726		
"Configuring WINS Servers and NetBIOS Node Type for the Client" on page 726		
"Configuring the BIMS server Information for the Client" on page 727		
"Configuring Gateways for the Client" on page 728		
"Configuring the TFTP Server Address and Bootfile Name for the Client" on page 728		
"Configuring Self-Defined DHCP Options" on page 729		

Creating a DHCP Address Pool

To create a DHCP address pool, use the following commands:

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*	Required No DHCP address pool is created by default.

Configuring an Address Allocation Mechanism

CAUTION: *You can configure either the static binding or dynamic address allocation for an 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.

Configure 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.

To configure the static binding in a DHCP address pool, use the following commands:

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Enter DHCP address pool view		**dhcp server ip-pool** *pool-name*	-
Bind IP addresses statically		**static-bind ip-address** *ip-address* [*mask-length* | **mask** *mask*]	Required No IP addresses are statically bound by default
Bind MAC addresses or IDs statically	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 ID	**static-bind client-identifier** *client-identifier*	

- *Use the* ***static-bind ip-address*** *command together with* ***static-bind mac-address*** *or* ***static-bind client-identifier*** *command to accomplish a static binding configuration.*
- *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.*

Configure 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 GW, FTP server and so forth 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.

To configure the dynamic address allocation, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter DHCP 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, meaning no assignable address
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 All addresses in the DHCP address pool assignable by default.

- *In DHCP address pool view, using the **network** command repeatedly overwrites the previous configuration.*
- *Using the **dhcp server forbidden-ip** command repeatedly can specify multiple IP address ranges not assignable.*

Configuring a Domain Name for the Client

You can specify a domain name in each DHCP address pool on the DHCP server. To configure the domain name in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter the DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
Specify the domain name 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.

To configure DNS servers in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
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 Windows 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, 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, then broadcasts it to get the destination IP address.

To configure WINS servers and NetBIOS node type in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
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

If b-node is specified for the client, you need to specify no 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.

To configure the BIMS server IP address, port number, and shared key in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
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 wanting to access hosts outside the local subnet request gateways to forward data. You can specify gateways in each address pool for clients. Up to eight gateways can be specified in a DHCP address pool.

To configure the gateways in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
Specify gateways	**gateway-list** *ip-address*&<1-8>	Required No gateway is specified by default.

Configuring the TFTP Server Address 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 autoconfiguration. The request process of the client is described below:

When a router starts up with an empty configuration file, the system sets the specified interface (Vlan-interface1) as the DHCP client to request from the DHCP server parameters such as the IP address and name of a TFTP server, bootfile name.

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.

Note that if the client cannot get related parameters, it will use the empty configuration file for system initialization.

When option 55 in the requesting client message contains parameters of option 66, option 67, or option 150, the DHCP server will return the IP address and name of the specified TFTP server, and bootfile name to the client.

To configure the IP address and name of the TFTP server and the bootfile name in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-

To do...	Use the command...	Remarks
Specify the TFTP server	**tftp-server ip-address** *ip-address*	Optional Not specified by default
Specify the name of the TFTP server	**tftp-server domain-name** *domain-name*	Optional Not specified by default
Specify the bootfile name	**bootfile-name** *bootfile-name*	Optional 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.
- Expand existing DHCP options. When the current DHCP options cannot meet the customer's requirements (for example, you cannot use the **dns-list** command to configure more than eight DNS server addresses), you can expand these options.

To configure a self-defined DHCP option in the DHCP address pool, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter DHCP address pool view	**dhcp server ip-pool** *pool-name*	-
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

CAUTION:

- *Be careful when configuring self-defined DHCP options because the configuration of these options may affect the DHCP operation process.*
- *When you use self-defined option (Option 51) to configure the IP address lease duration, convert the lease duration into seconds in hexadecimal notation.*

Table 32 Description of common options

Option	Corresponding Name in RFC	Corresponding command	Command option
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**
51	IP Address Lease Time	**expired**	**hex**
58	Renewal (T1) Time Value	**expired**	**hex**

Table 32 Description of common options

Option	Corresponding Name in RFC	Corresponding command	Command option
59	Rebinding (T2) Time Value	**expired**	**hex**
66	TFTP server name	**tftp-server**	**ascii**
67	Bootfile name	**bootfile-name**	**ascii**

Configuring the DHCP Server Security Functions

This configuration is necessary to secure DHCP services on the DHCP server.

Configuration Prerequisites

Before performing this configuration, you have finished the configuration tasks of the DHCP server.

Enabling Unauthorized DHCP Server Detection

There are unauthorized DHCP servers on networks, which reply DHCP clients with wrong IP addresses.

With this feature enabled, when receiving a DHCP message with the siaddr field not being 0 from a client, the DHCP server will record the value of the siaddr field in the message and the receiving interface. The administrator can use this information to check out any DHCP unauthorized servers.

To enable unauthorized DHCP server detection, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable unauthorized DHCP server detection	**dhcp server detect**	Required Disabled by default

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 via 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 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, the server will assign the IP address to the requesting client (The DHCP client probes the IP address by sending gratuitous ARP packets).

To configure IP address conflict detection, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
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.

Enabling the DHCP Server to Support Option 82

When the DHCP server receives a message with Option 82 from a relay agent, if the server supports Option 82, it will assign an IP address to the requesting client, and if the server does not support Option 82, it will ignore the message.

Configuration prerequisites

Before performing this configuration, you have finished the configuration tasks of the DHCP server.

Enabling the DHCP server to support Option 82

To enable the DHCP server to support Option 82, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the server to support Option 82	**dhcp server relay information enable**	Optional Enabled by default

To support Option 82, it is required to perform configuration on both the DHCP server and relay agent. Refer to "Configuring the DHCP Relay Agent to Support Option 82" on page 740 for related configuration details.

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** { **ip** *ip-address* | **pool** [*pool-name*] | **all** }	
Display information about assignable IP addresses	**display dhcp server free-ip**	
Display IP addresses excluded from dynamic allocation in the DHCP address pool	**display dhcp server forbidden-ip**	
Display information about bindings	**display dhcp server ip-in-use** { **ip** *ip-address* | **pool** [*pool-name*] | **all** }	
Display information about DHCP server statistics	**display dhcp server statistics**	
Display information about the address pool tree organization	**display dhcp server tree** { **pool** [*pool-name*] | **all** }	
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** { **ip** *ip-address* | **pool** [*pool-name*] | **all** }	
Clear information about DHCP server statistics	**reset dhcp server statistics**	

*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 Example

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.

Network requirements

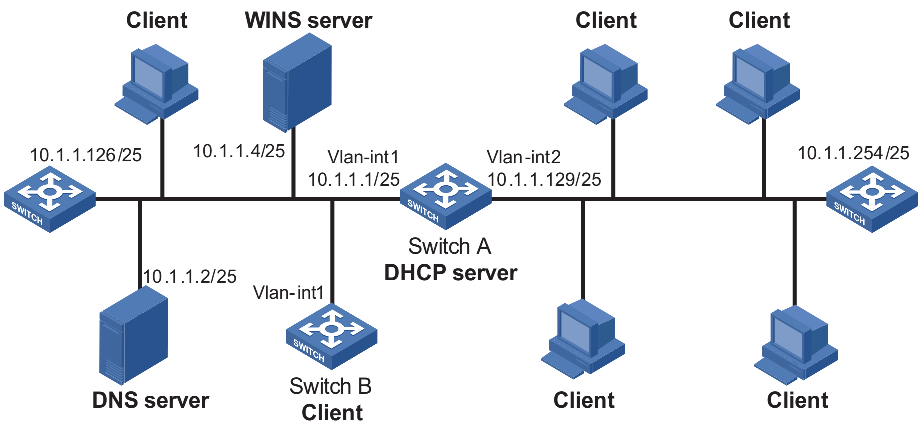
- The DHCP server (Switch A) assigns IP address to clients in subnet 10.1.1.0/24, which is subnetted into 10.1.1.0/25 and 10.1.1.128/25.
- The IP addresses of VLAN interfaces 1 and 2 on Switch A are 10.1.1.1/25 and 10.1.1.129/25 respectively.

- In the address pool 10.1.1.0/25, the address lease duration is ten days and twelve hours, domain name aabbcc.com, DNS server address 10.1.1.2, gateway 10.1.1.126, and WINS server 10.1.1.4.
- In the address pool 10.1.1.128/25, the address lease duration is five days, domain name aabbcc.com, DNS server address 10.1.1.2, and gateway address 10.1.1.254, and there is no WINS server address.
- The domain name and DNS server address on the subnets 10.1.1.0/25 and 10.1.1.128/25 are the same. Therefore, a domain name and DNS server address can be configured only for the subnet 10.1.1.0/24, and the subnet 10.1.1.0/25 and 10.1.1.128/25 can inherit the configuration of the subnet 10.1.1.0/24.

In this example, the number of requesting clients connected to Vlan-interface1 should not exceed 122, and that of clients connected to Vlan-interface2 should not exceed 124.

Network diagram

Figure 215 A DHCP network



Configuration procedure

1 Specify VLAN interfaces and IP addresses for VLAN interfaces (omitted).

2 Configure the DHCP server

# Enable DHCP

```
<Sysname> system-view
[Sysname] dhcp enable
```

# Exclude IP addresses (addresses of the DNS server, WINS server and gateways).

```
[Sysname] dhcp server forbidden-ip 10.1.1.2
[Sysname] dhcp server forbidden-ip 10.1.1.4
[Sysname] dhcp server forbidden-ip 10.1.1.126
[Sysname] dhcp server forbidden-ip 10.1.1.254
```

# Configure DHCP address pool 0 (address range, client domain name, and DNS server address).

```
[Sysname] dhcp server ip-pool 0
[Sysname-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[Sysname-dhcp-pool-0] domain-name aabbcc.com
[Sysname-dhcp-pool-0] dns-list 10.1.1.2
[Sysname-dhcp-pool-0] quit
```

# Configure DHCP address pool 1 (address range, gateway, lease duration, and WINS server).

```
[Sysname] dhcp server ip-pool 1
[Sysname-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.128
[Sysname-dhcp-pool-1] gateway-list 10.1.1.126
[Sysname-dhcp-pool-1] expired day 10 hour 12
[Sysname-dhcp-pool-1] nbns-list 10.1.1.4
[Sysname-dhcp-pool-1] quit
```

# Configure DHCP address pool 2 (address range, gateway, and lease duration).

```
[Sysname] dhcp server ip-pool 2
[Sysname-dhcp-pool-2] network 10.1.1.128 mask 255.255.255.128
[Sysname-dhcp-pool-2] expired day 5
[Sysname-dhcp-pool-2] gateway-list 10.1.1.254
```

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 into DOS window. Type **ipconfig/release** to relinquish the IP address and then **IPconfig/renew** to obtain another IP address.

51 DHCP Relay Agent Configuration

When configuring the DHCP relay agent, go to these sections for information you are interested in:

The DHCP relay agent configuration is supported only on VLAN interfaces.

Introduction to DHCP Relay Agent

Application Environment

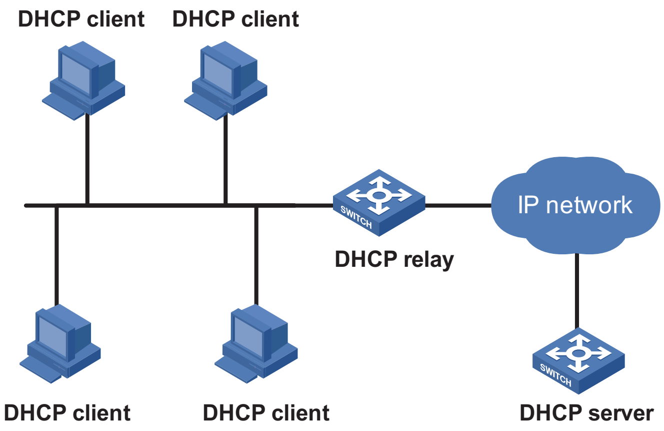
Since DHCP clients request IP addresses via broadcast messages, the DHCP server and clients must be on the same subnet. Therefore, a DHCP server must be available on each subnet. It is not practical.

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 216 shows a typical application of the DHCP relay agent.

Figure 216 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 "Dynamic IP Address Allocation Procedure" on page 718). The following describes the forwarding process on the DHCP relay agent.

1 After receiving a DHCP-DISCOVER or DHCP-REQUEST broadcast message from a DHCP client, the DHCP relay agent forwards the message to the designated DHCP server in unicast mode.

2 The DHCP server returns an IP address to the relay agent, which conveys it to the client via broadcast.

Configuring DHCP Relay Agent

DHCP Relay Agent Configuration Task List

Complete the following tasks to configure the DHCP relay agent:

Task	Remarks
"Enabling DHCP" on page 736	Required
"Enabling the DHCP Relay Agent on Interfaces" on page 736	Required
"Correlating a DHCP Server Group with Relay Agent Interfaces" on page 737	Required
"Configuring the DHCP Relay Agent to Send a DHCP-Release Request" on page 737	Optional
"Configuring the DHCP Relay Agent Security Functions" on page 738	Optional
"Configuring the DHCP Relay Agent to Support Option 82" on page 740	Optional

Enabling DHCP

Enable DHCP before performing other DHCP-related configurations.

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 Interfaces

With this task completed, upon receiving a DHCP request from an enabled interface, the relay agent will forward the request to a DHCP server for address allocation.

To enable the DHCP relay agent on interfaces, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**Interface** *interface-type interface-number*	-

To do...	Use the command...	Remarks
Enable the DHCP relay agent on the current interface	**dhcp select relay**	Required With DHCP enabled, interfaces work in the DHCP server mode.

If the DHCP client obtains an IP address via the DHCP relay agent, the address pool of the subnet which the IP address of the DHCP relay agent belongs to must be configured on the DHCP server. Otherwise, the DHCP client cannot obtain a correct IP address.

Correlating a DHCP Server Group with Relay Agent Interfaces

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.

To correlate a DHCP server group with relay agent interfaces, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Specify a DHCP server group number and servers in the group	**dhcp relay server-group** *group-id* **ip** *ip-address*	Required Not specified by default
Enter VLAN interface view	**interface** *interface-type interface-number*	-
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.

- *You can specify at most twenty DHCP server groups on Switch 8800s .*
- *Up to eight DHCP server addresses can be configured for each DHCP server group. The IP addresses of DHCP servers that belong to the DHCP server group 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 in the **dhcp relay server-select** command was specified by the **dhcp relay server-group** command.*

Configuring the DHCP Relay Agent to Send a DHCP-Release Request

Sometimes, you need to release a client's IP address manually on the DHCP relay agent. With this task completed, the DHCP relay agent can actively send 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.

With this feature enabled in system view, the DHCP-RELEASE request will be sent to those DHCP servers correlated with the DHCP relay agent interfaces.

To configure the DHCP relay agent in system view to send a DHCP-RELEASE request, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the DHCP relay agent in system view to send a DHCP-RELEASE request	**dhcp relay release ip** *client-ip*	Required

Configuring the DHCP Relay Agent Security Functions

Create static bindings and enable invalid IP address check

The DHCP relay agent dynamically records clients' IP-to-MAC bindings to generate a dynamic binding after clients got IP addresses. It also supports static binding, 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 on it (both dynamic and static bindings). If not, the client cannot access outside networks via the DHCP relay agent.

To create a static binding and enable invalid IP address check, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a static binding	**dhcp relay security static** *ip-address mac-address*	Optional No static binding is created by default
Enter VLAN interface view	**interface** *vlan-interface interface-number*	-
Enable invalid IP address check	**dhcp relay address-check** { **enable | disable** }	Required Disabled by default

- *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.*
- *Before executing the* ***dhcp relay address-check enable*** *command on the DHCP relay interface connected to the DHCP server, you need to configure the static binding between the IP address and MAC address of the DHCP server. Otherwise, the DHCP client will fail to obtain an IP address.*

Configure 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, the DHCP relay agent can update dynamic bindings at a specified interval.

The DHCP relay agent use its own MAC address and the IP address to be assigned to a client to regularly send a DHCP-REQUEST message to the DHCP server. If the server returns a DHCP-ACK message, which means IP address to be assigned to the client is assignable now, the DHCP relay agent will update its bindings by aging out the binding entry of the client's 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.

To configure dynamic binding update interval, use the following commands:

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)

A large number of binding entries may result in a slow refreshing speed, so you are recommended to use the default refreshing interval.

Enable unauthorized DHCP servers detection

There are invalid DHCP servers on networks, which reply DHCP clients with wrong IP addresses. These invalid DHCP servers are unauthorized DHCP servers.

With this feature enabled, upon receiving a DHCP message with the siaddr field (IP address of the server assigning IP addresses to clients) not being 0 from a client, the DHCP relay agent will record the value of the siaddr field and the information on the interface receiving the DHCP message. The administrator can use this information to check out any DHCP unauthorized servers.

To enable unauthorized DHCP server detection, use the following commands:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable unauthorized DHCP server detection	**dhcp relay server-detect**	Required Disabled by default

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 recorded information of a DHCP server is cleared, a new record will be put for the DHCP server.

Configuring the DHCP Relay Agent to Support Option 82

Introduction to Option 82

Option 82 is the relay agent option in the Options field of the DHCP message. It records the location information of the DHCP client. When a DHCP relay agent receives a client's request, it adds Option 82 to the request message so that the administrator can locate the DHCP client to further implement security control and accounting.

Option 82 involves at most 255 sub-options. At least one sub-option must be defined. Now Switch 8800s support two sub-options: sub-option 1 and sub-option 2.

Option 82 has no unified definition. Its padding formats vary with vendors. Currently, two padding formats are supported: normal and verbose.

The padding contents for sub-options in the normal padding format are:

sub-option 1: Padded with the VLAN ID and interface number related to the interface that received the client's request.

sub-option 2: Padded with the MAC address of the interface that received the client's request.

The padding contents for sub-options in the verbose padding format are:

sub-option 1: Padded with specified access node identifier, type of the interface that received the client's request, interface number, PVC identifier (used when the interface type is ATM), and VLAN ID.

sub-option 2: Padded with the MAC address of the interface that received the client's request.

Handling strategies for Option 82 on the relay agent

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 Option 82 before forwarding the reply to the client.

If a client's request message has...	Handling strategy	Padding format	The DHCP relay agent will...
Option 82	Drop	-	Drop the message.
	Keep	-	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.

If a client's request message has...	Handling strategy	Padding format	The DHCP relay agent will...
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.

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

Use the following commands for this configuration:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface vlan-interface** *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 the padding format for Option 82	**dhcp relay information format** { **normal** | **verbose** [**node-identifier** { **mac** | **sysname** | **user-defined** *node-identifier* }] }	Optional **normal** by default

- *To support Option 82, it is required to perform related configuration on both the DHCP server and relay agent. Refer to "Enabling the DHCP Server to Support Option 82" on page 731 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** { **interface** *interface-type interface-number* | **all** }	Available in any view
Display information about bindings of DHCP relay agents	**display dhcp relay security** [**dynamic** | **static** | *ip-address*]	Available in any view

To do...	Use the command...	Remarks
Display statistics information about bindings of DHCP relay agents	**display dhcp relay security statistics**	Available in any view
Display information about the refreshing interval for entries of dynamic IP-to-MAC bindings	**display dhcp relay security tracker**	Available in any view
Display information about the configuration of a specified or all DHCP server groups	**display dhcp relay server-group** { *group-id* | **all** }	Available in any view
Display packet statistics on relay agent	**display dhcp relay statistics** [**server-group** { *group-id* | **all** }]	Available in user view
Clear packet statistics from relay agent	**reset dhcp relay statistics** [**server-group** *group-id*]	Available in user view

DHCP Relay Agent Configuration Example

Prerequisites

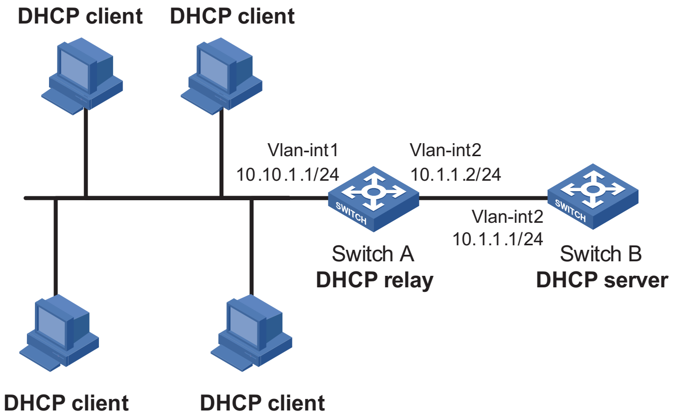
- Vlan-interface1 on the DHCP relay agent (Switch A) connects to the network where DHCP clients reside. The IP address of Vlan-interface1 is 10.10.1.1/24 and IP address of Vlan-interface2 is 10.1.1.2/24.
- The IP address of the DHCP server 10.1.1.1/24.
- Switch A forwards messages between DHCP clients and the DHCP server, so that the DHCP client can obtain an IP address of the network segment 10.10.1.0/24 and related configuration information from the DHCP server.

Network requirements

Enable Switch A to forward DHCP messages so that the DHCP clients can obtain IP addresses and corresponding configuration information from the DHCP server.

Network diagram

Figure 217 Network diagram for DHCP relay agent



Configuration procedure

# Enable DHCP.

```
<Sysname> system-view
[Sysname] dhcp enable
```

# Enable the DHCP relay agent on Vlan-interface1.

```
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] dhcp select relay
[Sysname-Vlan-interface1] quit
```

# Configure DHCP server group 1 with the DHCP server 10.1.1.1, and correlate the DHCP server group 1 with Vlan-interface1.

```
[Sysname] dhcp relay server-group 1 ip 10.1.1.1
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] dhcp relay server-select 1
```

- *Performing the configuration on the DHCP server is also required to guarantee the client-server communication via the relay agent. Since the DHCP server configuration varies with devices, it is not mentioned here.*
- *If the DHCP relay agent and server are on different subnets, routes in between must be reachable.*

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.

52 DNS CONFIGURATION

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

- "DNS Overview" on page 745
- "Configuring Static Domain Name Resolution" on page 747
- "Configuring Dynamic Domain Name Resolution" on page 747
- "Displaying and Maintaining DNS" on page 747
- "DNS Configuration" on page 745
- "Troubleshooting DNS Configuration" on page 751

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. Each time the DNS server receives a name query it checks its static DNS database before looking up the dynamic DNS database. Reduction of the searching time in the dynamic DNS database would increase efficiency. Some frequently used addresses can be put in the static DNS database.

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 DNS database 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 in 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 DNS server. This process continues until a result, whether success or failure, is returned.
4. The DNS client returns the resolution result to the application after receiving a response from the DNS server.

Figure 218 Dynamic domain name resolution

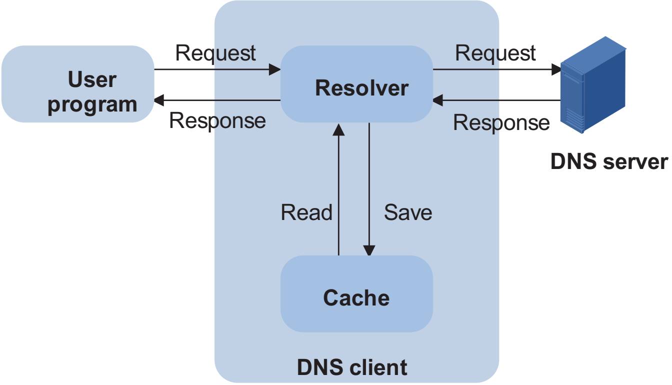


Figure 218 shows the relationship between user program, DNS client, and DNS server.

The resolver and cache comprise the DNS client. The user program and DNS client can run on the same machine or different machines, while the DNS server and the DNS client usually must run on different machines.

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 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 as a host name and add a DNS suffix before query. The original domain name (for example, aabbcc) is used if the query fails.
- 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 as a fully qualified domain name and return the query result, success or a failure. Hence, the dot "." at the end of the domain name is called the terminating symbol.

Currently, the device supports static and dynamic DNS services.

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.

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 host name and IP address in the static DNS database	**ip host** *hostname ip-address*	Required No mapping between host name and IP address is configured in the static DNS database by default.

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
Configure an IP address for the DNS server	**dns server** *ip-address*	Required No IP address is configured for the DNS server by default.
Configure DNS suffixes	**dns domain** *domain-name*	Optional No DNS suffix is configured by default

- *You may configure up to six DNS servers and ten DNS suffixes.*
- *You can use the **dns domain** command to configure a DNS suffix with the maximum length of 238 characters. Since a valid DNS suffix is a character string separated by dots, with each separated part (label) containing no more than 63 characters, any part exceeding this length may result in failure to generate packets.*

Displaying and Maintaining DNS

To do...	Use the command...	Remarks
Display the static DNS database	**display ip host**	Available in any view
Display the DNS server information	**display dns server** [**dynamic**]	Available in any view
Display the DNS suffixes	**display dns domain** [**dynamic**]	Available in any view
Display the information in the dynamic domain name cache	**display dns dynamic-host**	Available in any view

To do...	Use the command...	Remarks
Clear the information in the dynamic domain name cache	**reset dns dynamic-host**	Available in user view

DNS Configuration Example

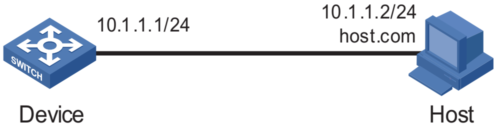
Static DNS Configuration Example

Network requirements

Device uses the static domain name resolution to access Host with IP address 10.1.1.2 through domain name host.com.

Network diagram

Figure 219 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 device 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=255 time=1 ms
    Reply from 10.1.1.2: bytes=56 Sequence=2 ttl=255 time=4 ms
    Reply from 10.1.1.2: bytes=56 Sequence=3 ttl=255 time=3 ms
    Reply from 10.1.1.2: bytes=56 Sequence=4 ttl=255 time=2 ms
    Reply from 10.1.1.2: bytes=56 Sequence=5 ttl=255 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
```

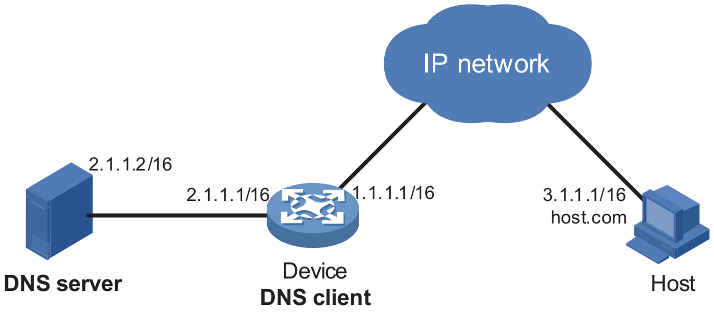
Dynamic DNS Configuration Example

Network requirements

- The IP address of the DNS server is 2.1.1.2/16 and the DNS suffix is com.
- Device serving as a DNS client uses the dynamic domain name resolution and DNS suffix to access the host with the domain name being host.com and the IP address 3.1.1.1/16.

Network diagram

Figure 220 Network diagram for dynamic domain name resolution



Configuration procedure

- *Before performing the following configuration, make sure that there is a route between the device and the host, and configurations are done on both the device and the host. For the IP addresses of the interfaces, see Figure 220.*
- *This configuration may vary with different DNS servers. The following configuration is performed on Windows 2000 server.*

1 Configure DNS server

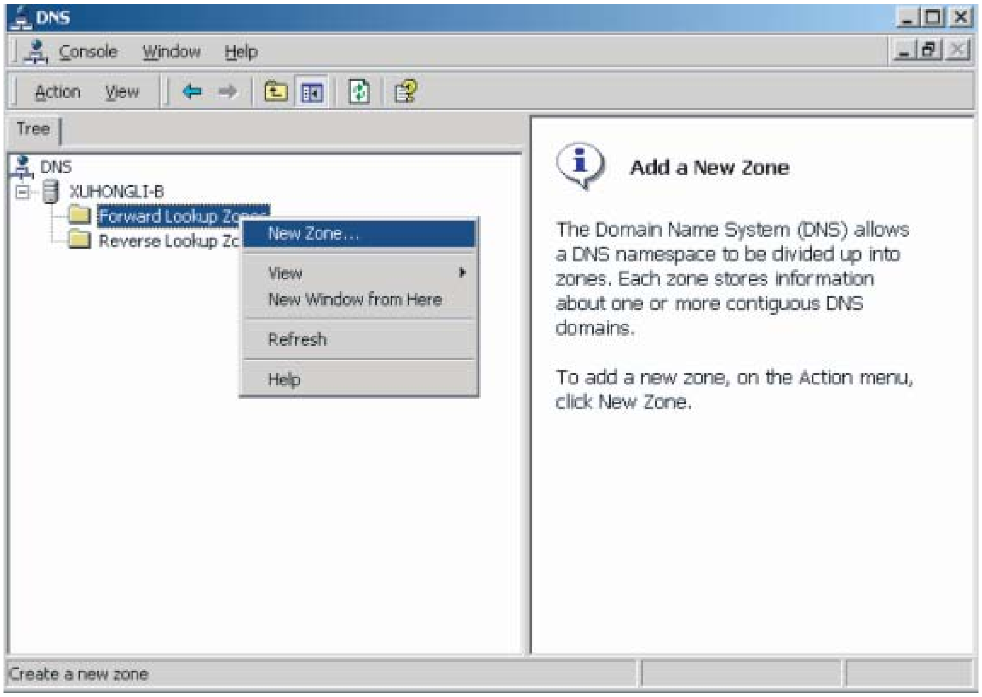
# Enter DNS server configuration page.

Select **Start** > **Programs** > **Administrative Tools** > **DNS**.

# Create zone com.

In Figure 221, right click **Forward Lookup Zones**, select **New zone**, and then follow the instructions to create a new zone com.

Figure 221 Create a zone



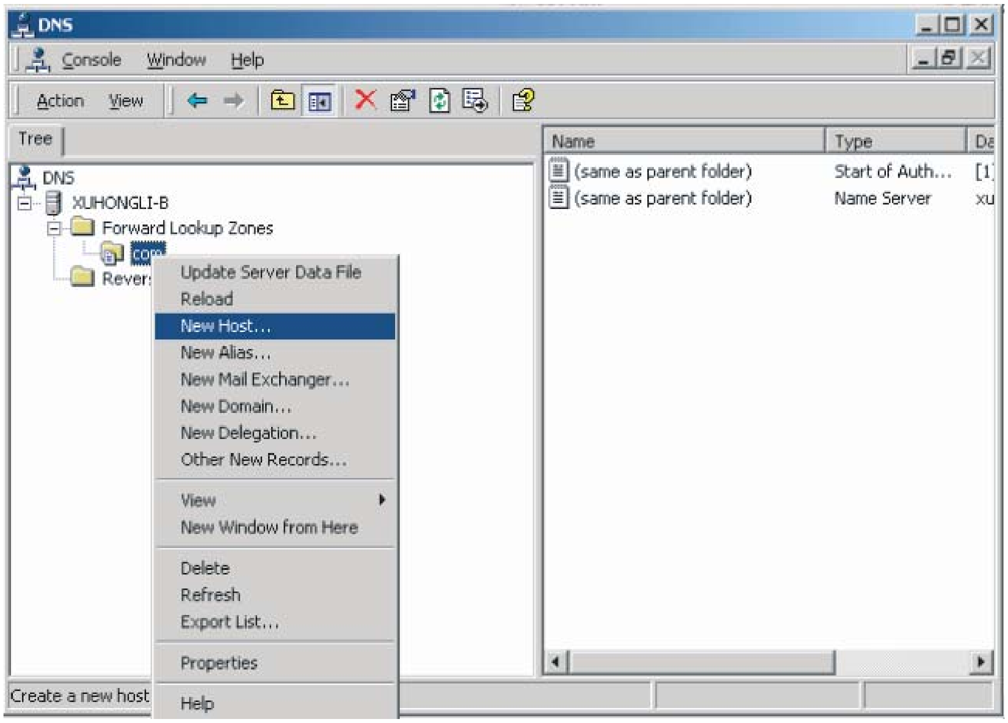
Add a New Zone

The Domain Name System (DNS) allows a DNS namespace to be divided up into zones. Each zone stores information about one or more contiguous DNS domains.

To add a new zone, on the Action menu, click New Zone.

# Create a mapping between host name and IP address.

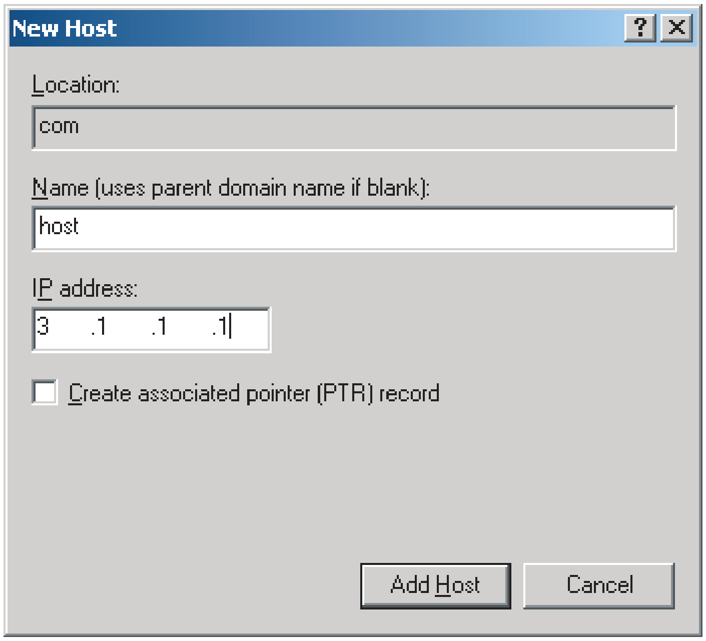
Figure 222 Add a host



Name	Type	Da
(same as parent folder)	Start of Auth...	[1]
(same as parent folder)	Name Server	xu

In Figure 222, right click zone **com**, and then select **New Host** to bring up a dialog box as shown in Figure 223. Enter host name host and IP address 3.1.1.1.

Figure 223 Add a mapping between domain name and IP address



Location:
com

Name (uses parent domain name if blank):
host

IP address:
3 .1 .1 .1

□ Create associated pointer (PTR) record

2 Configure DNS client Device

# Enable dynamic domain name resolution.

```
<Sysname> system-view
[Sysname] dns resolve
```

# Configure IP address 2.1.1.2 for the DNS server

```
[Sysname] dns server 2.1.1.2
```

# Configure com as the DNS suffix

```
[Sysname] dns domain com
```

Execute the **ping host** command on the device to verify that the communication between the device and the host is normal and that the corresponding destination IP address is 3.1.1.1.

Troubleshooting DNS Configuration

Symptom

After enabling the dynamic domain name resolution, the user cannot get the correct IP address.

Analysis

The DNS client should be used in cooperation with the DNS server to obtain a correct IP address through DNS resolution.

Solution

- Use the **display dns dynamic-host** command to check that the specified domain name is in the cache.

- If there is no defined domain name, check that dynamic domain name resolution is enabled and the DNS client can communicate with the DNS server.
- Check the mapping between the domain name and IP address is correct on the DNS server.

53 DNS CONFIGURATION

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

- "DNS Overview" on page 753
- "Configuring Static Domain Name Resolution" on page 755
- "Configuring Dynamic Domain Name Resolution" on page 755
- "Displaying and Maintaining DNS" on page 755
- "DNS Configuration" on page 753
- "Troubleshooting DNS Configuration" on page 759

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. Each time the DNS server receives a name query it checks its static DNS database before looking up the dynamic DNS database. Reduction of the searching time in the dynamic DNS database would increase efficiency. Some frequently used addresses can be put in the static DNS database.

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 DNS database 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 in 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 DNS server. This process continues until a result, whether success or failure, is returned.
4. The DNS client returns the resolution result to the application after receiving a response from the DNS server.

Figure 224 Dynamic domain name resolution

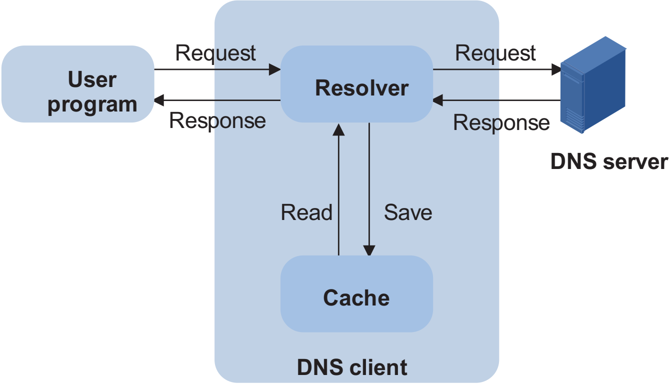


Figure 224 shows the relationship between user program, DNS client, and DNS server.

The resolver and cache comprise the DNS client. The user program and DNS client can run on the same machine or different machines, while the DNS server and the DNS client usually must run on different machines.

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 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 as a host name and add a DNS suffix before query. The original domain name (for example, aabbcc) is used if the query fails.
- 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 as a fully qualified domain name and return the query result, success or a failure. Hence, the dot "." at the end of the domain name is called the terminating symbol.

Currently, the device supports static and dynamic DNS services.

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.

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 host name and IP address in the static DNS database	**ip host** *hostname ip-address*	Required No mapping between host name and IP address is configured in the static DNS database by default.

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
Configure an IP address for the DNS server	**dns server** *ip-address*	Required No IP address is configured for the DNS server by default.
Configure DNS suffixes	**dns domain** *domain-name*	Optional No DNS suffix is configured by default

- *You may configure up to six DNS servers and ten DNS suffixes.*
- *You can use the* ***dns domain*** *command to configure a DNS suffix with the maximum length of 238 characters. Since a valid DNS suffix is a character string separated by dots, with each separated part (label) containing no more than 63 characters, any part exceeding this length may result in failure to generate packets.*

Displaying and Maintaining DNS

To do...	Use the command...	Remarks
Display the static DNS database	**display ip host**	Available in any view
Display the DNS server information	**display dns server** [**dynamic**]	Available in any view
Display the DNS suffixes	**display dns domain** [**dynamic**]	Available in any view
Display the information in the dynamic domain name cache	**display dns dynamic-host**	Available in any view

To do...	Use the command...	Remarks
Clear the information in the dynamic domain name cache	**reset dns dynamic-host**	Available in user view

DNS Configuration Example

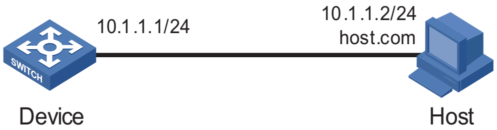
Static DNS Configuration Example

Network requirements

Device uses the static domain name resolution to access Host with IP address 10.1.1.2 through domain name host.com.

Network diagram

Figure 225 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 device 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=255 time=1 ms
    Reply from 10.1.1.2: bytes=56 Sequence=2 ttl=255 time=4 ms
    Reply from 10.1.1.2: bytes=56 Sequence=3 ttl=255 time=3 ms
    Reply from 10.1.1.2: bytes=56 Sequence=4 ttl=255 time=2 ms
    Reply from 10.1.1.2: bytes=56 Sequence=5 ttl=255 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
```

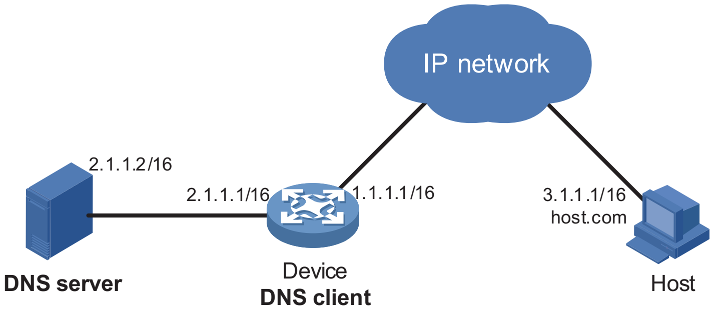
Dynamic DNS Configuration Example

Network requirements

- The IP address of the DNS server is 2.1.1.2/16 and the DNS suffix is com.
- Device serving as a DNS client uses the dynamic domain name resolution and DNS suffix to access the host with the domain name being host.com and the IP address 3.1.1.1/16.

Network diagram

Figure 226 Network diagram for dynamic domain name resolution



Configuration procedure

- *Before performing the following configuration, make sure that there is a route between the device and the host, and configurations are done on both the device and the host. For the IP addresses of the interfaces, see Figure 226.*
- *This configuration may vary with different DNS servers. The following configuration is performed on Windows 2000 server.*

1 Configure DNS server

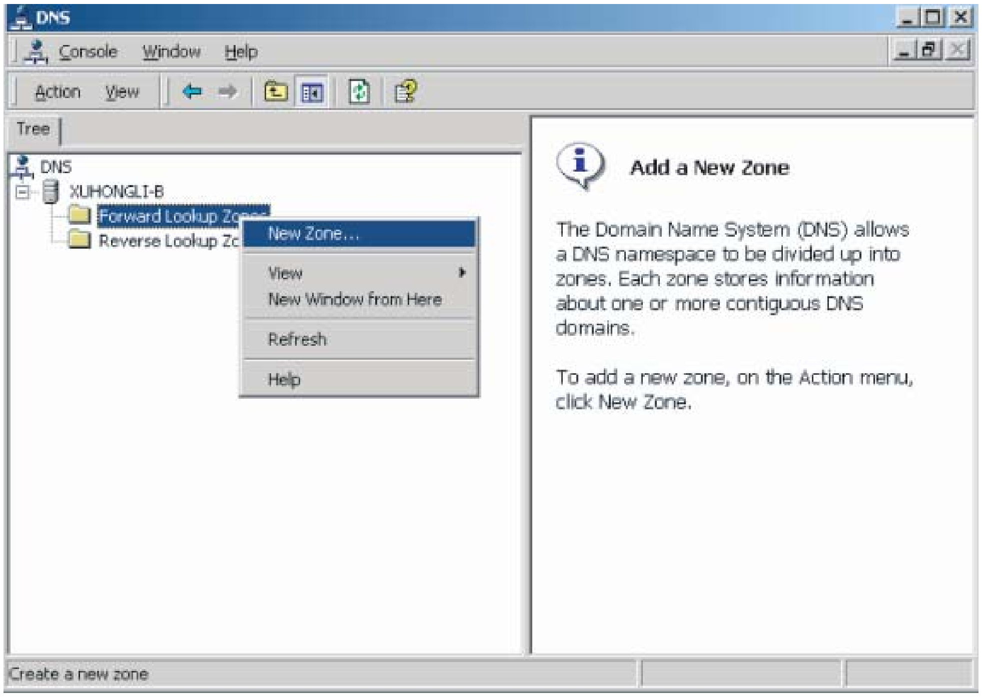
# Enter DNS server configuration page.

Select **Start** > **Programs** > **Administrative Tools** > **DNS**.

# Create zone com.

In Figure 227, right click **Forward Lookup Zones**, select **New zone**, and then follow the instructions to create a new zone com.

Figure 227 Create a zone



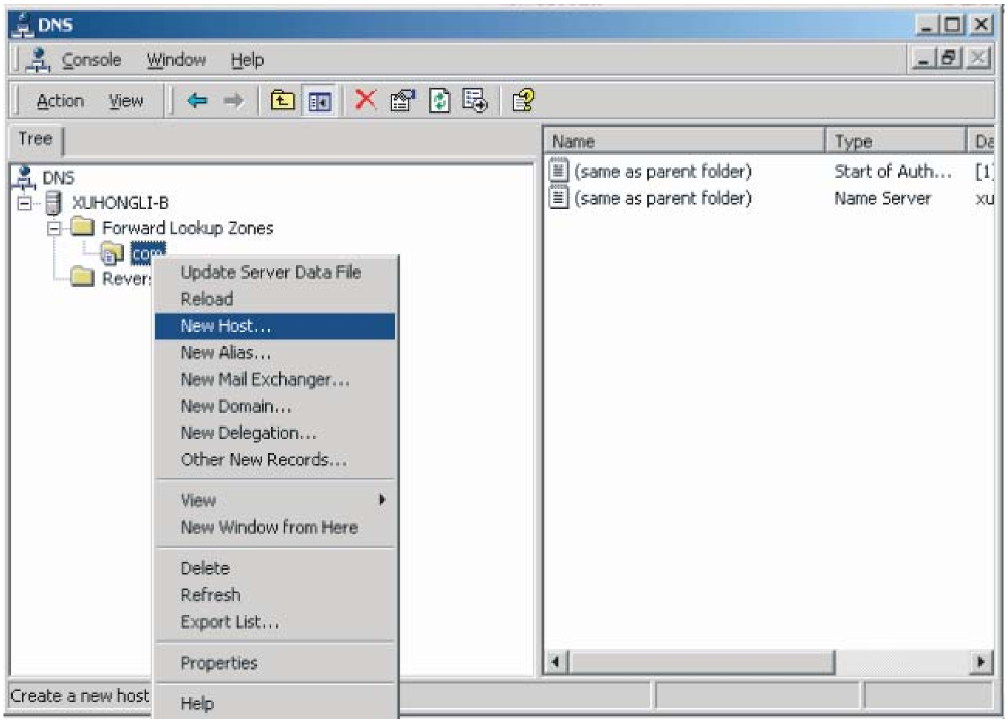
Add a New Zone

The Domain Name System (DNS) allows a DNS namespace to be divided up into zones. Each zone stores information about one or more contiguous DNS domains.

To add a new zone, on the Action menu, click New Zone.

# Create a mapping between host name and IP address.

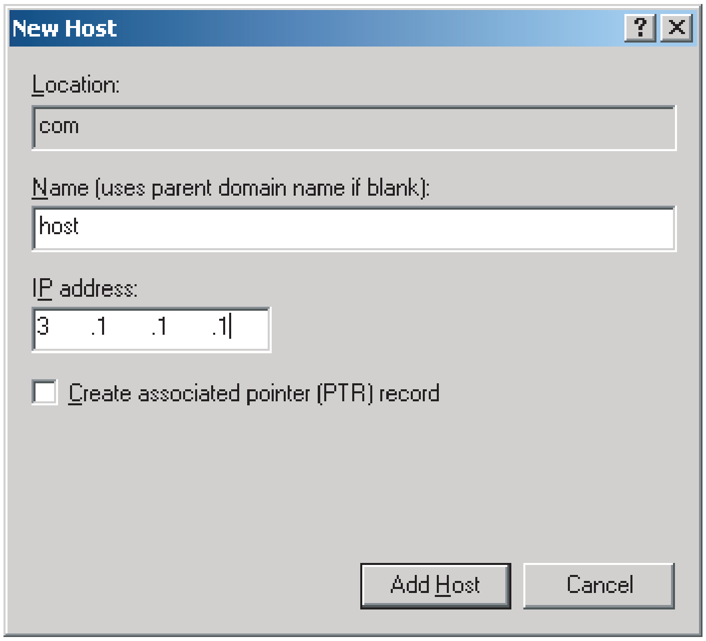
Figure 228 Add a host



Name	Type	Da
(same as parent folder)	Start of Auth...	[1]
(same as parent folder)	Name Server	xu

In Figure 228, right click zone **com**, and then select **New Host** to bring up a dialog box as shown in Figure 229. Enter host name host and IP address 3.1.1.1.

Figure 229 Add a mapping between domain name and IP address



Location:
com

Name (uses parent domain name if blank):
host

IP address:
3 .1 .1 .1

□ Create associated pointer (PTR) record

2 Configure DNS client Device

# Enable dynamic domain name resolution.

```
<Sysname> system-view
[Sysname] dns resolve
```

# Configure IP address 2.1.1.2 for the DNS server

```
[Sysname] dns server 2.1.1.2
```

# Configure com as the DNS suffix

```
[Sysname] dns domain com
```

Execute the **ping host** command on the device to verify that the communication between the device and the host is normal and that the corresponding destination IP address is 3.1.1.1.

Troubleshooting DNS Configuration

Symptom

After enabling the dynamic domain name resolution, the user cannot get the correct IP address.

Analysis

The DNS client should be used in cooperation with the DNS server to obtain a correct IP address through DNS resolution.

Solution

- Use the **display dns dynamic-host** command to check that the specified domain name is in the cache.

- If there is no defined domain name, check that dynamic domain name resolution is enabled and the DNS client can communicate with the DNS server.
- Check the mapping between the domain name and IP address is correct on the DNS server.

54 VRRP CONFIGURATION

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

- *The term router and the icon router in this document refer to a router in a generic sense or a Switch 8800 running routing protocols.*
- *At present, the interfaces that VRRP involves can only be VLAN interfaces for Switch 8800s.*

Introduction to VRRP

This section covers these topics:

Overview

Normally, you can configure a default route to the gateway for every host on a network, allowing all packets destined to the external networks to be sent over the default route to the gateway. This enables hosts on a network to communicate with external networks. However, when the gateway fails, all the hosts using the gateway as the default next-hop router are isolated from the external network.

Apparently, this approach to enabling hosts on a network to communicate with external networks is easy to configure but it imposes a very high requirement of performance stability on the device acting as the gateway. A common way to improve system reliability is to use more egress gateways, introducing the problem of routing among the multiple egresses.

Virtual router redundancy protocol (VRRP) was designed to address this problem. Deploying VRRP on multicast and broadcast LANs such as Ethernet, you can assure that the system can still provide highly reliable default links without changing configurations when a device fails.

There are two VRRP versions: VRRPv2 and VRRPv3. VRRPv2 is based on IPv4, while VRRPv3 is based on IPv6. The two versions implement the same functions but provide different commands.

Basic Concepts of VRRP

This section introduces some concepts used throughout this document:

- "VRRP standby group" on page 762
- "VRRP priority" on page 763
- "Preemption mode" on page 763
- "Interface tracking" on page 763
- "Authentication mode" on page 763

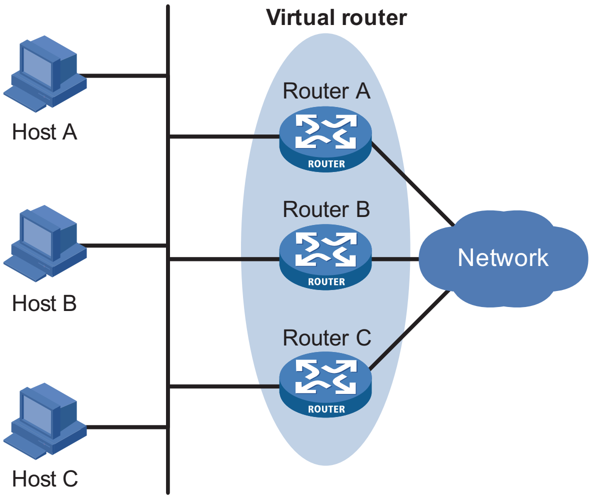
VRRP standby group

VRRP combines a group of routers on a LAN (including a master and multiple backups) into a virtual router called standby group.

The VRRP standby group has the following features:

- 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 standby group use a certain election mechanism to elect the gateway. Once the router acting as the gateway fails, the other routers in the standby group elect a new gateway to undertake the responsibility of the failed router.

Figure 230 Network diagram for a virtual router



As shown in Figure 230, 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.

In fact, only one of the three routers acts 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 standby group resides or the IP address of an interface on a router in the standby group. In the latter case, the router is called the IP address owner.*
- *In a VRRP standby group, there can only be one IP address owner.*

VRRP priority

VRRP determines the role (master or backup) of each router in the standby group by priority. A router with a higher priority has more opportunity to become the master.

Preemption mode

- In non-preemption mode, once a router in the standby group becomes the master, it stays as the master as long as it operates normally, even if a backup router is assigned a higher priority later.
- In preemption mode, once a backup router finds its priority higher than that of the router acting as the master, it becomes the master. Accordingly, the original master becomes a backup.

Interface tracking

The interface tracking function expands the backup functionality of VRRP. It provides backup not only when the interface to which a standby group is assigned fails but also when other interfaces on the router become unavailable. This is achieved by tracking interfaces. When a monitored interface goes down, the priority of the router owning the interface is automatically decreased by a specified value*,* allowing a higher priority router in the standby group to become the master.

Authentication mode

VRRP provides two authentication modes:

- Simple: Simple text authentication
- MD5: MD5 authentication

On a secure network, you can configure the routers not to perform authentication. In this case, neither the routers sending VRRP packets nor the routers receiving the VRRP packets perform authentication.

On a network where potential threats are present, you can set the authentication mode to simple. In this case, a router fills the authentication key into the VRRP packet before sending the packet out, while the router receiving the VRRP packet compares the authentication key in the packet with its own. If they are the same, the packet is considered genuine and legitimate; otherwise, the packet is considered illegitimate and is discarded.

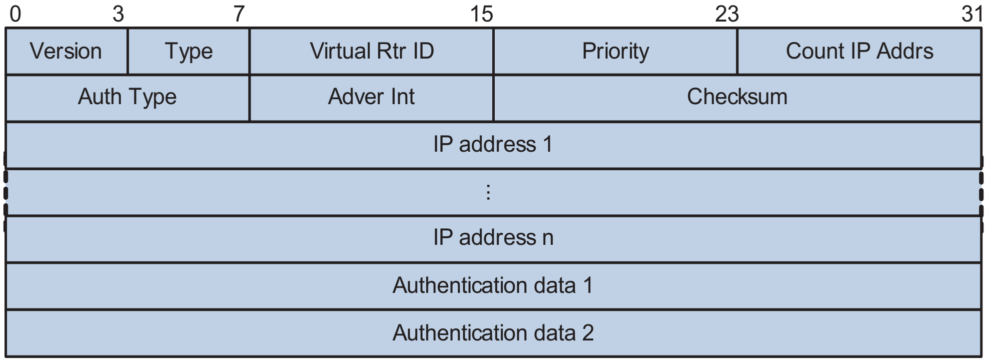
On an insecure network, you can set the authentication mode to MD5. This allows the router to encrypt VRRP packets using the authentication key and the MD5 algorithm and then save the encrypted packet in the authentication header (AH). The router receiving the VRRP packet uses the authentication key to decrypt and validate the packet.

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 231 IPv4-based VRRP packet format

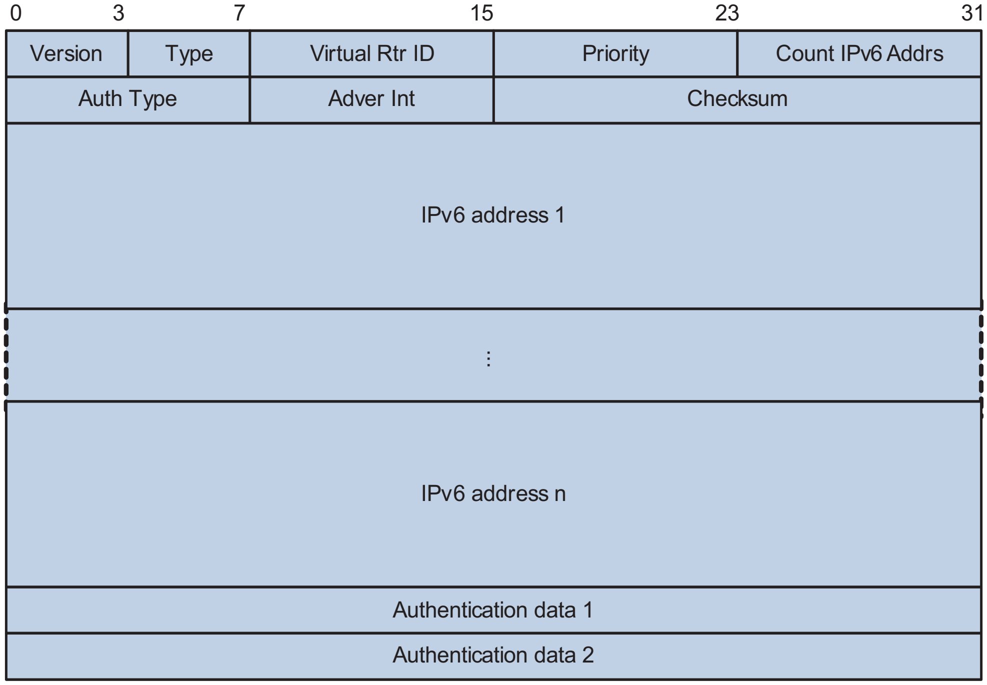


As shown in Figure 231, an IPv4-based VRRP packet consists of the following fields:

- 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): Number of the virtual router, that is, number of the standby group. It ranges from 1 to 255.
- Priority: Priority of the router in the standby group, in the range 0 to 255. A greater value represents a higher priority. The priority of 0 is reserved for special purposes, while 255 is reserved for the IP address owner.
- Count IP Addrs: Number of virtual IP addresses for the standby group. A standby group can have multiple virtual IP addresses.
- Auth Type: Authentication type. 0 means no authentication, 1 means simple 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 standby group. The allowed number is given by the Count IP Addrs field.
- 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 232 IPv6-based VRRP packet format



As shown in Figure 232, 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): Number of the virtual router, that is, number of the standby group. It ranges from 1 to 255.
- Priority: Priority of the router in the standby group, in the range 0 to 255. A greater value represents a higher priority. The priority of 0 is reserved for special purposes, while 255 is reserved for the IP address owner.
- Count IPv6 Addrs: Number of virtual IPv6 addresses for the standby group. A standby group can have multiple virtual IPv6 addresses.
- Auth Type: Authentication type. 0 means no authentication, 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 standby group. The allowed number is given by the Count IPv6 Addrs field.
- Authentication Data: Authentication key. Currently, this field is used only for simple authentication and is 0 for any other authentication modes.

Principles of VRRP

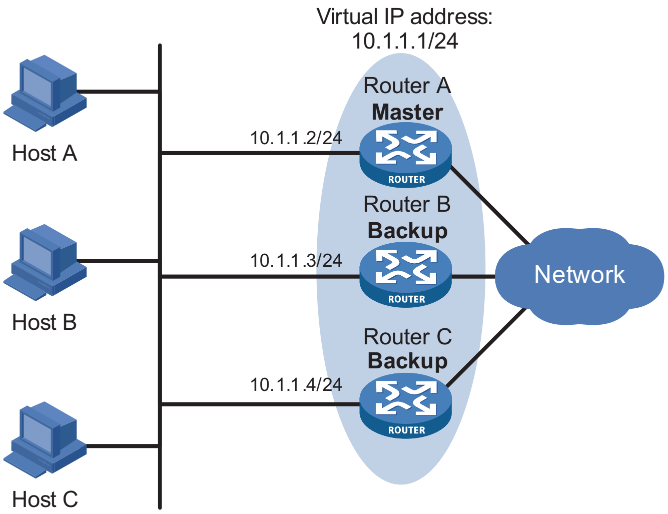
1. With VRRP enabled, the routers determine their respective roles in the standby group by priority. The router with the highest priority becomes the master, while the others are the backups. The master sends VRRP advertisement packets periodically to notify the backups that it is working properly, and each of the backups starts a timer to wait for advertisement packets from the master.
2. In preemption mode, when a backup receives a VRRP advertisement, it compares the priority in the packet with that of its own. If its priority is lower, it remains a backup; otherwise, it becomes the master.
3. In non-preemption mode, the router in the standby group remains as a master or backup as long as the master does not fail. The backup will no become the master even if the former is configured with a higher priority.
4. If the timer of a backup expires but the backup still does not receive any VRRP advertisement packet, it considers that the master fails and starts the election process to elect a new master for forwarding packets.

Operation Modes of VRRP (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 standby group, in which each router holds different priorities and the one with the highest priority becomes the master, as shown in Figure 233.

Figure 233 VRRP in master/backup mode



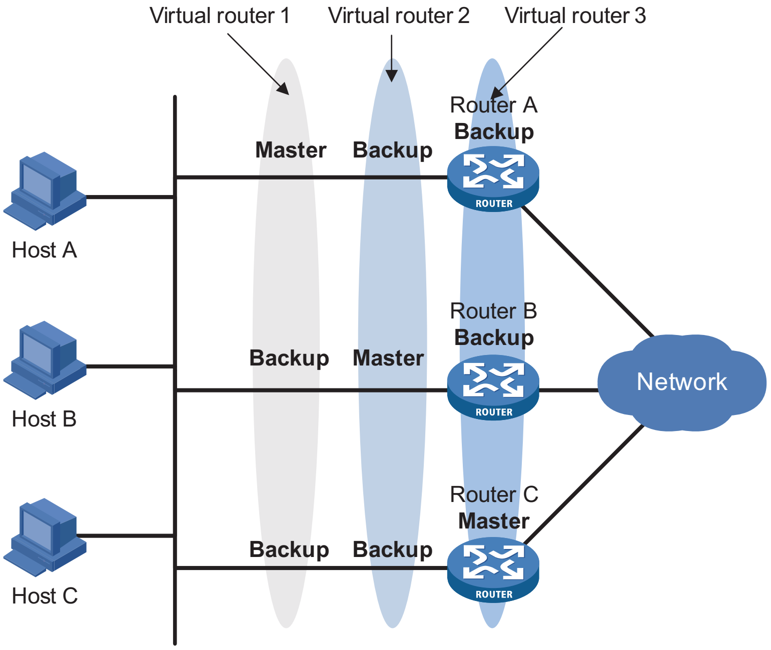
At the beginning, Router A is the master and therefore can forward packets to external networks, while Router B and Router C are backups and are thus in the state of listening. If Router A fails, Router B and Router C will elect for the new master. The new master takes over the forwarding task to provide services to hosts on the LAN.

Load balancing

You can create more than one standby group on an interface of a router, allowing the router to be the master of one standby 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 standby groups, each of which includes a master and one or more backups. The masters of the standby groups can be assumed by different routers, as shown in Figure 234.

Figure 234 VRRP in load balancing mode



A router can be in multiple standby groups and hold a different priority in different group.

In Figure 234, three standby groups are present:

- Standby group 1: Router A is the master; Router B and Router C are the backups.
- Standby group 2: Router B is the master; Router A and Router C are the backups.
- Standby 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 standby group 1, 2, and 3 as the default gateways respectively. When configuring VRRP priorities, ensure that each router holds such a priority in each standby group that it will take the expected role in the group.

Configuring VRRP for IPv4

Complete these tasks to configure VRRP for IPv4:

Task	Remarks
"Enabling Users to Ping Virtual IP Addresses of Standby Groups" on page 768	Optional
"Configuring the Association Between MAC Address and Virtual IP Address" on page 768	Optional

Task	Remarks
"Creating Standby Group and Configuring Virtual IP Address" on page 769	Required
"Configuring Priority, Preemption Mode and Interface Tracking for a Standby Group" on page 769	Optional
"Configuring VRRP Packet Attributes" on page 770	Optional

CAUTION: *VRRP is not supported on the VLAN interfaces of Super VLAN. Do not configure VRRP on this type of interfaces.*

Enabling Users to Ping Virtual IP Addresses of Standby Groups

According to VRRP, the virtual IP addresses of a standby group cannot be pinged successfully. Thus, a user connected to the switch is unable to rely on the ping command to judge whether or not an IP address has been used by the standby group. This may result in a user configuring the same IP address for the host and for the standby group. In this case, all the packets in this network segment will be sent to the host, instead of being correctly forwarded.

You can, however, follow the steps below to enable a user to successfully ping the virtual IP addresses of standby groups:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable users to ping virtual IP address of the standby group	**vrrp ping-enable**	Optional Enabled by default.

CAUTION: *Configure this function before creating a standby group. Otherwise, your configuration will fail.*

Configuring the Association Between MAC Address and Virtual IP Address

There are two types of association between MAC address and virtual IP address:

- Virtual IP address is associated with virtual switch MAC address
- Virtual IP address is associated with real MAC address of the interface

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 MAC address and virtual IP 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 standby group. Otherwise, you cannot modify the mapping between the MAC address and the virtual IP address.*

Creating Standby Group and Configuring Virtual IP Address

Configuration prerequisites

Before creating standby group and configuring virtual IP address, you should first configure the IP address of 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 standby group and configure virtual IP address:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Create standby group and configure virtual IP address of the standby group	**vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*	Required Standup group is not created by default.

CAUTION:

- *For Switch 8800s, the maximum number of VRRPv2 standby groups on an interface is 16, the maximum number of virtual IP addresses in a standby group is 16 and the maximum number of standby groups on a switch is 64.*
- *A standby group is removed after you remove all the virtual IP addresses in it. In addition, configurations on that standby group no longer take effect.*
- *The virtual IP address of the standby group cannot be 0.0.0.0, 255.255.255.255, loopback address, non A/B/C address and 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 standby group operate normally; otherwise. the state of the standby group is always* ***Initialize****.*

Configuring Priority, Preemption Mode and Interface Tracking for a Standby Group

Configuration prerequisites

Before you configure these features, you should first create a standby group on the interface and configure virtual IP address for it.

Configuration procedure

By configuring priority, preemption mode and interface tracking for a standby group, you can decide which switch in the standby group serves as the Master.

Follow these steps to configure priority, preemption mode and interface tracking for a standby group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure switch priority in the standby group	**vrrp vrid** *virtual-router-id* **priority** *priority-value*	Optional 100 by default.

To do...	Use the command...	Remarks
Configure the switch in the standby group to work in preemption mode and configure preemption delay	**vrrp vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	Optional The switch in the standby group works in preemption mode and the preemption delay is 0 seconds by default. If the switch in the standby group works in non preemption 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.

CAUTION:

- *The priority of an IP address owner is always 255 and not configurable.*
- *Interface tracking is not configurable to an IP address owner.*
- *Tracked interfaces can only be VLAN interfaces.*
- *The priority of a device is restored if the state of the interface under tracking changes from down to up.*

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create the standby group and configure the 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 VLAN interface view	**interface** *interface-type interface-number*	-
Configure the authentication mode and authentication key when the standby 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 standby 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 may configure different authentication modes and authentication keys for the standby groups on an interface. However, the members of the same standby group must use the same authentication mode and authentication key.*
- *Factors like excessive traffic or different timer setting on switches 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.

Displaying and Maintaining VRRP for IPv4

To do...	Use the command...	Remarks
Display VRRP status	**display vrrp** [**verbose**] [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Display VRRP statistics	**display vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Remove VRRP statistics	**reset vrrp statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in user view

Configuring VRRP for IPv6

Complete these tasks to configure VRRP for IPv6:

Task	Remarks
"Enabling Users to Ping Virtual IPv6 Addresses of Standby Groups" on page 771	Optional
"Configuring the Association Between MAC Address and Virtual IPv6 Address" on page 772	Optional
"Creating Standby Group and Configuring Virtual IPv6 Address" on page 772	Required
"Configuring Priority, Preemption Mode and Interface Tracking for a Standby Group" on page 773	Optional
"Configuring VRRP Packet Attributes" on page 773	Optional

CAUTION: *VRRP is not supported on the VLAN interfaces of Super VLAN. Do not configure VRRP on this type of interfaces.*

Enabling Users to Ping Virtual IPv6 Addresses of Standby Groups

According to VRRPv3, the virtual IPv6 addresses of a standby group cannot be pinged successfully. Thus, a user connected to the switch is unable to rely on the ping command to judge whether or not an IPv6 address has been used by the standby group. This may result in a user configuring the same IPv6 address for the host and for the standby group. In this case, all the packets in this network segment will be sent to the host, instead of being correctly forwarded.

You can, however, follow the steps below to enable a user to successfully ping the virtual IPv6 addresses of standby groups:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable a user to ping virtual IPv6 address of the standby group	**vrrp ipv6 ping-enable**	Optional Enabled by default

CAUTION: *You should configure this function before creating a standby group. Otherwise, you cannot ping the virtual IPv6 addresses of standby groups.*

Configuring the Association Between MAC Address and Virtual IPv6 Address

Two types of association are available between MAC address and Virtual IPv6 address:

- Virtual IPv6 address is associated with the MAC address of virtual switch
- Virtual IPv6 address is associated with the real MAC address of an interface

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 MAC address and virtual IPv6 address	**vrrp ipv6 method** { **real-mac** | **virtual-mac** }	Optional The virtual MAC address of the standby group is associated with the virtual IPv6 address by default.

CAUTION: *You should configure this function before creating a standby group. Otherwise, you cannot modify the mapping between the MAC address and the virtual IP address.*

Creating Standby Group and Configuring Virtual IPv6 Address

Configuration prerequisites

Before creating standby group and configuring virtual IPv6 address, you should first configure the 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 standby group and configure its virtual IPv6 address:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Create standby group and configure its virtual IPv6 address	**vrrp ipv6 vrid** *virtual-router-id* **virtual-ip** *virtual-address* [**link-local**]	Required No standby group is created by default. The first virtual IPv6 address of the standby group must be a link local address. Only one link local address is allowed in a standby group, and must be removed the last.

CAUTION:

- *For Switch 8800s, the maximum number of VRRPv3 standby groups on an interface is 16, the maximum number of virtual IP addresses in a standby group is 16 and the maximum number of standby groups on a switch is 64.*

- *A standby group is removed after you remove all the virtual IPv6 addresses in it. In addition, configurations on that standby group no longer take effect.*

Configuring Priority, Preemption Mode and Interface Tracking for a Standby Group

Configuration prerequisites

Before configuring these features, you should first create the standby group and configure the virtual IPv6 address.

Configuration procedure

By configuring standby group priority, preemption mode and interface tracking, you can decide which switch in the standby group serves as the Master.

Follow these steps to configure priority, preemption mode and interface tracking for a standby group:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure the priority of the switch in the standby group	**vrrp ipv6 vrid** *virtual-router-id* **priority** *priority-value*	Optional 100 by default
Configure the switch in the standby to work in preemption mode and configure preemption delay of the standby group	**vrrp ipv6 vrid** *virtual-router-id* **preempt-mode** [**timer delay** *delay-value*]	Optional The switch in the standby group works in preemption mode and the preemption delay is zero seconds by default. If the switch in the standby group works in non preemption mode, the preemption delay changes to zero seconds automatically.
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 priority of an IP address owner is always 255 and not configurable.*
- *Interface tracking is not configurable on an IP address owner.*
- *Tracked interfaces can only be VLAN interfaces.*
- *The priority of a device is reset if the state of the interface under tracking changes from down to up.*

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create the standby group and configure the 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 VLAN interface view	**interface** *interface-type interface-number*	-
Configure the authentication mode and authentication key when the standby groups send and transmit 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 standby 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 standby groups on an interface. However, the members of the same standby group must use the same authentication mode and authentication key.*
- *Factors like excessive traffic or different timer setting on switches 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 delay for preemption.*

Displaying and Maintaining VRRP for IPv6

To do...	Use the command...	Remarks
Display VRRP status	**display vrrp ipv6 [verbose]** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Display VRRP statistics	**display vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in any view
Remove VRRP statistics	**reset vrrp ipv6 statistics** [**interface** *interface-type interface-number* [**vrid** *virtual-router-id*]]	Available in user view

IPv4-Based VRRP Configuration Example

This section provides these configuration examples:

- "Single VRRP Standby Group Configuration Example" on page 774
- "VRRP Interface Tracking Configuration Example" on page 777
- "Multiple VRRP Standby Groups Configuration Example" on page 780

Single VRRP Standby Group Configuration Example

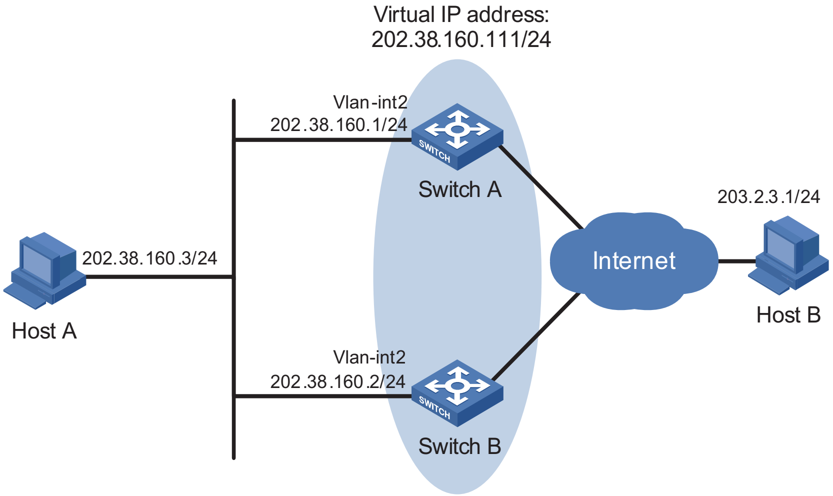
Network requirements

- 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 standby group 1 with the virtual IP address of 202.38.160.111.

- 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.

Network diagram

Figure 235 Network diagram for single VRRP standby group configuration



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 2/1/4
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Configure the priority of Switch A in the standby group 1 as 110.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 priority 110
```

# Configure Switch A to work in preemption mode and configure the preemption delay to five seconds.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet 2/1/4
```

```
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Configure Switch A to work in preemption mode and configure the preemption delay to five seconds.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
[SysnameB-Vlan-interface2] return
```

3 Verify the configuration

After the configuration, host B can be pinged through on host A. You can use the **display vrrp** command to verify the configuration.

# Display detailed information of standby group 1 on Switch A.

```
<SysnameA> display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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 standby group 1 on Switch B.

```
<SysnameB> display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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      : 202.38.160.111
 Master IP       : 202.38.160.1
```

The above information indicates that in standby 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** command to view the detailed information of the standby group on Switch B.

If Switch A fails, the detailed information of standby group 1 on Switch B is displayed.

```
<SysnameB> display vrrp verbose
 IPv4 Standby Information:
 Run Method       : VIRTUAL-MAC
 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       : 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.

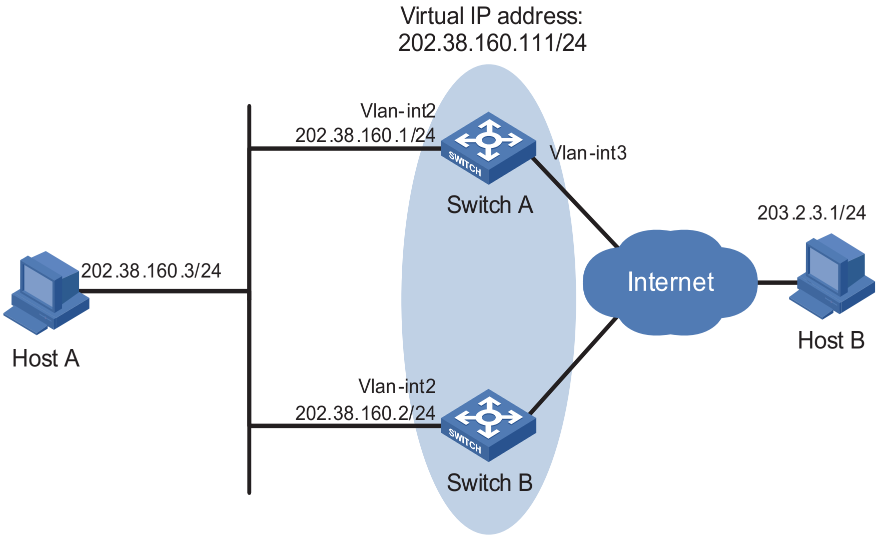
VRRP Interface Tracking Configuration Example

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 standby group 1 with the virtual IP address of 202.38.160.111.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A is in work, but when its interface Vlan-interface3 which connects to the internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 236 Network diagram for interface tracking in VRRP



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 2/1/4
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Set the priority of Switch A in the standby group to 110.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 priority 110
```

# Configure the authentication mode of the standby group as MD5 and authentication key as abc123.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 authentication-mode md5 abc123
```

# Configure the master to send VRRP packets every five seconds.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

# Set the interface to be tracked.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 track vlan-interface 3 reduced 30
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet 2/1/4
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Configure the authentication mode of the standby group as MD5 and authentication key as abc123.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 authentication-mode md5 abc123
```

# Configure the master to send VRRP packets every five seconds.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 timer advertise 5
[SysnameB-Vlan-interface2] return
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp** command to verify the configuration.

# Display detailed information of standby group 1 on Switch A.

```
<SysnameA> display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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 IF        : Vlan-interface3            Pri Reduced     : 30
 Virtual IP      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1
```

# Display detailed information of standby group 1 on Switch B.

```
<SysnameB> display vrrp verbose
 IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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      : 202.38.160.111
 Master IP       : 202.38.160.1
```

The above information indicates that in standby 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 is in work, but when its Vlan-interface3 that connects to the Internet is not available, you can still ping through Host B on Host A. Use the **display vrrp** command to view the detailed information of the standby group.

# If Vlan-interface3 on Switch A is not available, the detailed information of standby group 1 on Switch A is displayed.

```
<SysnameA> display vrrp verbose
IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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 IF         : Vlan-interface3            Pri Reduced     : 30
 Virtual IP       : 202.38.160.111
 Master IP        : 202.38.160.2
```

# If Vlan-interface3 on Switch A is not available, the detailed information of standby group 1 on Switch B is displayed.

```
<SysnameB> display vrrp verbose
 IPv4 Standby Information:
 Run Method       : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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       : 202.38.160.111
 Virtual MAC      : 0000-5e00-0101
 Master IP        : 202.38.160.2
```

The above information indicates that if Vlan-interface3 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.

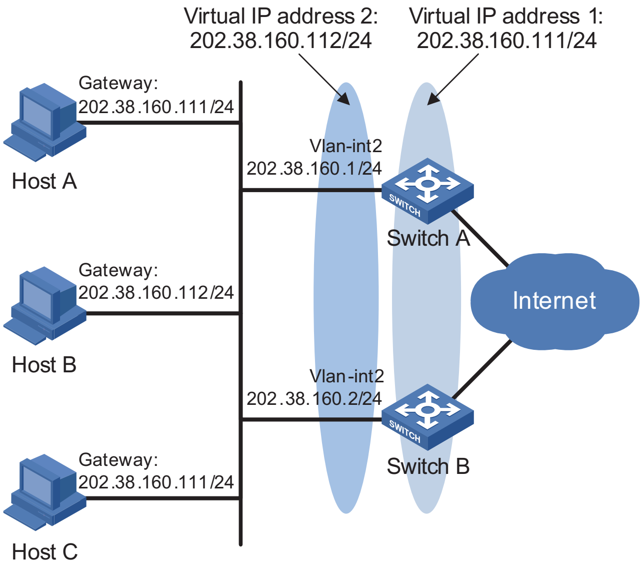
Multiple VRRP Standby Groups Configuration Example

Network requirements

- In the segment 202.28.160.0/24, some hosts use 202.38.160.111/24 as their default gateway and some hosts use 202.38.160.112/24 as their default gateway.
- Load sharing and mutual backup between default gateways can be implemented by using VRRP standby groups.

Network diagram

Figure 237 Network diagram for multiple VRRP standby groups configuration



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 2/1/3
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Set the priority of Switch A in standby group 1 to 110.

```
[SysnameA-Vlan-interface2] vrrp vrid 1 priority 110
```

# Create standby group 2 and configure its virtual IP address as 202.38.160.112.

```
[SysnameA-Vlan-interface2] vrrp vrid 2 virtual-ip 202.38.160.112
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet 2/1/3
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

# Create standby group 1 and configure its virtual IP address as 202.38.160.111.

```
[SysnameB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

# Create standby group 2 and configure its virtual IP address as 202.38.160.112.

```
[SysnameB-Vlan-interface2] vrrp vrid 2 virtual-ip 202.38.160.112
```

# Set the priority of Switch B in standby group 2 to 110.

```
[SysnameB-Vlan-interface2] vrrp vrid 2 priority 110
[SysnameB-Vlan-interface2] return
```

3 Verify the configuration

You can use the **display vrrp** command to verify the configuration.

# Display detailed information of the standby group on Switch A.

```
<SysnameA> display vrrp verbose
IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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      : 202.38.160.111
 Virtual MAC     : 0000-5e00-0101
 Master IP       : 202.38.160.1

 Interface       : Vlan-interface2
 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      : 202.38.160.112
 Master IP       : 202.38.160.2
```

# Display detailed information of the standby group on Switch B.

```
<SysnameB> display vrrp verbose
IPv4 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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      : 202.38.160.111
 Master IP       : 202.38.160.1

 Interface       : Vlan-interface2
```

```
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          : 202.38.160.112
Virtual MAC         : 0000-5e00-0102
Master IP           : 202.38.160.2
```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and the host with the default gateway of 202.38.160.111/24 accesses the Internet through Switch A; in standby group 2 Switch A is the backup, Switch B is the master and the host with the default gateway of 202.38.160.112/24 accesses the Internet through Switch B.

IPv6-Based VRRP Configuration Example

This section provides these configuration examples:

- "Single VRRP Standby Group Configuration Example" on page 783
- "VRRP Interface Tracking Configuration Example" on page 786
- "Multiple VRRP Standby Group Configuration Example" on page 789

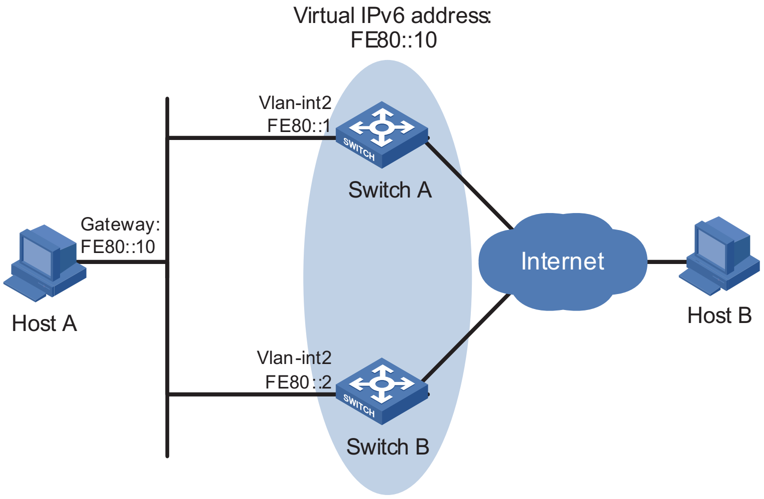
Single VRRP Standby Group Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using FE80::10 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IPv6 address of 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.

Network diagram

Figure 238 Network diagram for single IPv6 VRRP standby group configuration



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] ipv6
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 2/1/6
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ipv6 address fe80::1 link-local
[SysnameA-Vlan-interface2] ipv6 address 1::1 64
```

# Create a standby group 1 and set its virtual IP address to fe80::10.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Set the priority of Switch A in standby group 1 to 110.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

# Set Switch A to work in preemption mode.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode
```

# Enable Switch A to send RA messages.

```
[SysnameA-Vlan-interface2] undo ipv6 nd ra halt
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] ipv6
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet 2/1/6
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ipv6 address fe80::2 link-local
[SysnameB-Vlan-interface2] ipv6 address 1::2 64
```

# Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Enable Switch B to send RA messages.

```
[SysnameB-Vlan-interface2] undo ipv6 nd ra halt
[SysnameB-Vlan-interface2] return
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp ipv6** command to verify the configuration.

# Display detailed information of standby group 1 on Switch A.

```
<SysnameA> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1
```

# Display detailed information of standby group 1 on Switch B.

```
<SysnameB> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
 Master IP       : FE80::1
```

The above information indicates that in standby 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** command to view the detailed information of the standby group on Switch B.

# If Switch A fails, the detailed information of standby group 1 on Switch B is displayed.

```
<SysnameB> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 Interface       : Vlan-interface2
 VRID            : 1                      Adver. Timer    : 100
 Admin Status    : UP                     State           : Master
 Config Pri      : 100                    Run Pri         : 100
 Preempt Mode    : YES                    Delay Time      : 0
 Auth Type       : NONE
 Virtual IP      : FE80::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.

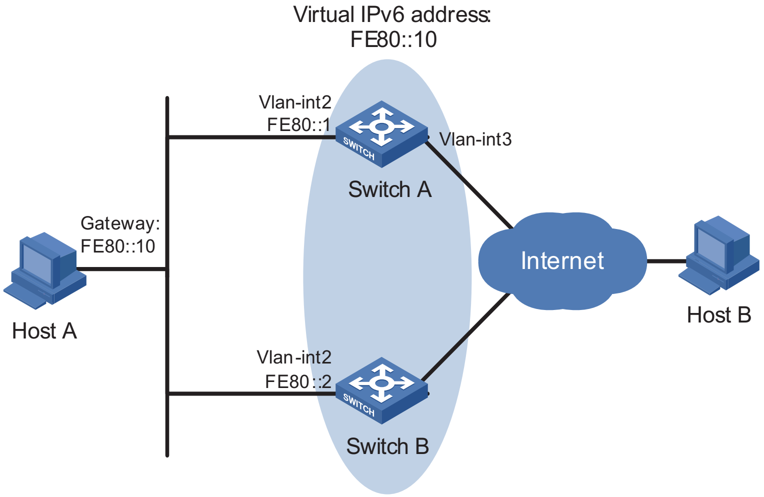
VRRP Interface Tracking Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using FE80::10 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IP address of FE80::10.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A is in work, but its Vlan-interface3 which connects to the Internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 239 Network diagram for IPv6 VRRP interface tracking



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] ipv6
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 1/5
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ipv6 address fe80::1 link-local
[SysnameA-Vlan-interface2] ipv6 address 1::1 64
```

# Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Set the priority of Switch A in standby group 1 to 110.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

# Set the authentication mode for standby group 1 to SIMPLE and authentication key to hello.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simple hello
# Set the VRRP advertisement interval to 500 centiseconds.
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

# Set Switch A work in preemption mode. The preemption delay is five seconds.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

# Set the interface to be tracked.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 track interface vlan-int
erface 3 reduced 30
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] ipv6
[SysnameB] vlan 2
[SysnameB-vlan2] port ethernet 1/5
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ipv6 address fe80::2 link-local
[SysnameB-Vlan-interface2] ipv6 address 1::2 64
```

# Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Set the authentication mode for standby group 1 to SIMPLE and authentication key to hello.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simple hello
# Set the VRRP advertisement interval to 500 centiseconds.
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

# Set Switch B to work in preemption mode. The preemption delay is five seconds.

```
[SysnameB-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** command to verify the configuration.

# Display detailed information of standby group 1 on Switch A.

```
<SysnameA> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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        : Vlan-interface3      Pri Reduced     : 30
 Virtual IP      : FE80::10
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1
```

# Display detailed information of standby group 1 on Switch B.

```
<SysnameB> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
 Master IP       : FE80::1
```

The above information indicates that in standby 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 is in work, but its interface Vlan-interface3 is not available, you can still ping through Host B on Host A. You can use the **display vrrp ipv6** command to view the detailed information of the standby group.

# If Switch A is in work, but its interface Vlan-interface3 is not available, the detailed information of standby group 1 on Switch A is displayed.

```
<SysnameA> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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        : Vlan-interface3      Pri Reduced     : 30
 Virtual IP      : FE80::10
 Master IP       : FE80::2
```

# If Switch A is in work, but its interface Vlan-interface3 is not available, the detailed information of standby group 1 on Switch B is displayed.

```
<SysnameB> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
Virtual MAC       : 0000-5e00-0201
Master IP         : FE80::2
```

The above information indicates that if Vlan-interface3 on Switch A is not available, the priority of Switch A reduces 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.

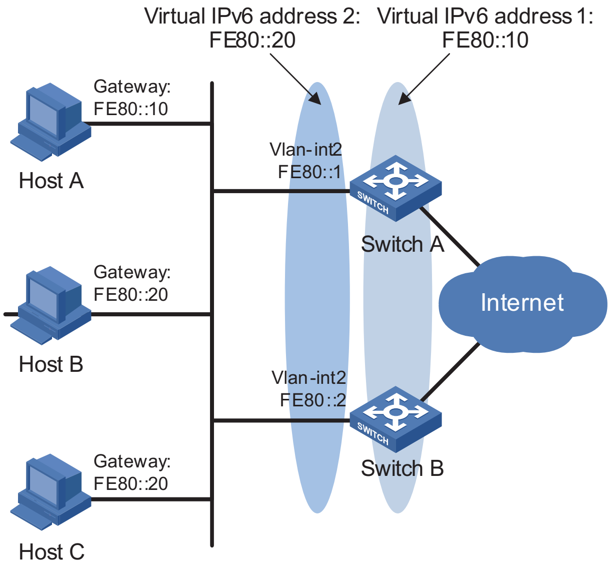
Multiple VRRP Standby Group Configuration Example

Network requirements

- In the network, some hosts use FE80::10 as their default gateway and some hosts use FE80::20 as their default gateway.
- Load sharing and mutual backup between default gateways can be implemented by using VRRP standby groups.

Network diagram

Figure 240 Network diagram for multiple IPv6 VRRP standby group configuration



Configuration procedure

1 Configure Switch A

# Configure VLAN 2.

```
<SysnameA> system-view
[SysnameA] ipv6
[SysnameA] vlan 2
[SysnameA-vlan2] port ethernet 2/1/6
[SysnameA-vlan2] quit
[SysnameA] interface vlan-interface 2
[SysnameA-Vlan-interface2] ipv6 address fe80::1 link-local
[SysnameA-Vlan-interface2] ipv6 address 1::1 64
```

# Create standby group 1 and set its virtual IP address to FE80::10.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Set the priority of Switch A in standby group 1 to 110.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

# Create standby group 2 and set its virtual IP address to FE80::20.

```
[SysnameA-Vlan-interface2] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-local
[SysnameA-Vlan-interface2] return
```

2 Configure Switch B

# Configure VLAN 2.

```
<SysnameB> system-view
[SysnameB] ipv6
[SysnameB-vlan2] port ethernet 2/1/6
[SysnameB-vlan2] quit
[SysnameB] interface vlan-interface 2
[SysnameB-Vlan-interface2] ipv6 address fe80::2 link-local
[SysnameB-Vlan-interface2] ipv6 address 1::2 64
```

# Create standby group 1 and set its virtual IP address to FE80::10.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

# Create standby group 2 and set its virtual IP address to FE80::20.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-local
```

# Set the priority of Switch B in standby group 2 to 110.

```
[SysnameB-Vlan-interface2] vrrp ipv6 vrid 2 priority 110
[SysnameB-Vlan-interface2] return
```

3 Verify the configuration

You can use the **display vrrp ipv6** command to verify the configuration.

# Display detailed information of the standby group on Switch A.

```
<SysnameA> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
 Virtual MAC     : 0000-5e00-0201
 Master IP       : FE80::1

 Interface       : Vlan-interface2
 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      : FE80::20
 Master IP       : FE80::2
```

# Display detailed information of the standby group on Switch B.

```
<SysnameB> display vrrp ipv6 verbose
 IPv6 Standby Information:
 Run Method      : VIRTUAL-MAC
 Virtual IP Ping : Enable
 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
 Master IP       : FE80::1

 Interface       : Vlan-interface2
 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      : FE80::20
 Virtual MAC     : 0000-5e00-0202
 Master IP       : FE80::2
```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and the host with the default gateway of FE80::10 accesses the Internet through Switch A; in standby group 2 Switch A is the backup, Switch B is the master and the host with the default gateway of FE80::20 accesses the Internet through Switch B.

Multiple standby groups are commonly used in actual networking. In IPv6 network, you need to manually configure the default gateway for VRRP standby group to share load.

Troubleshooting VRRP

Symptom 1:

The console screen displays error prompts frequently.

Analysis:

This error is probably due to the inconsistent configuration of the other device in the standby group, or that 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 standby group.

Analysis:

- If presence of multiple masters only lasts a short period, this is normal and requires no manual intervention.
- If it lasts long, you must ensure that these masters can receive VRRP packets and the packets received are legitimate.

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.

55 GR CONFIGURATION

When configuring Graceful Restart (GR), go to these sections for information you are interested in:

- "Introduction to Graceful Restart" on page 793
- "Configuring Graceful Restart" on page 797
- "Displaying and Maintaining Graceful Restart" on page 799
- "Graceful Restart Configuration Examples" on page 799

Throughout this chapter, the term router in this document refers to a router in a generic sense or a Switch 8800 running routing protocols.

Introduction to Graceful Restart

Graceful Restart Overview

Graceful Restart ensures the continuity of packet forwarding when a routing protocol restarts.

The mechanism of Graceful Restart works as follows: when the routing protocol on a Graceful Restart capable device restarts, the device can notify its neighbors to temporarily preserve its adjacency with them and the routing information. After the routing protocol restart is finished, the neighbors help the device to synchronize its routing information and to restore it to the state prior to the restart in minimal time. The routing and forwarding remain highly stable during the restart, the packet forwarding paths remain the same, and the whole system can forward IP packets continuously. Hence, it is called "Graceful Restart".

Basic Mechanism of Graceful Restart

Basic concepts in Graceful Restart

A router with the Graceful Restart feature enabled is called a Graceful Restart capable router. It can perform a Graceful Restart when its routing protocol restarts. 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, which 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 across 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 and routing information sent from the GR Restarter for a period as specified by the GR Time.

Graceful Restart communication procedure

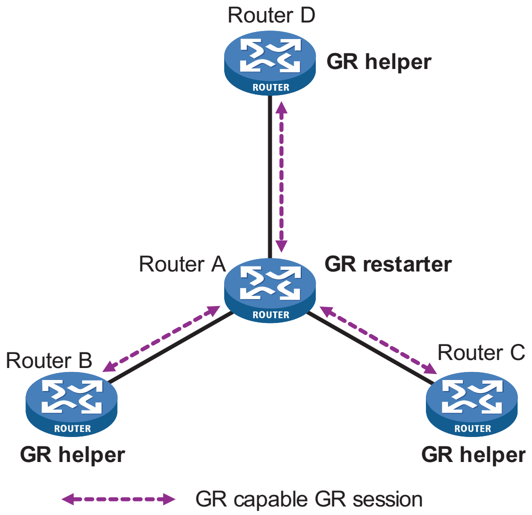
Configure a device as GR Restarter in a network. This device and its GR Helper must support GR or be GR capable. Thus, when GR Restarter restarts, its GR Helper can know its restart process.

- *In some cases, GR Restarter and GR Helper can replace with each other.*
- *If a router is to act as a Graceful Restarter, it must have the ability to preserve the routing information in the routing table (forwarding table). Routers that fail to meet this can only act as a GR Helper.*

The communication procedure between the GR Restarter and the GR Helper works as follows:

1 A GR session is established between the GR Restarter and the GR Helper.

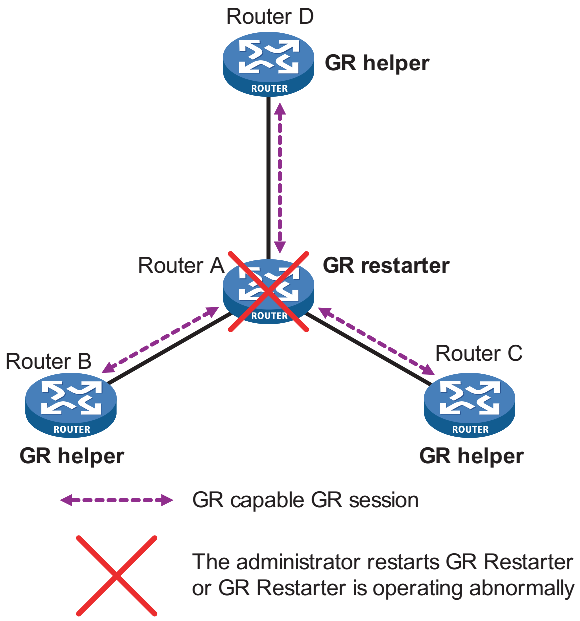
Figure 241 A GR session is established between the GR Restarter and the GR Helper



As illustrated in Figure 241, 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 GR Restarter restarting

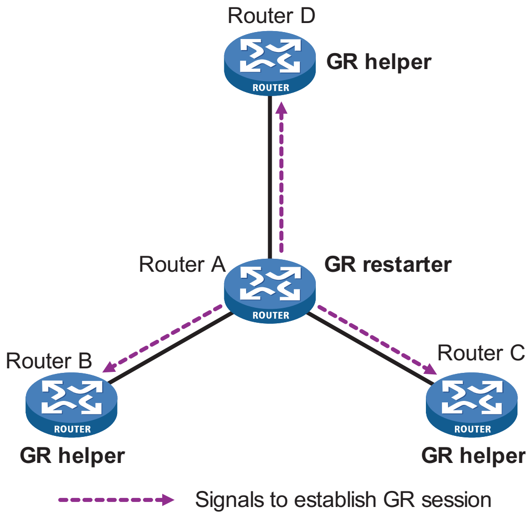
Figure 242 Restarting process for the GR Restarter



As illustrated in Figure 242. 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 GR Restarter signaling to GR Helper

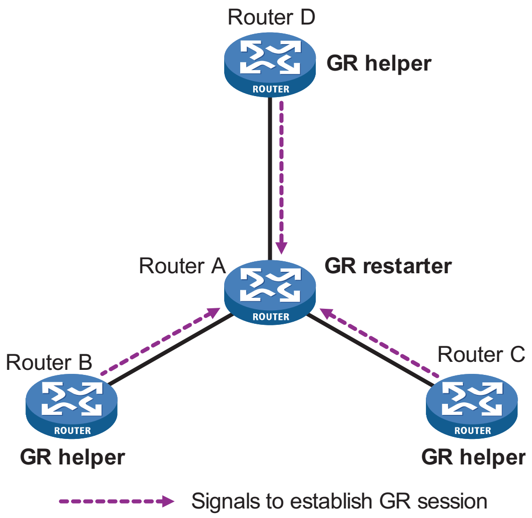
Figure 243 The GR Restarter signals to the GR Helper(s) after restart



As illustrated in Figure 243, after the GR Restarter has recovered, it will signal to all its neighbors and will reestablish GR Session.

4 The GR Restarter obtaining topology and routing information from the GR Helper

Figure 244 The GR Restarter obtains topology and routing information from the GR Helper



As illustrated in Figure 244, 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

Graceful restart is currently implemented in BGP, OSPF, IS-IS, and LDP.

BGP-based Graceful Restart

1. To establish a session with its peer, a BGP-based GR Restarter needs to send an OPEN message first to its peer including its Graceful Restart Capability.
2. Upon receipt of this message, the peer is aware that the sending router is capable of Graceful Restart. It will use the OPEN messages to exchange the Graceful Restart Capability with the GR Restarter and to establish a GR session with it. If neither party has the Graceful Restart Capability, the session established between them will not be Graceful Restart Capable.
3. The GR session between the GR Restarter and the GR Helper goes down when a BGP restarts. A Graceful Restart Capable GR Helper will mark all routes associated with the GR Restarter as stale. However, during the configured GR Time, it still uses these routes in packet forwarding, ensuring that no packet will be lost when routing information from its peer is recollected.
4. After the restart, the GR Restarter will reestablish the GR session with its peer and send new GR messages notifying the completion of restart. Routing information can be exchanged between the peers and used by the GR Restarter in creating the new routing table (forwarding table) in place of the stale routing information. Thus the BGP routing convergence is complete.

OSPF-based Graceful Restart

After an OSPF-based GR Restarter restarts its OSPF, it needs to perform the following two tasks in order to update its link-state database with its neighbor.

1. To obtain once again effective OSPF adjacency while retaining the original one.
2. To obtain once again link-state database information.

Before the restart, the GR Restarter originates Grace-LSA negotiation GR capability. During the restart, the GR Helper continues to advertise its adjacency with the GR Restarter.

After the restart, the GR Restarter will send an OSPF GR signal to its neighbor so that the adjacency is not reset. In this way, the GR Restarter can restore its adjacency with its neighbor upon receiving the responses from the latter.

Upon the reestablishment of the adjacency, the GR Restarter will update its link-state database with all of its GR Capable neighbors and exchange routing information with them. After that, the GR Restarter will update its own routing table (forwarding table) based on the new routing information and delete the stale routes. In this way, the OSPF routing convergence is complete.

Configuring Graceful Restart

One device can act as both GR Restarter and GR Helper at the same time.

Graceful Restart configuration can be normally achieved by enabling the Graceful Restart Capability. The time parameters can be configured if necessary, but under normal circumstances the defaults should be used.

Configuring BGP-based Graceful Restart

Follow these steps to configure BGP-based GR on the GR Restarter and the GR Helper:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable BGP, and enter its view	**bgp** *as-number*	Required Disabled by default
Enable Graceful Restart 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: *End-of-RIB = End of Router-Information-Base*

- *In general the maximum time allowed for the peer to reestablish a BGP session should be less than the Holdtime carried in the OPEN message. Upon detection of the termination of the session, the GR Helper will wait for a period of time as specified by the GR Time before it reestablishes the BGP session.*
- *The End-of-RIB marker can be used to indicate that the updated routing information has been sent.*

Configuring OSPF-based Graceful Restart

Configuring OSPF-based GR Restarter

Follow these steps to configure the OSPF-based 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** *vpn-instance-name*] *	Required Disabled by default
Enable opaque LSA capability	**opaque-capability enable**	Required Disabled by default.
Enable the use of link-local signaling	**enable link-local-signaling**	Required Disabled by default
Enable out-of-band resynchronization	**enable out-of-band-resynchronization**	Required Disabled by default
Enable GR for OSPF	**graceful-restart**	Required Disabled by default

Configuring OSPF-based GR Helper

Follow these steps to configure OSPF-based GR Helper:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable OSPF, and enter OSPF view	**ospf** [*process-id* | **router-id** *router-id* | **vpn-instance** *instance-name*] *	Required Disabled by default
Enable the use of link-local signaling	**enable link-local-signaling**	Required Disabled by default
Enable out-of-band resynchronization	**enable out-of-band-resynchronization**	Required Disabled by default
Configure for which OSPF neighbors the current 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.

Restarting OSPF Graceful Restart

You can perform the following configuration on a routing switch to restart the GR process for OSPF protocol, but ensure that the routing switch has been enabled with the following capabilities first:

- LLS (link local signaling)
- OOB (out of band resynchronization)
- Opaque LSA

Following the step to restart the OSPF GR process:

To do...	Use the command...	Remarks
Restart OSPF Graceful Restart	**reset ospf** [*process-id*] **process graceful-restart**	Required Available in user view

Displaying and Maintaining Graceful Restart

To do...	Use the command...	Remarks
Display the Graceful Restart state for IS-IS	**display isis graceful-restart status** [**level-1** | **level-2**] [*process-id* | **vpn-instance** *vpn-instance-name*]	Available in any view

Graceful Restart Configuration Examples

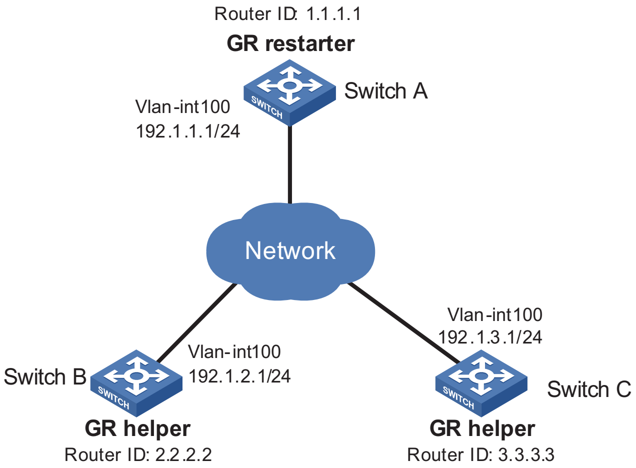
OSPF-based Graceful Restart Configuration Examples

Network requirements

- Switch A, Switch B and Switch C belong to the same autonomous system and the same OSPF domain. They are connected through OSPF.
- Switch A acts as the GR Restarter, and Switch B and Switch C are the GR Helpers and remain OOB synchronized with Switch A through GR mechanism.

Network diagram

Figure 245 OSPF-based Graceful Restart configuration example



Configuration procedure

1 Configure Switch A.

```
<SysnameA> system-view
[SysnameA] interface vlan-interface 100
[SysnameA-Vlan-interface100] ip address 192.1.1.1 255.255.255.0
[SysnameA-Vlan-interface100] quit
[SysnameA] router id 1.1.1.1
[SysnameA] ospf 100
[SysnameA-ospf-100] enable link-local-signaling
[SysnameA-ospf-100] enable out-of-band-resynchronization
[SysnameA-ospf-100] graceful-restart
[SysnameA-ospf-100] area 0
[SysnameA-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SysnameA-ospf-100-area-0.0.0.0] return
```

2 Configure Switch B.

```
<SysnameB> system-view
[SysnameB] acl number 2000
[SysnameB-acl-basic-2000] rule 10 permit source 192.1.1.1 0.0.0.0
[SysnameB-acl-basic-2000] quit
[SysnameB] interface vlan-interface 100
[SysnameB-Vlan-interface100] ip address 192.1.2.1 255.255.255.0
[SysnameB-Vlan-interface100] ospf dr-priority 0
[SysnameB-Vlan-interface100] quit
[SysnameB] router id 2.2.2.2
[SysnameB] ospf 100
[SysnameB-ospf-100] enable link-local-signaling
[SysnameB-ospf-100] enable out-of-band-resynchronization
[SysnameB-ospf-100] graceful-restart
[SysnameB-ospf-100] area 0
[SysnameB-ospf-100-area-0.0.0.0] network 192.1.2.0 0.0.0.255
```

3 Configure Switch C.

```
<SysnameC> system-view
[SysnameC] acl number 2000
[SysnameC-acl-basic-2000] rule 10 permit source 192.1.1.1 0.0.0.0
[SysnameC-acl-basic-2000] quit
[SysnameC] interface vlan-interface 100
[SysnameC-Vlan-interface100] ip address 192.1.3.1 255.255.255.0
[SysnameC-Vlan-interface100] ospf dr-priority 2
[SysnameC-Vlan-interface100] quit
[SysnameC] router id 3.3.3.3
[SysnameC] ospf process 100
[SysnameC-ospf-100] enable link-local-signaling
[SysnameC-ospf-100] enable out-of-band-resynchronization
[SysnameC-ospf-100] graceful-restart
[SysnameC-ospf-100] area 0
[SysnameC-ospf-100-area-0.0.0.0] network 192.1.3.0 0.0.0.255
```

56 ACL OVERVIEW

Unless otherwise stated, ACLs refer to both IPv4 ACLs and IPv6 ACLs throughout this document.

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 address, destination address, and port number.

They can apply to firewall, QoS, and wherever traffic identification is desired.

This chapter covers these topics:

- "Time-Based ACL" on page 801
- "IPv4 ACL" on page 801
- "IPv6 ACL" on page 803

Time-Based ACL

Time-based ACLs allow you to control the period during which a rule can take effect by referencing a time range in the rule.

The 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 comes active.

CAUTION: *On the Switch 8800s, the active state of ACL rules must be consistent on the I/O Module and interface cards.*

IPv4 ACL

This section covers these topics:

- "IPv4 ACL Classification" on page 801
- "IPv4 ACL Naming" on page 802
- "IPv4 ACL Match Order" on page 802
- "IP Fragments Filtering with IPv4 ACL" on page 803

IPv4 ACL Classification

IPv4 ACLs, identified by ACL numbers, fall into the following four categories:

- Basic IPv4 ACL, based on source IP address. Basic ACLs are numbered 2000 through 2999.
- Advanced IPv4 ACL, based on source IP address, destination IP address, protocol carried on IP, and other Layer 3 or Layer 4 protocol header information. Advanced ACLs are numbered 3000 through 3999.

- Ethernet frame header ACL, based on Layer 2 protocol header fields such as source MAC address, destination MAC address, 802.1p priority, and link layer protocol type. Ethernet frame header ACLs are numbered 4000 through 4999.
- User-defined ACL, based on customized information of protocol headers such as IP and MPLS. User-defined ACLs are numbered 5000 through 5999.

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 the name of the ACL.

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

Each ACL is a sequential collection of rules defined with different matching criteria. The order in which a packet is matched against the rules may affect how the packet is handled.

At present, the following two match orders are available:

- **config**: where packets are compared against ACL rules in the order in which they are configured.
- **auto**: where depth-first match is performed. 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 source IP address wildcard first and compare packets against the rule configured with more zeros in the source IP address wildcard prior to other rules.
2. If two rules are present with the same number of zeros in their source IP address wildcards, compare packets against the rule configured first prior to the other.

For example, the rule with the source IP address wildcard 0.0.0.255 is compared prior to the rule with the source IP address wildcard 0.0.255.255.

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 source IP address wildcard first and compare packets against the rule configured with more zeros in the source IP address wildcard prior to other rules.
2. If two rules are present with the same number of zeros in their source IP address wildcards, look at the destination IP address wildcards in the rules in addition. Then, compare packets against the rule configured with more zeros in the destination IP address wildcard prior to the other.
3. If the numbers of zeros in the destination IP address wildcards are the same, compare packets against the rule configured first prior to the other.

For example, the rule with the source IP address wildcard 0.0.0.255 is compared prior to the rule with the source IP address wildcard 0.0.255.255.

Depth-first match for an Ethernet frame header IPv4 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 prior to other rules.
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 prior to the other.
3. If the numbers of ones in the destination MAC address masks are the same, the one configured first is compared prior to the other.

For example, the rule with source MAC address mask FFFF-FFFF-0000 is compared prior to the rule with source MAC address mask FFFF-0000-0000.

The match order for a user-defined ACL can only be ***config****.*

The comparison of a packet against an ACL stops once a match is found. The packet is then processed as per the rule.

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.

To address the risk, the following packet filtering functions are delivered:

- IP-based filtering on all fragments.
- Standard match and exact match for ACLs containing advanced information such as TCP/UDP port number and ICMP type. The default approach is standard match.

- *Standard match considers only Layer 3 information.*
- *Exact match considers all header information defined in ACL rules.*

IPv6 ACL

This section covers these topics:

- "IPv6 ACL Classification" on page 803
- "IPv6 ACL Naming" on page 804
- "IPv6 ACL Match Order" on page 804

IPv6 ACL Classification

IPv6 ACLs, identified by ACL numbers, fall into the following three categories:

- Basic IPv6 ACL, based on source IPv6 address. Basic IPv6 ACLs are numbered 2000 through 2999.

- Advanced IPv6 ACL, based on source IPv6 address, destination IPv6 address, protocol carried on IP, and other Layer 3 or Layer 4 protocol header fields. Advanced ACLs are numbered 3000 through 3999.

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 the name of the ACL.

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, IPv6 ACLs are sequential collections of rules defined with different matching parameters. The order in which a packet is matched against the rules in an IPv6 ACL may affect how the packet is handled.

Like in IPv4 ACLs, the following two match orders are available in IPv6 ACLs:

- **config**: where rules are compared against in the order in which they are configured.
- **auto**: where depth-first match is performed.

The depth-first mechanism performed by IPv6 ACLs is to match packets against the rule that specifies a narrower address range first. This is done by comparing prefix lengths: the smaller the prefix length, the narrower the address range.

Consider two IPv6 addresses, 2050:6070::/96 and 2050:6070::/64. In the **auto** match approach, packets are matched against the rule with the address of 2050:6070::/96 first, because that address specifies a narrower address range compared with 2050:6070::/64. In case two rules with the same prefix length are defined in an IPv6 ACL, the one configured first is compared prior to the other one.

The comparison of a packet against an ACL stops once a match is found. The packet is then processed as per the rule.

57 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-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 state of a specified or all time ranges	**display time-range** { **all** | *time-name* }	Optional Available in any view

A time range can be one of the following:

- Periodic time range created using the **time-range** *time-name start-time* **to** *end-time days* command. A time range thus created recurs periodically on the day or days of the week.
- Absolute time range created using the **time-range** *time-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-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.

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
[Sysname] display time-range test
Current time is 13:27:32 4/16/2005 Saturday
Time-range : test ( Inactive )
 08:00 to 18:00 working-day
```

Configuring a Basic IPv4 ACL

Basic IPv4 ACLs filter packets based on source IP address. They are numbered in the range 2000 to 2999.

Configuration Prerequisites

If you want to reference a time range to 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 and enter basic IPv4 ACL view	**acl number** *acl-number* [**match-order** { **auto** | **config** }]	Required The default match order is **config**.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } [**fragment** | **logging** | **source** { *sour-addr sour-wildcard* | **any** } | **time-range** *time-name* | **vpn-instance** *vpn-instance-name*] *	Required To create multiple rules, repeat this step.
Set a rule numbering step	**step** *step-value*	Optional The default step is 5.
Create an IPv4 ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

Note that:

- You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- When defining ACL rules, you need not always assign them IDs. The system can automatically assign rule IDs starting with 0 and increasing in certain rule numbering steps. A rule ID thus assigned is greater than the current highest rule ID. For example, if the rule numbering step is 5 and the current highest rule ID is 28, the next rule will be numbered 30. For detailed information about step, refer to the **step** command in the *Switch 8800 Command Reference Guide*.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first order rather than by rule number.

CAUTION:

- *You can modify the match order of an ACL with the **acl number** acl-number **match-order** { **auto** | **config** } command but only when it does not contain any rules.*
- *The rule specified in the **rule comment** command must have existed.*
- *For common I/O Modules, matching packets against an ACL rule with the **VPN-Instance** keyword or the **logging** keyword specified is not supported.*

Configuration Example

Create IPv4 ACL 2000 to deny the packets with source address 1.1.1.1 to pass.

```
<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, 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 filter packets based on source IP address, destination IP address, protocol carried on IP, and other protocol header fields, such as the TCP/UDP source port, TCP/UDP destination port, 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 to 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 and enter advanced IPv4 ACL view	**acl number** *acl-number* [**match-order** { **auto** | **config** }]	Required The default match order is **config**.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [**destination** { *dest-addr dest-wildcard* | **any** } | **destination-port** *operator port1* [*port2*] | **dscp** *dscp* | **established** | **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-name* | **tos** *tos* | **vpn-instance** *vpn-instance-name*] *	Required To create multiple rules, repeat this step.
Set a rule numbering step	**step** *step-value*	Optional The default step is 5.
Create an IPv4 ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.

When defining ACL rules, you need not assign them IDs. The system can automatically assign rule IDs, starting with 0 and increasing in certain rule numbering steps. A rule ID thus assigned is greater than the current highest rule ID. For example, if the rule numbering step is five and the current highest rule ID is 28, the next rule will be numbered 30. For detailed information about step, refer to the **step** command in the *Switch 8800 Command Reference Guide*.

You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.

CAUTION:

- *You can modify the match order of an ACL with the **acl number** acl-number **match-order** { **auto** | **config** } command but only when it does not contain any rules.*
- *The rule specified in the **rule comment** command must have existed.*
- *For common I/O Modules, matching packets against an ACL rule with the **VPN-Instance** keyword or the **logging** keyword specified is not supported.*
- *For common interface cards, matching packets against an ACL rule with the **reflective** keyword specified is not supported.*

Configuration Example

Create IPv4 ACL 3000, permitting TCP packets with port number 80 sent from 129.9.0.0 to 202.38.160.0 to pass.

```
<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, 1 rule,
Acl’s step is 5
 rule 0 permit tcp source 129.9.0.0 0.0.255.255 destination 202.38.1
60.0 0.0.2.255 destination-port eq www (5 times matched)
```

Configuring an Ethernet Frame Header ACL

Ethernet frame header ACLs filter 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 to 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 and enter Ethernet frame header ACL view	**acl number** *acl-number* [**match-order** { **auto** | **config** }]	Required The default match order is **config**.
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-name* **| type** *type-code type-wildcard*] *	Required To create multiple rules, repeat this step.
Set a rule numbering step	**step** *step-value*	Optional The default step is 5.
Create an ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.

When defining ACL rules, you need not assign them IDs. The system can automatically assign rule IDs, starting with 0 and increasing in certain rule numbering steps. A rule ID thus assigned is greater than the current highest rule ID. For example, if the rule numbering step is five and the current highest rule ID is

28, the next rule will be numbered 30. For detailed information about step, refer to the **step** command in the *Switch 8800 Command Reference Guide*.

You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first order rather than by rule number.

CAUTION:

- *You can modify the match order of an ACL with the* ***acl number*** *acl-number* ***match-order*** *{* ***auto*** *|* ***config*** *} command but only when it does not contain any rules.*
- *The rule specified in the* ***rule comment*** *command must have existed.*
- *When you create an Ethernet frame header ACL, do not specify the* ***LSAP*** *keyword or set the type-code argument to 0800, 86DD, 8847, 8848 or 8100.*
- *For the default flow template to be used, the destination mask corresponding to the* ***type*** *keyword must be FFFF.*

Configuration Example

Create IPv4 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, 1 rule,
Acl’s step is 5
rule 0 deny cos excellent-effort(5 times matched)
```

Configuring a User-Defined ACL

User-defined ACLs allow you to customize rules based on information of protocol headers such as IP. When defining a user-defined ACL rule, you need to specify an offset in bytes on which a match operation should start from the beginning of a packet header and in addition, specify a mask. When comparing a packet against the rule, the system ANDs the mask with the corresponding bytes in the packet and compare the result with the rule.

User-defined ACLs are numbered in the range 5000 to 5999.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure a user-defined IPv4 ACL:

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Create and enter user-defined IPv4 ACL view	**acl number** *acl-number*	Required

To do...	Use the command...	Remarks
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } [{ **ipv4** | **ipv6** | **l2** | **l4** | **l5**] *rule-string rule-mask offset* }&<1-8>] [**time-range** *time-name*]	Required To create multiple rules, repeat this step.
Create an ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

You will fail to create a user-defined ACL rule if its permit/deny statement is exactly the same as another rule. Unlike other types of ACLs, however, you can modify a user-defined ACL rule.

When defining user-defined ACL rules, you need not assign them IDs. The system can automatically assign rule IDs starting with 0 and increasing in rule numbering steps of five. A rule ID thus assigned is greater than the current highest rule ID. For example, if the current highest rule ID is 28, the next rule will be numbered 30. For detailed information about step, refer to the **step** command in the *Switch 8800 Command Reference Guide*.

For a user-defined ACL, the match order can only be **config**.

For D-type modules, matching packets against a user-defined ACL that the *offset* is set from the beginning of the Layer 5 header is not supported.

***CAUTION:** The rule specified in the **rule comment** command must have existed.*

Configuration Example

Configure used-defined ACL 5500.

```
<Sysname> system-view
[Sysname] acl number 5500
[Sysname-acl-user-5500] rule 0 permit l2 0806 ffff 20 time-range t1
```

# Verify the configuration.

```
[Sysname-acl-user-5500] display acl 5500
User defined ACL  5500, 1 rule,
Acl's step is 5
 rule 0 permit l2 0806 ffff 20 time-range t1 (Active)
```

Displaying and Maintaining IPv4 ACLs

To do...	Use the command...	Remarks
Display information about a specified or all IPv4 ACLs	**display acl** { *acl-number* | **all** | **name** *acl-name* }	Available in any view
Display the configuration and state of a specified or all time ranges	**display time-range** { *time-name* | **all** }	Available in any view
Clear the statistics about the specified or all IPv4 ACLs except for user-defined IPv4 ACLs	**reset acl counter** { *acl-number* | **all**}	Available in user view

IPv4 ACL Configuration Examples

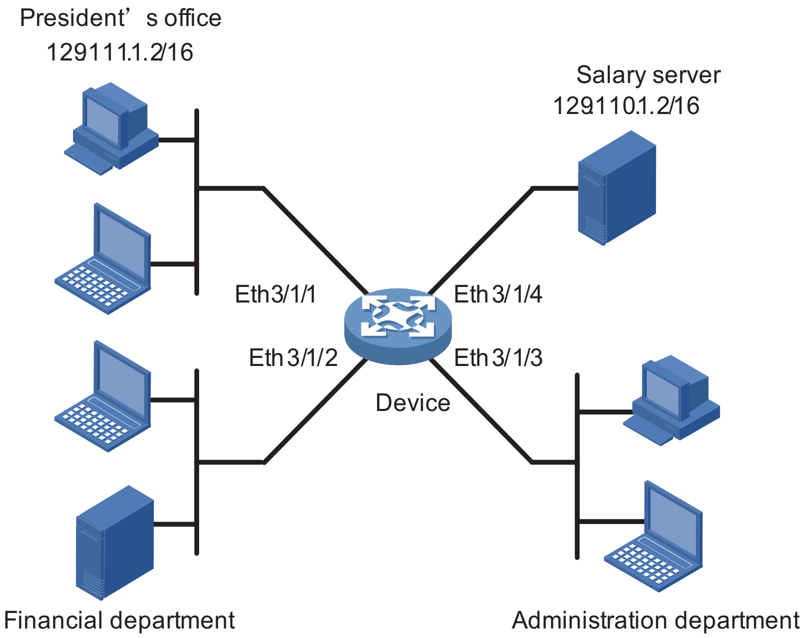
IPv4 ACL Configuration Examples

Network Requirements

A company interconnects its departments through the Device. The President's Office uses IP address 129.111.1.2; the salary server of the Finance Department uses IP address 129.110.1.2.

Configure an ACL to deny accesses of all departments but the President's Office to the salary server during office hours from 8:00 to 18:00 in working days.

Network Diagram



Configuration Procedure

1 Create a time range for office hours

# Create a periodic time range spanning 8:00 to 18:00 in working days.

```
<Sysname> system-view
[Sysname] time-range trname 8:00 to 18:00 working-day
```

2 Define an ACL to control accesses to the salary server

# Create and enter the view of ACL 3000.

```
[Sysname] acl number 3000
```

# Create a rule to control access of the President's Office to the salary server.

```
[Sysname-acl-adv-3000] rule 1 permit ip source 129.111.1.2 0.0.0.0
[Sysname-acl-adv-3000] quit
```

# Create a rule to control accesses of other departments to the salary server.

```
[Sysname] acl number 3001
[Sysname-acl-adv-3001] rule 1 deny ip source any destination 129.110
.1.2 0.0.0.0 time-range trname
[Sysname-acl-adv-3001] quit
```

3 Define and apply a QoS policy on interfaces

# Configure traffic classification rules and traffic behaviors.

```
[Sysname] traffic classifier test_permit
[Sysname-classifier-test_permit] if-match acl 3000
[Sysname-classifier-test_permit] quit
[Sysname] traffic behavior test_permit
[Sysname-behavior-test_permit] filter permit
[Sysname-behavior-test_permit] quit
[Sysname] traffic classifier test_deny
[Sysname-classifier-test_deny] if-match acl 3001
[Sysname-classifier-test_deny] quit
[Sysname] traffic behavior test_deny
[Sysname-behavior-test_deny] filter deny
[Sysname-behavior-test_deny] quit
```

# Configure a QoS policy.

```
[Sysname] qos policy test
[Sysname-qospolicy-test] classifier test_permit behavior test_permit
[Sysname-qospolicy-test] classifier test_deny behavior test_deny
[Sysname-qospolicy-test] quit
```

# Apply the QoS policy at the inbound direction of Ethernet 3/1/1, Ethernet 3/1/2 and Ethernet 3/1/3.

```
[Sysname] interface Ethernet 3/1/1
[Sysname-Ethernet3/1/1] qos apply policy test inbound
[Sysname-Ethernet3/1/1] quit
[Sysname] interface Ethernet 3/1/2
[Sysname-Ethernet3/1/2] qos apply policy test inbound
[Sysname-Ethernet3/1/2] quit
[Sysname] interface Ethernet 3/1/3
[Sysname-Ethernet3/1/3] qos apply policy test inbound
[Sysname-Ethernet3/1/3] quit
```

58 IPv6 ACL Configuration

When configuring IPv6 ACLs, go to these sections for information you are interested in:

- "Creating a Time Range" on page 815
- "Configuring a Basic IPv6 ACL" on page 815
- "Configuring an Advanced IPv6 ACL" on page 816
- "Displaying and Maintaining IPv6 ACLs" on page 818
- "IPv6 ACL Configuration Examples" on page 818

Creating a Time Range

Refer to section "Creating a Time Range" on page 805

Configuring a Basic IPv6 ACL

Basic IPv6 ACLs filter packets based on source IPv6 address. They are numbered in the range 2000 to 2999.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an IPv6 ACL:

Table 33 Network diagram for ACL configuration

To do...	Use the command...	Remarks
Enter system view	**system-view**	--
Create and enter basic IPv6 ACL view	**acl ipv6 number** *acl6-number* [**match-order** { **auto** | **config** }]	Required The default match order is **config**.
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-name*] *	Required To create multiple rules, repeat this step.
Set a rule numbering step	**step** *step-value*	Optional The default step is 5.
Create an IPv6 ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.

When defining ACL rules, you need not assign them IDs. The system can automatically assign rule IDs starting with 0 and increasing in certain rule numbering steps. A rule ID thus assigned is greater than the current highest rule ID. For example, if the rule numbering step is five and the current highest rule ID is 28, the next rule will be numbered 30.

You may use the **display acl ipv6** command to verify rules configured in an ACL. If the match order for this IPv6 ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.

CAUTION:

- *You can modify the match order of an IPv6 ACL with the **acl ipv6 number** acl6-number **match-order** { **auto** | **config** } command but only when it does not contain any rules.*
- *The rule specified in the **rule comment** command must have existed.*

Configuration Example

Create IPv6 ACL 2000 to permit IPv6 packets with source address 2030:5060::9050/64 to pass while denying IPv6 packets with source address 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, 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 ACLs filter packets based on the source IPv6 address, destination IPv6 address, protocol carried on IP, and other protocol header fields such as the TCP/UDP source port, TCP/UDP destination port, 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 to 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**	--

To do...	Use the command...	Remarks
Create and enter advanced IPv6 ACL view	**acl ipv6 number** *acl6-number* [**match-order** { **auto** | **config** }]	Required The default match order is **config**.
Create or modify a rule	**rule** [*rule-id*] { **deny** | **permit** } *protocol* [**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-name*] *	Required To create multiple rules, repeat this step.
Set a rule numbering step	**step** *step-value*	Optional The default step is 5.
Create an ACL description	**description** *text*	Optional
Create a rule description	**rule** *rule-id* **comment** *text*	Optional

You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.

When defining ACL rules, you need not assign them IDs. The system can automatically assign rule IDs, starting with 0 and increasing in certain rule numbering steps. A rule ID thus assigned is greater than the current highest rule ID. For example, if the rule numbering step is 5 and the current highest rule ID is 28, the next rule will be numbered 30.

You may use the **display acl ipv6** command to verify rules configured in an IPv6 ACL. If the match order for this IPv6 ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.

CAUTION:

- *You can modify the match order of an IPv6 ACL with the* ***acl ipv6 number*** *acl6-number* ***match-order*** *{* ***auto*** *|* ***config*** *} command but only when it does not contain any rules.*
- *The rule specified in the* ***rule comment*** *command must have existed.*
- *When creating an IPv6 ACL rule, you cannot specify the* ***fragment*** *keyword and the protocol argument at the same time.*

Configuration Example

Create IPv6 ACL 3000 to permit the TCP packets with the source address 2030:5060::9050/64 to pass.

```
<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, 1 rule,
 Acl’s step is 5
 rule 0 permit tcp source 2030:5060::9050/64 (5 times matched)
```

Displaying and Maintaining IPv6 ACLs

To do...	Use the command...	Remarks
Display information about a specified or all IPv6 ACLs	**display acl ipv6** { *acl6-number* | **all** | **name** *acl6-name* }	Available in any view
Display the configuration and status about the specified or all time ranges	**display time-range** { *time-name* | **all** }	Available in any view
Clear the statistics about a specified or all IPv6 ACLs	**reset acl ipv6 counter** { *acl6-number* | **all** }	Available in user view

IPv6 ACL Configuration Examples

IPv6 Configuration Examples

Network Requirements

Perform packet filtering in the inbound direction of interface Ethernet 1/3/1 to deny all IPv6 packets but those with source addresses in the range 4050::9000 to 4050::90FF.

Configuration Procedure

# Create an IPv6 ACL 2000 as follows: # Enter system view.

```
<Sysname> system-view
```

# Create an ACL rule, permitting the packets with the source IP addresses in the range 4050::9000 to 4050::90FF.

```
[Sysname] acl ipv6 number 2000
[Sysname-acl6-basic-2000] rule permit source 4050::9000/120
```

# Create an ACL rule, denying the packets with any source IP addresses.

```
[Sysname] acl ipv6 number 2001
[Sysname-acl6-basic-2001] rule deny source any
```

Configure IPv6 packet filtering at the inbound direction of Ethernet 3/1/1

# Configure a traffic classification rule and a traffic behavior, permitting the packets with the source IP addresses in the range 4050::9000 to 4050::90FF.

```
[Sysname] traffic classifier c_permit
[Sysname-classifier-c_permit] if-match acl ipv6 2000
[Sysname-classifier-c_permit] quit
[Sysname] traffic behavior b_permit
[Sysname-behavior-b_permit] filter permit
[Sysname-behavior-b_permit] quit
```

# Configure a traffic classification rule and a traffic behavior, denying the packets with any source IP addresses.

```
[Sysname] traffic classifier c_deny
[Sysname-classifier-c_deny] if-match acl ipv6 2001
[Sysname-classifier-c_deny] quit
[Sysname] traffic behavior b_deny
[Sysname-behavior-b_deny] filter deny
[Sysname-behavior-b_deny] quit
```

# Configure and apply the QoS policy.

```
[Sysname] qos policy test
[Sysname-qospolicy-test] classifier c_permit behavior b_permit
[Sysname-qospolicy-test] classifier c_deny behavior b_deny
[Sysname-qospolicy-test] quit
```

Configure IPv6 packet filtering at the inbound direction of Ethernet 1/3/1

```
[Sysname] inter Ethernet1/3/1
[Sysname- Ethernet1/3/1] qos apply policy test inbound
[Sysname- Ethernet1/3/1] quit
[Sysname]
```

59 FLOW TEMPLATE CONFIGURATION

This chapter covers these topics:

- "Configuring a Flow Template" on page 821
- "Displaying and Maintaining Flow Templates" on page 824
- "Flow Template Configuration Examples" on page 824

Configuring a Flow Template

Follow these steps to create a flow template and apply it to an interface:

To do...		Use the command...	Remarks
Enter system view		**system-view**	--
Create a flow template	Create a basic flow template	**flow-template** *flow-template-name* **basic**{**customer-vlan-id** | **dip** | **dipv6** | **dmac** | **dport** | **dscp** | **ethernet-protocol** | **fragments** | **icmp-code** | **icmp-type** | **icmpv6-code** | **icmpv6-type** | **ip-precdence** | **ip-protocol** | **ipv6-dscp** | **ipv6-fragment** | **ipv6-protocol** | **service-cos** | **service-vlan-id** | **sip** | **sipv6** | **smac** | **sport** | **tcp-flag | tos** } *****	Optional Use either command.
	Create an extended flow template	**flow-template** *flow-template-name* **extend** { **ipv4** *offset-max-value length-max-value* | **ipv6** *offset-max-value length-max-value* | **l2** *offset-max-value length-max-value* | **l4** *offset-max-value length-max-value* | **l5** *offset-max-value length-max-value* } *	
Enter interface view or port group view	Enter interface view	**interface** *interface-type interface-number*	Required
	Enter port group view	**port-group** { **aggregation** *agg-id* | **manual** *port-group-name* }	

To do...	Use the command...	Remarks
Apply the flow template to the interface or port group	**flow-template** *flow-template-name*	Optional The default one applies by default.

CAUTION: *When one of the following situations occurs, you cannot configure user-defined flow templates on interfaces:*

- *B-type and C-type modules have IPv6 unicast and mix-insertion enabled on the virtual interfaces of VLANs.*
- *A Loopback interface for tunneling is configured on the ports of D-type modules.*
- *B-type and C-type modules have MLD, MLD Snooping and IPv6 PIM enabled on the virtual interfaces of VLANs.*
- *B-type, C-type and D-type modules have the RSVP protocol enabled on the virtual interfaces of VLANs.*
- *B-type, C-type and D-type modules are configured with VPLS redirection to the VPLS service modules.*

Once a user-defined flow template is configured on an interface, only after the corresponding functions are removed can the user-defined flow template be deleted.

- *By default, the default flow template is referenced on interfaces.*
- *On an interface you can reference only one flow template and this flow template must be one already created.*
- *The default flow template includes such fields as sip, dip, ip-protocol, sport, dport, tcp-flag, icmp-type, icmp-code, service-vlan-id, and ethernet-protocol.*
- *For modules suffixed with D, the default flow template also includes sipv6, dipv6, ipv6-protocol, and ipv6-fragment fields.*
- *Referencing a user-defined flow template on an interface may cause VLAN ACLs not function at the inbound direction of the interface.*
- *Configuring user-defined flow templates is prohibited on the even numbered interface of the 3C17526 module. However, user-defined flow templates configured on the odd numbered ports will function at the outbound direction of all ports.*
- *User-defined flow templates are not available for the POS interface.*

The total size of all fields (in bytes) in a flow template should be less than 16 bytes. Table 34 shows the size of every field.

Table 34 Description on the size of every field

Field	Byte number	Remarks
customer-vlan-id	4 or 8	Usually 8 bytes (4 bytes when the ethernet-protocol field is configured)
dip	4	-

Table 34 Description on the size of every field

Field	Byte number	Remarks
dipv6	10	10 bytes in a flow template In fact, the field is 16-byte long.
dmac	6	-
dscp	1	1 byte, no matter these three fields are configured respectively or together
ip-precedence	1	
tos	1	
ethernet-protocol	6	-
fragments	0 or 2	0 byte for B-type or C-type modules; 2 bytes for D-type modules
icmp-type	2	2 bytes, no matter they are configured respectively or together; Total 2 bytes when they are configured with the sport field; Total 2 bytes (for B-type or C-type modules) or 4 bytes (for D-type modules) when they are configured with the dport field.
icmp-code	2	
icmpv6-type	2	2 bytes, no matter they are configured respectively or together; Total 2 bytes when they are configured with the sport field.
icmpv6-code	2	
sport	2	-
dport	2	-
ip-protocol	1	-
ipv6-dscp	0	-
ipv6-fragment	1	1 byte, no matter they are configured respectively or together
ipv6-protocol	1	
service-cos	0, 2 or 4	0 byte when the customer-vlan-id field is configured; 2 bytes when the ethernet-protocol field is configured; or 4 bytes
service-vlan-id	0 or 2	2 bytes for B-type or C-type modules; 0 byte for D-type modules
sip	4	-
sipv6	0	10 bytes in a flow template In fact, the field is 16-byte long.
smac	6	-
tcp-flag	1	-

Displaying and Maintaining Flow Templates

To do...	Use the command...	Remarks
Display the configuration information of a specified or all user-defined flow templates	**display flow-template user-defined** [*flow-template-name*]	Available in any view
Display information about the flow templates referenced to interfaces.	**display flow-template interface** [*interface-type interface-number*]	Available in any view

Flow Template Configuration Examples

Network requirements

Create flow templates and apply them to interfaces.

Configuration procedure

# Create basic flow template **aaa** and extended flow template **bbb**.

```
<Sysname> system-view
[Sysname] flow-template aaa basic customer-vlan-id
[Sysname] flow-template bbb extend 14 2 4 15 10 7
```

# Reference flow template aaa on interface Ethernet 4/2/1 and flow template bbb on interface Ethernet 3/2/1.

```
[Sysname] interface ethernet 4/2/1
[Sysname-Ethernet4/2/1] flow-template aaa
[Sysname-Ethernet4/2/1] quit
[Sysname] interface ethernet 3/2/1
[Sysname-Ethernet3/2/1] flow-template bbb
[Sysname-Ethernet3/2/1] quit
```

# Display information about flow template aaa.

```
[Sysname] display flow-template user-defined aaa
user-defined flow template: basic
 name:aaa, index:1, total reference counts:1
 fields: customer-vlan-id
```

# Display information about all user-defined flow templates.

```
[Sysname] display flow-template user-defined
user-defined flow template: basic
 name:aaa, index:1, total reference counts:1
 fields: customer-vlan-id
user-defined flow template: extend
 name:bbb, index:2, total reference counts:1
 fields: 14 2 4 15 10 7
```

# Display information about the user-defined flow templates referenced to interfaces.

```
[Sysname] display flow-template interface
Interface: Ethernet4/2/1
user-defined flow template: basic
 name:aaa, index:1, total reference counts:1
 fields: customer-vlan-id
```

```
Interface: Ethernet3/2/1
user-defined flow template: extend
 name:bbb, index:2, total reference counts:1
 fields: 14 2 4 15 10 7
```

# Delete flow template aaa. As it is being referenced by interface Ethernet 1/0, remove it from the interface first.

```
[Sysname] interface ethernet 4/2/1
[Sysname-Ethernet4/2/1] undo flow-template
[Sysname-Ethernet4/2/1] quit
[Sysname] undo flow-template name aaa
```

60 QoS Overview

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

Introduction

Quality of Service (QoS) measures the service performance of service providers in terms of client satisfaction. Instead of giving accurate marks, QoS emphasizes analyzing what good or imperfect services are, and they come in what kind of circumstances, so as to provide a cutting edge improvement.

In the Internet, QoS evaluates service performance for network packet forwarding. Due to various services offered by the network, the evaluation for QoS will be based on different aspects accordingly. Generally, QoS evaluates the service performance for those network core requirements during packet forwarding process, such as delay, jitter and packet loss ratio.

Traditional Packets Forwarding Application

On traditional IP networks, the devices treat all packets identically and handle them with the first in, first out (FIFO) policy, assigning forwarding resources by arrival sequence of packets. All the packets share the resources of the network devices. How many resources the packets can obtain will completely depend on the time they arrive.

This service policy is called Best-effort, which delivers the packets to their destination as it can, without any assurance and guarantee for delivery delay, jitter, packet loss ratio, reliability and so on for packet forwarding.

The traditional Best-Effort service policy is only suitable for applications insensitive to bandwidth and delay, such as WWW, file transfer and e-mail.

New Requirements Caused by New Applications

With the fast development of the network, more and more networks access the Internet. The Internet has been expanded in terms of its scale, coverage and users quantities. More and more users use Internet as their data transmission platform to implement various applications.

Apart from traditional applications of WWW, e-mail and FTP, network users try to expand some new applications, such as tele-education, telemedicine, video telephone, videoconference and Video-on-Demand (VoD), on the Internet. And the enterprise users expect to connect their regional branches together to develop some operational applications through VPN technology, for instance, to access the database of the company or monitor their remote equipment via Telnet.

Those new applications have one thing in common, i.e. high requirements for bandwidth, delay, and jitter. For instance, videoconference and VOD need the assurance of wide bandwidth, low delay and jitter. As for mission-critical applications, such as transaction and Telnet, they may not require wide bandwidth but do require lower delay and be handled by priority during congestion.

The new emerging applications demand higher service performance of IP network. Better network services during packets forwarding are required other than simply delivering the packets to their destination, such as providing user-specific bandwidth, reducing packet loss ratio, avoiding congestion, regulating network traffic, setting priority of the packets. To meet those requirements, the network should be provided with better service capability.

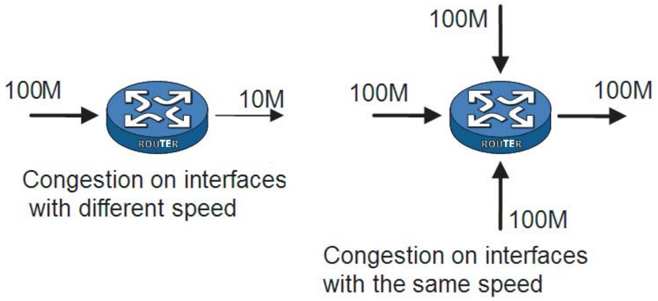
Congestion: Causes, Impact, and Countermeasures

Network congestion is a key factor to degrade the service quality of the traditional network. Congestion refers to such a fact that the service rates are decreased due to relative deficiency of the resources supply (leading to extra delay).

Causes

Congestion will easily occur in complex packet switching circumstances in the Internet, with two cases illustrated in the following figure:

Figure 246 Traffic congestion causes


100M
10M
ROUTER
Congestion on interfaces with different speed
100M
100M
100M
ROUTER
100M
Congestion on interfaces with the same speed

1 The packet streams enter a device from a high speed link and are forwarded via a low speed link;

2 The packet streams enter a device from several interfaces with a same speed and are forwarded through an interface with the same speed as well.

When traffic arrives at wire speed, congestion may occur for network resource bottleneck.

Besides the bottleneck of link bandwidth, congestion will also be caused by resources deficiency in normal packet forwarding, such as the deficiency of assignable processor time, buffer and memory. In addition, congestion may occur if the arrival traffic is not managed efficiently and the assignable network resources are inadequate.

Impact

Congestion may cause the following negative effects:

- Increase the delay and jitter of packet transmission
- Packet re-transmission caused by high delay
- Decrease the efficient throughput of network and lower the utilization of network resources
- Intensified congestion can occupy too many network resources (especially in memory), and the irrational assignment of resources even can lead to resource block and breakdown for the system

It is obvious that congestion will make the traffics unable to obtain the resources in time and degrade the service performance accordingly. No one wants congestion, but it occurs frequently in complex environments where packet switching and multi-users applications coexist. So it needs to be treated cautiously.

Countermeasure

A direct way to solve resources deficiency problem is to increase the bandwidth of network; however, it cannot resolve all the problems caused by congestion.

A more effective method to solve the problem of QoS is to enhance the functions of traffic control and resource allocation in the network, and to provide differentiated services for applications with different service requirement in order to allocate and use resources rightly. During the process of resources allocation and traffic control, the direct or indirect factors that might cause network congestion should be controlled with best effort to reduce the probability of congestion. As congestion occurs, resource allocation should be balanced according to features and demands of applications, to minimize the effects on QoS by congestion.

Traffic Management Technologies

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: It is a prerequisite for differentiated service, to identify the interested objects based on a certain matching rule.
- Traffic policing: polices the specification of particular traffics entering the switch. When the traffics exceed the specification, then some restriction or punishment measures can be taken to protect the commercial benefits of carriers and to prevent network resources from being damaged.
- Traffic shaping: A traffic control measure of actively adjusting the output speed of traffics, generally it can enable the traffic to adapt to the network resources supplied by the downstream switch, to prevent the unwanted packet dropping and congestion. Same as traffic policing, traffic shaping is implemented at the IP layer.
- Congestion management: handles resource competition during network congestion. Generally, it stores the packets in the queue first, and then takes a dispatching algorithm to assign the forwarding sequence of packets.
- Congestion avoidance: Exceeding congestion consumes network resources. Congestion avoidance can monitor the usage status of network resources, and as congestion becomes worse, actively take the policy of dropping packets through adjusting traffic to resolve the overloading of the network.

Among those traffic management technologies, traffic classification is the basis. It is a prerequisite for differentiated services, which identifies the interested packets with certain matching rule. As for traffic policing, traffic shaping, congestion management and congestion avoidance, they implement management to network traffic and allocated resources from different aspects respectively to realize the differentiated service.

Normally, QoS provides the following functions:

- Traffic classification
- Access control
- Traffic policing and shaping
- Congestion management
- Congestion avoidance

61 TRAFFIC CLASSIFICATION AND TRAFFIC SHAPING CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "Traffic Classification Overview" on page 831
- "Traffic Shaping Overview" on page 832
- "Traffic Evaluation and Token Bucket" on page 832
- "Traffic Shaping Configuration" on page 835

Traffic Classification Overview

Traffic classification

Traffic classification is the prerequisite and foundation for differentiated services, which uses certain rules to identify the packets with certain features.

To discriminate flows, you can set traffic classification rules using the priority bits of ToS (type of service) field in the IP packet header. Alternatively, the network administrator may define a traffic classification policy, for instance, integrating information such as source IP address, destination IP address, MAC address, IP protocol, or port number of the applications to classify the traffic. In general, it can be a narrow range defined by a quintuple (source IP address, source port number, destination IP address, destination port number and the Transport Protocol), or can be all packets to a network segment.

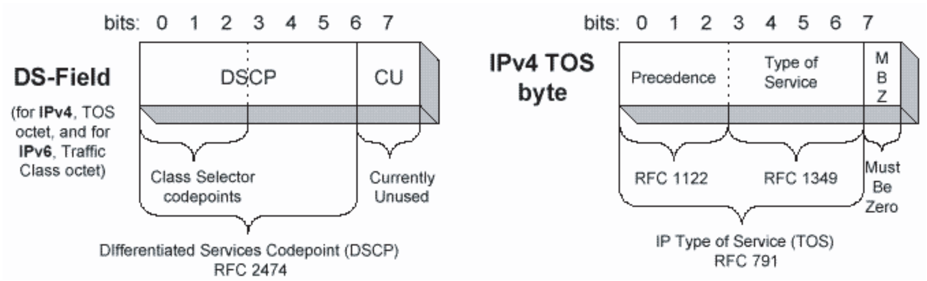
In general, while packets being classified on the network border, the precedence bits in the ToS byte of IP header are set so that IP precedence can be used as a direct packet classification standard within the network. The queuing technologies can use IP precedence to handle the packets. Downstream network can receive the packets classification results from upstream network selectively, or re-classify the packets with its own standard.

Traffic classification is used to provide differentiated service, so it must be associated with certain kinds of traffic policing or resource-assignment mechanisms. To adopt what kind of traffic policing action will depend on the current stage and load status of the network. For example, to police the packets according to the committed rate when they enter the network, to make traffic shaping before they flow out the nodes, to perform queuing management in the event of congestion and to employ congestion avoidance when congestion becomes worse.

Priority

Several priorities are described as follows:

Figure 247 DS field and ToS byte


bits: 0 1 2 3 4 5 6 7
DS-Field
(for IPv4, TOS octet, and for IPv6, Traffic Class octet)
DSCP
CU
Class Selector codepoints
Currently Unused
Differentiated Services Codepoint (DSCP) RFC 2474
bits: 0 1 2 3 4 5 6 7
IPv4 TOS byte
Precedence
Type of Service
M B Z
RFC 1122
RFC 1349
Must Be Zero
IP Type of Service (TOS) RFC 791

As shown in Figure 247, the ToS byte of IP header contains 8 bits: the first three bits (0 to 2) indicates IP precedence, valued in the range 0 to 7; the following 4 bits (3 to 6) indicates ToS priority, valued in the range 0 to 15. In RFC2474, the ToS field of IP packet header is redefined as DS field, where the DiffServ code point (DSCP) priority is indicated by the first 6 bits (0 to 5), valued in the range 0 to 63. The remaining 2 bits (6 and 7) are reserved.

Traffic Shaping Overview

If no restrictions are imposed on the traffics from the users, bursting data sent by mass users continuously will make the network become more congested. Thus for more efficient network function and better network service for more users, the traffics from the users must be restricted, for example, to restrict a traffic can only acquire the specific assigned resources in certain time interval so as to prevent the network congestion caused by excess burst.

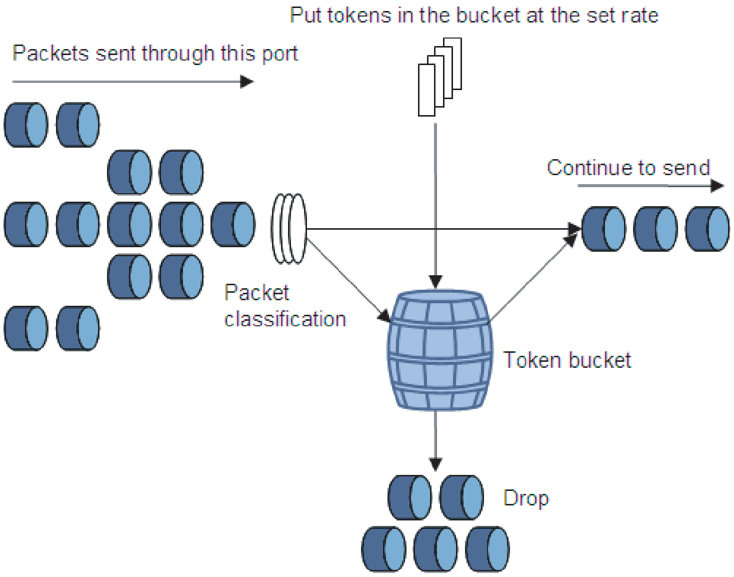
Traffic shaping is a traffic monitoring policy to restrict the traffic and resources through comparing with the traffic specification. To know whether the traffic exceeds the specification or not is a prerequisite for traffic shaping. Then based upon the evaluation result you can implement a regulation policy. Usually, Token Bucket is used to value the traffic specification.

Traffic Evaluation and Token Bucket

Token bucket features

Token Bucket can be regarded as a container to reserve Token, which has certain capacity. The system will put Tokens into the Bucket at a defined rate. In case the Bucket is full, the extra Tokens will overflow and no more Tokens will be added.

Figure 248 Measuring the traffic with Token Bucket


Put tokens in the bucket at the set rate
Packets sent through this port
Continue to send
Packet classification
Token bucket
Drop

Measuring the traffic with Token Bucket

Whether or not the token quantity of the Token Bucket can satisfy the packets forwarding is the basis for Token Bucket to measure the traffic specification. If enough tokens are available for forwarding packets, traffic is regarded conforming the specification (generally, one token is associated to the forwarding ability of one bit), otherwise, non-conform or excess.

When measuring the traffic with Token Bucket, these parameters are included:

- Mean rate: The rate of putting Token into Bucket, i.e. average rate of the permitting traffic. Generally set as CIR (Committed Information Rate).
- Burst size: Token Bucket's capability, i.e. the maximum traffic size of every burst. Generally, it is set as CBS (Committed Burst Size), and the bursting size must be greater than the maximum packets size.

A new evaluation will be made when a new packet arrives. If there are enough tokens in bucket for each evaluation, it shows that traffics are within the bound, and at this time the amount of tokens appropriate for the packets forwarding rights, need to be taken out. Otherwise, it shows that too many tokens have been used, and traffic specifications are exceeded.

Complicated evaluation

Two Token Buckets can be configured to evaluate conditions that are more complex and to implement more flexible regulation policy. For example, Traffic Policing (TP) has four parameters, as follows:

- CIR (Committed information rate)
- CBS (Committed burst size)
- PIR (Peak information rate)
- EBS (Excess burst size)

It uses two token buckets, with the token-putting rate of every bucket set as CIR and PIR and the capability of every bucket set as CBS and EBS (CBS < EBS, called C bucket and E bucket), which represents different bursting class permitted. In each evaluation, you may use different traffic control policies for different situations, such as "C bucket has enough tokens"; "Tokens of C bucket are deficient, but those of E bucket are enough"; "Tokens of C bucket and E bucket are all deficient".

Traffic policing

Typically, traffic policing is used to monitor the specification of certain traffic entering the network and keep it within a reasonable bound, or it will make "penalty" on the exceeding traffic so as to protect network resources and profits of carriers. For example, it can restrict HTTP packets to occupy network bandwidth of no more than 50%. Once finding the traffic of a connection exceeds, it may drop the packets or reset the precedence of packets.

Traffic policing allows you to define match rules based on IP precedence or DiffServ code point (DSCP). It is widely used by ISP to police the network traffic. TP also includes the traffic classification service for the policed traffics, and depending upon the different evaluation results, it will implement the pre-configured policing actions, which are described as the following:

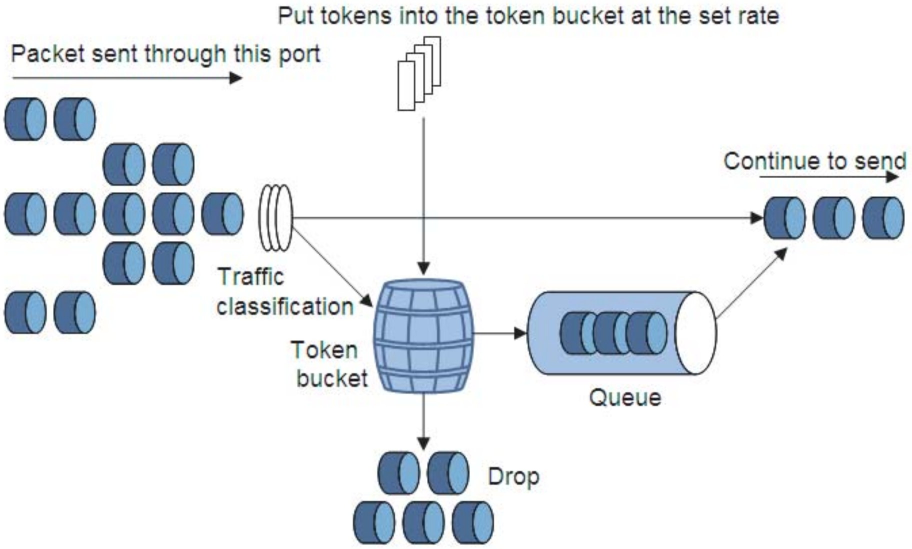
- Forward: For example, continue to forward the packets evaluated as "conform".
- Drop: For example, dropping the packets evaluated as "not conform".

Traffic shaping

Traffic shaping is an active way to adjust the traffic output rate.

The main difference between traffic shaping and traffic policing is: the packets to be dropped in traffic policing will be stored during traffic shaping - generally they will be put into buffer or queues, as shown in Figure 249. Once there are enough tokens in token bucket, those stored packets will be evenly sent. Another difference is that traffic shaping may intensify delay, yet traffic policing seldom does so.

Figure 249 TS diagram


Put tokens into the token bucket at the set rate
Packet sent through this port
Continue to send
Traffic classification
Token bucket
Queue
Drop

For example, Switch A sends packets to Switch B. Switch B implements TP on those packets, and directly drops exceeding traffic.

To reduce unnecessary packet drop, GTS can be applied to the packets on the egress interface of Switch A. The packets beyond the traffic specifications of GTS are stored in Switch A. While sending the next set of packets, GTS takes out those packets from buffer queues and send them. Thus, all the packets sent to Switch B accord with the traffic specification of Switch B.

- *Traffic shaping cannot be configured on ports with even port numbers on the 3C17526 module, and traffic shaping configured on a port with an odd port number takes effect on the incoming packets of both the port with the odd port number and the port with the even port number (the odd port number plus one).*
- *Traffic shaping is not available to POS interfaces.*

Traffic Shaping Configuration

Configuring Traffic Shaping

Traffic shaping includes the following two types:

- Queue-based traffic shaping: set traffic shaping parameters for packets in a queue.
- Traffic shaping applicable to all the traffic: set traffic shaping parameters for all the traffic.

Configuring queue-based traffic shaping

Follow these steps to configure queue-based traffic shaping:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure traffic shaping on the specific port or ports in the specific port group		**qos gts queue** *queue-number* **cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]]	Required Generally, CBS is CIR*62.5. EBS is 0 by default.
Display the traffic shaping configuration information		**display qos gts interface** [*interface-type interface-number*]	Optional Available in any view

- *In traffic shaping configuration, the outgoing ports of 3C17526, 3C17532, and 3C17538 modules support four queues, that is, the queue-number argument is in the range of 0 to 3; this argument is in the range of 0 to 7 for the other modules.*
- *For the description on the default value of CBS in the Switch 8800 Command Reference Guide.*

Configuring traffic shaping applicable to all traffics

Follow these steps to configure traffic shaping applicable to all traffics:

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure traffic shaping on the specific port or ports in the specific port group		**qos gts any cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]]	Required Generally, CBS is CIR*62.5. EBS is 0 by default.
Display the traffic shaping configuration information		**display qos gts interface** [*interface-type interface-number*]	Optional Available in any view

For the description on the default value of CBS, refer to the Switch 8800 Command Reference Guide.

Traffic shaping configuration example

Configure TS on Ethernet 1/1/1 to shape the outgoing packets with traffic rate exceeding 500 kbps on the port.

# Enter system view.

```
<Sysname> system-view
```

# Enter Ethernet interface view.

```
[Sysname] interface ethernet1/1/1
```

# Configure traffic shaping parameters.

```
[Sysname-Ethernet1/1/1] qos gts any cir 500
```

62 QoS Policy Configuration

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "QoS Policy Overview" on page 839
- "QoS Policy Configuration Procedure" on page 840
- "Configuring QoS Policy" on page 840
- "Displaying and Maintaining QoS Policy" on page 846

QoS Policy Overview

A QoS policy includes three elements: class, traffic behavior, and policy. You can bind the specified class to a traffic behavior through the QoS policy so as to configure QoS conveniently.

Class

A class is used to identify traffic.

Class elements include class name and rule.

You can define a series of rules by executing some commands to classify packets. Also you can use commands to specify the relationship between rules: **and** and **or**.

- and: The device considers a packet belongs to a class only when the packet matches all rules.
- or: The device considers a packet belongs to a class as long as the packet matches one of the rules in the class.

Traffic behavior

A traffic behavior is used to define QoS actions conducted for packets.

Traffic behavior elements include traffic behavior name and actions defined in the traffic behavior.

Users can use commands to define multiple actions in a traffic behavior.

Policy

A policy is used to bind a specific class to a specific traffic behavior.

Policy elements include policy name and name of the bound class and traffic behavior.

QoS Policy Configuration Procedure

Follow these steps to configure QoS policy:

1. Define the class and define a group of traffic classification rules in class view.
2. Define the traffic behavior and define a group of QoS actions in traffic behavior view.
3. Define the policy, and define the corresponding traffic behavior for the class in use in policy view.
4. Apply QoS policy.

Configuring QoS Policy

Configuration Prerequisites

- The class name and rules of the class are defined in a policy.
- The traffic behavior name and actions in the traffic behavior are defined,
- The policy name is defined.
- The interface to which policy is applied is defined.

Defining a Class

Before you can define a class, you must first create its class name. Then you can configure matching rules in this class view.

Configuration procedure

Follow these steps to define a class:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Define a class and enter the class view	**traffic classifier** *tcl-name* [**operator** { **and | or** }]	Required The **operator** keyword defaults to **and**. That is, the relation between the matching rules in the class view is logic AND.
Define rules to match packets	**if-match** *match-criteria*	Required
Display class information	**display traffic classifier user-defined** [*tcl-name*]	Optional Available in any view

Configuration example

1 Network requirements

Configure a class **test** to match packets with the destination MAC address 0050-ba27-bed3.

2 Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Define a class and enter class view.

```
[Sysname] traffic classifier test
```

# Configure the classification rule.

```
[Sysname-classifier-test] if-match destination-mac 0050-ba27-bed3
[Sysname-classifier-test]
```

*With the **operator** keyword set to **and**, the if-match statements or the parameters in an if-match statement cannot conflict with each other.*

Defining a Traffic Behavior

To define a traffic behavior, you must first create a traffic behavior name and then configure features for it in the traffic behavior view.

Configuration procedure

Follow these steps to define traffic behavior:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Define a traffic behavior and enter traffic behavior view	**traffic behavior** *behavior-name*	Required *behavior-name*: Traffic behavior name.

To do...	Use the command...	Remarks
Configure the traffic statistics action	**accounting**	Required Configure corresponding traffic behaviors as needed.
Configure the traffic policing action	**car cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**pir** *peak-information-rate*] [**red** *action*]	
Configure the traffic filtering action	**filter** { **deny** | **permit** }	
Configure the action of creating an outer VLAN tag	**nest top-most vlan-id** *vlan-id-value*	
Configure the 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* | **link-aggregation group** *group-number*]] } }	
Configure the DSCP marking action	**remark dscp** *dscp-value*	
Configure the 802.1p precedence marking action	**remark dot1p** *8021p*	
Configure the drop precedence marking action	**remark drop-precedence** *drop-precedence-value*	
Configure the local precedence marking action	**remark local-precedence** *local-precedence*	
Configure the action of marking service provider network VLAN ID	**remark service-vlan-id** *vlan-id-value*	
Configure the action of obtaining other precedence values through the corresponding priority mapping table	**primap pre-defined** { **dscp-lp** | **dscp-dp** | **dscp-dot1p** | **dscp-dscp** | **color+dscp-dscp** | **color+dscp-dp** | **color+dscp-lp** | **color+dscp-dot1p** | **color+lp-dot1p** }	
Display traffic behavior information	**display traffic behavior user-defined** [*behavior-name*]	Optional Available in any view

Note that:

- For the description on the default values of CIR, CBS, EBS, and PIR, refer to the *Switch 8800 Command Reference Guide*.
- When the traffic redirecting action is configured, if the outbound interface to be redirected to is bound with an NAT virtual interface, packets sent from this outbound interface are redirected to the L3+NAT card, thus resulting in traffic redirecting failure.
- Multiple traffic actions can be used together. How ever, the traffic filtering action defined by the **filter deny** command can only be used in conjunction

with the traffic statistics action, and the action of redirecting traffic to the CPU cannot be used in conjunction with any other traffic actions.

- The tunnel redirecting policy can be configured only in port view on D-type modules.
- Only D-type modules support the action of creating an outer VLAN tag and the action of marking service provider network VLAN ID.
- B-type modules cannot mark one CoS value separately. Four CoS values (dot1p, dscp, drop-precedence, and local-precedence) must be marked together on B-type modules.
- On B-type modules, the priority marking action cannot be configured in conjunction with the traffic policing action or the traffic statistics action.
- On all modules, in the action of obtaining other precedence values through a colored priority mapping table, the DSCP precedence and the 802.1p precedence must be modified at the same time and cannot be modified separately.
- On all the modules, the 802.1p precedence marking action, the local precedence marking action, and the drop precedence marking action cannot be configured with the action of obtaining other precedence values through an uncolored priority mapping table at the same time.
- When configuring the action of redirecting traffic to an aggregation group and applying the corresponding policy to a port, make sure that the action of marking service provider network VLAN ID is configured at the same time. If the corresponding policy is applied in VLAN view, the action of marking service provider network VLAN ID is unnecessary to be configured at the same time.
- When configuring the action of redirecting traffic to the next hop, make sure that the type of ACL rules is consistent with that of the address the traffic is to be redirected to. That is, you cannot redirect traffic to an IPv6 address using IPv4 ACL rules or redirect traffic to an IPv4 address using IPv6 ACL rules.
- The interface the traffic is to be redirected to must be an I/O Module interface.
- The aggregation CAR action cannot be configured in conjunction with the traffic statistics action.
- With the action of redirecting traffic to the CPU configured, any of the other traffic actions cannot be configured.
- The action of marking service provider network VLAN ID (or the action of creating an outer VLAN tag) cannot be configured with any traffic redirecting action except the action of redirecting traffic to an aggregation group.
- The action of marking service provider network VLAN ID cannot be configured in conjunction with the action of creating an outer VLAN tag.

Configuration example

1 Network requirements

Configure a traffic behavior **test** and perform traffic policing with the CAR being 100 kbps.

2 Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Define a traffic behavior and enter traffic behavior view.

```
[Sysname] traffic behavior test
```

# Configure the traffic behavior.

```
[Sysname-behavior-test] car cir 100
```

Defining a Policy

A policy defines the mapping relationship between a class and a traffic behavior (configured with multiple QoS actions).

In a policy, multiple class-to-traffic-behavior mappings are configured, and these mapping are executed according to the order they are configured.

Follow these steps to specify the traffic behavior for a class in the policy:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Define the policy and enter the policy view	**qos policy** *policy-name*	Required
Specify the traffic behavior for the class in the policy	**classifier** *tcl-name* **behavior** *behavior-name*	Required *tcl-name*: Name of a defined class. *behavior-name*: Name of a defined traffic behavior.
Display configuration information of the specified class of the specified policy and the traffic behavior associated with the class	**display qos policy user-defined** [*policy-name* [**classifier** *tcl-name*]]	Optional Available in any view

If an ACL is defined in a traffic classification rule in a QoS policy, the processing methods vary with different interfaces:

- For forwarding by software, if the **if-match** references the **deny** action in ACL rule, exit this **if-match** and continue to search for the subsequent rules;
- For ASIC forwarding, neglect actions in ACL rule, and reference the action defined in traffic behavior. Only the classification domain in ACL is used for packet matching.

Applying a Policy

Configuration procedure

The **qos apply policy** command maps a policy to a specific interface. One policy mapping can be applied on multiple interfaces. Only one policy can be applied on one direction (inbound or outbound) of an interface.

Follow these steps to apply a policy to the interface:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...		Use the command...	Remarks
Enter interface view or port group view	Enter interface view	**interface** *interface-type interface-number*	Use either the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Apply a policy on the interface		**qos apply policy** *policy-name* { **inbound** | **outbound** [**dynamic**] }	Required
Display the policy configuration information and operating condition of the specified interface or all the interfaces		**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**]	Optional Available in any view
Display configuration information of the specified class (or all the classes) of the specified policy (or all the policies) and configuration information of behaviors associated with these classes		**display qos policy user-defined** [*policy-name* [**classifier** *tcl-name*]]	

- *If a QoS policy is applied on the outbound direction of an interface, the QoS policy is not valid on a local packet (The following are the definition and functions of a local packet: some internal packets are the important protocol packets to maintain the normal operation of a device. Therefore, to ensure these packets to be sent successfully, they are defined as local packets, which QoS does not process, thus reducing the risk of discarding these packets because of the configuration of QoS. Commonly used local packets are: link maintenance packet, ISIS, OSPF, RIP, BGP, LDP, RSVP and SSH.)*
- *QoS policies cannot be applied to ports with even port numbers on 3C17526 modules, and QoS policies applied to a port with an odd port number take effect on the incoming packets of both the port with the odd port number and the port with the even port number (the odd port number plus one).*

Configuration example

1 Network requirements

Configure a policy **test** and in the policy specify the traffic behavior for the data belonging to the class **test_class** as **test_behavior**, and then apply the policy to the inbound interface Ethernet1/1/1.

2 Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Define a policy and enter policy view,

```
[Sysname] qos policy test
[Sysname-qospolicy-test]
```

# Specify the traffic behavior for the class.

```
[Sysname-qospolicy-test] classifier test_class behavior test_behavior
[Sysname-qospolicy-test] quit
```

# Enter Ethernet interface view.

```
[Sysname] interface ethernet 1/1/1
```

# Apply the policy to the port.

```
[Sysname-Ethernet1/1/1] qos apply policy test inbound
```

Displaying and Maintaining QoS Policy

Follow these steps to display and maintain QoS policy:

To do...	Use the command...	Remarks
Display configuration information of specified class of specified policy and behavior associated with these classes	**display qos policy user-defined** [*policy-name* [**classifier** *tcl-name*]]	Available in any view
Display policy configuration information and operating condition on specified or all interfaces	**display qos policy interface** [*interface-type interface-number*] [**inbound** | **outbound**]	
Display the configured traffic behavior information	**display traffic behavior user-defined** [*behavior-name*]	
Display the configured class information	**display traffic classifier user-defined** [*tcl-name*]	

63 HARDWARE-BASED CONGESTION MANAGEMENT CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

Congestion Management Overview

As to a network device, congestion will occur on the interface where the arrival rate of packets is faster than the sending rate. If there is no enough buffer capacity to store those packets and then a part of them will be lost, which may cause the packet retransmission from the device because of timeout, and lead to a vicious circle.

The key to congestion management is how to define a dispatching policy for resources to decide the forwarding order of packets when congestion occurs. This chapter describes hardware-based congestion management configuration.

SP Queuing

Strict priority (SP) queues include:

- Basic SP queues: a basic SP queue contains multiple queues, with each queue corresponding to a different priority. These queues are scheduled in the descending order of priority.
- Multi-mode SP queues: the queue scheduling modes are extended from that of basic SP queues.

Multi-mode SP queues operate in one of the following three modes:

- SP mode 0: in this mode, multi-mode SP queues are all basic SP queues and are scheduled in the descending order of priority.
- SP mode 1: in this mode, when the remaining external memory space is sufficient, SP queue scheduling algorithm is adopted; when the remaining external memory space is 0, the scheduling algorithm can preferentially forward the packets stored in the internal memory of the chip even if packets with higher priority are waiting to be scheduled in the external memory.
- SP mode 2: in this mode, packets stored in the internal memory of the chip are forwarded preferentially; if no packets are stored in the internal memory of the chip, all the packets are scheduled using the SP queue scheduling algorithm. The disadvantage of SP mode 2 is that the bus bandwidth of the external memory is decreased.

Currently, only SP mode 0 (that is, SP queue scheduling algorithm) is available to Switch 8800s .

WRR Queuing

Weighted round robin (WRR) queues include:

- Basic WRR queues: a basic WRR queue contains multiple queues. You can configure weight, percentage or byte count for each queue and WRR schedules these queues based on the user-defined parameters.
- Group-based WRR queues: all the queues in a group-based WRR queue are scheduled in the mix of WRR queue scheduling algorithm and SP queue scheduling algorithm. You can allocate all an output queue to WRR priority queue group 1, WRR priority queue group 2, or SP queue group as required. Queues are scheduled as follows: each group selects a candidate queue according to its own queue scheduling algorithm, and then the three candidate queues are scheduled using the SP algorithm.
- WRR queues with the maximum delay: the queue scheduling algorithm for WRR queues with the maximum delay is similar to that for the basic WRR queues except that the WRR queues with the maximum delay assure that packets in the queue with the highest priority are transmitted through the queue within the specified maximum delay.

Currently, Switch 8800s support only group-based WRR queues.

Configuring SP Queues

Configuration Procedure

Follow these steps to configure SP queues:

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure SP queues		**qos sp**	Optional
Display SP queuing information		**display qos sp interface** [*interface-type interface-number*]	Optional Available in any view

Configuration Examples

Network requirements

Configure SP queuing on Ethernet 1/1/1.

Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Configure SP queues on Ethernet 1/1/1.

```
[Sysname]interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] qos sp
```

Configuring Group-based WRR Queues

With a queue on a port configured as a group-based WRR queue, the queue scheduling algorithm on the current port is the mix of WRR queue scheduling algorithm and SP queue scheduling algorithm. The queues which are not configured as group-based WRR queues are allocated to the SP queue group.

Configuration Procedure

Group-based WRR queue configuration task list

Follow these steps to configure group-based WRR queues:

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Enable WRR queuing		**qos wrr**	Required
Configure group-based WRR queues		**qos wrr** *queue-id* **group** { { **1** | **2** } **weight** *schedule-value* | **sp** }	Required
Display WRR queuing configuration		**display qos wrr interface** [*interface-type interface-number*]	Optional Available in any view

- *POS interfaces do not support WRR.*
- *In group-based WRR queue configuration, the outgoing ports of 3C17526, 3C17532, and 3C17538 modules support four queues, that is, the queue-id argument is in the range of 0 to 3; this argument is in the range of 0 to 7 for the other modules.*

Configuration Examples

Network requirements

- Configure the queues on a port as WRR queues.
- Allocate queue 1, queue 3, and queue 4 to WRR group 1, with the weight 1, 5, and 10.
- Allocate queue 5 and queue 6 to WRR group 2, with the weight 20 and 10.

Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Configure WRR queues on Ethernet 1/1/1.

```
[Sysname]interface Ethernet 1/1/1
[Sysname-Ethernet1/1/1] qos wrr
[Sysname-Ethernet1/1/1] qos wrr 1 group 1 weight 1
[Sysname-Ethernet1/1/1] qos wrr 3 group 1 weight 5
[Sysname-Ethernet1/1/1] qos wrr 4 group 1 weight 10
[Sysname-Ethernet1/1/1] qos wrr 5 group 2 weight 20
[Sysname-Ethernet1/1/1] qos wrr 6 group 2 weight 10
```

64 PRIORITY MAPPING

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

Priority Mapping Overview

When packets enter the switch, the switch allocate a series of parameters including 802.1p precedence, DSCP, local precedence, and drop precedence to the packets according to the capability of the switch and the corresponding rules.

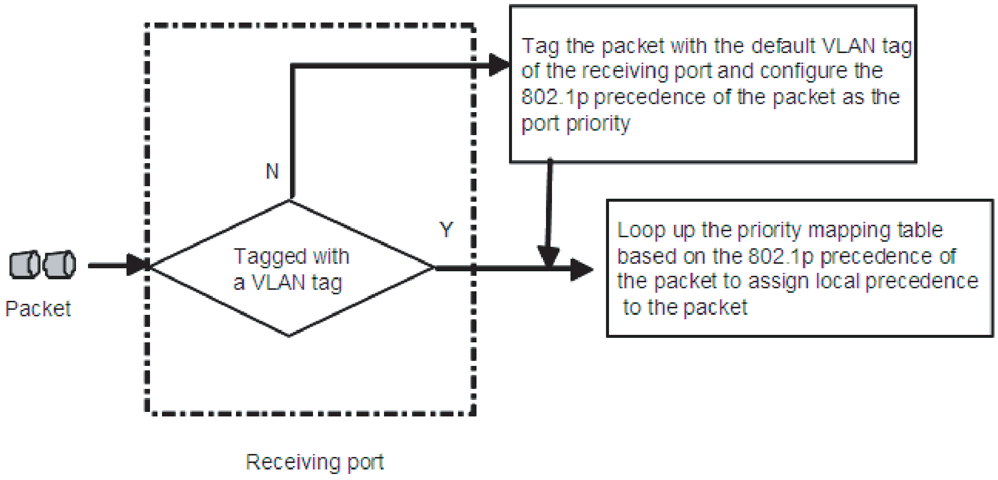
The local precedence and drop precedence are defined as follows:

- Local precedence: locally significant precedence that the switch assigns to the packets. A local precedence corresponds to an output queue ID.
- Drop precedence: parameter referenced for packet drop. The drop precedence for red packets, yellow packets, and green packets is 2, 1, and 0.

Switch 8800s can be configured to trust the packet priority. With packet priority trust mode configured, the switch looks up the priority mapping table based on the priority of a packet to assign priority parameters to the packet.

The packet priority mapping process on a Switch 8800 is shown in Figure 250.

Figure 250 Priority mapping process


Tag the packet with the default VLAN tag of the receiving port and configure the 802.1p precedence of the packet as the port priority
N
Y
Packet
Tagged with a VLAN tag
Loop up the priority mapping table based on the 802.1p precedence of the packet to assign local precedence to the packet
Receiving port

Switch 8800s provide multiple priority mapping tables. All the priority mapping tables and their default values are as follows:

- **dot1p-lp**: 802.1p-precedence-to-local-precedence mapping table.
- **dot1p-dp**: 802.1p-precedence-to-drop-precedence mapping table.
- **dscp-lp**: DSCP-to-local-precedence mapping table.
- **dscp-dp**: DSCP-to-drop-precedence mapping table.
- **dscp-dot1p**: DSCP-to-802.1p-precedence mapping table.
- **dscp-dscp**: DSCP-to-DSCP mapping table.
- **exp-rpr**: EXP-to-RPR-precedence mapping table.
- **dot1p-rpr**: 802.1p-precedence-to-RPR-precedence mapping table.
- **ippre-rpr**: IP-precedence-to-RPR-precedence mapping table.
- **green+dscp-dscp**: DSCP-to-DSCP mapping table for green packets.
- **yellow+dscp-dscp**: DSCP-to-DSCP mapping table for yellow packets.
- **red+dscp-dscp**: DSCP-to-DSCP mapping table for red packets.
- **green+dscp-dp**: DSCP-to-drop-precedence mapping table for green packets.
- **yellow+dscp-dp**: DSCP-to-drop-precedence mapping table for yellow packets.
- **red+dscp-dp**: DSCP-to-drop-precedence mapping table for red packets.
- **green+dscp-lp**: DSCP-to-local-precedence mapping table for green packets.
- **yellow+dscp-lp**: DSCP-to-local-precedence mapping table for yellow packets.
- **red+dscp-lp**: DSCP-to-local-precedence mapping table for red packets.
- **green+dscp-dot1p**: DSCP-to-802.1p-precedence mapping table for green packets.
- **yellow+dscp-dot1p**: DSCP-to-802.1p-precedence mapping table for yellow packets.
- **red+dscp-dot1p**: DSCP-to-802.1p-precedence mapping table for red packets.
- **green+lp-dot1p**: Local-precedence-to-802.1p-precedence mapping table for green packets.
- **yellow+lp-dot1p**: Local-precedence-to-802.1p-precedence mapping table for yellow packets.
- **red+lp-dot1p**: Local-precedence-to-802.1p-precedence mapping table for red packets.

*Use the **display qos map-table** command to view the default value of a priority mapping table.*

Configuring a Priority Mapping Table

The priority mapping tables in the switch can be modified as required.

Follow these steps to configure a priority mapping table:

1. Enter priority mapping table view;
2. Configure mapping table parameters.

Configuration Prerequisites

New priority mapping relationship 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 mapping table view	**qos map-table** { **dot1p-lp** | **dot1p-dp** | **dscp-lp** | **dscp-dp** | **dscp-dot1p** | **dscp-dscp** | **exp-rpr** | **dot1p-rpr** | **ippre-rpr** | **green+dscp-dscp** | **yellow+dscp-dscp** | **red+dscp-dscp** | **green+dscp-dp** | **yellow+dscp-dp** | **red+dscp-dp** | **green+dscp-lp** | **yellow+dscp-lp** | **red+dscp-lp** | **green+dscp-dot1p** | **yellow+dscp-dot1p** | **red+dscp-dot1p** | **green+lp-dot1p** | **yellow+lp-dot1p** | **red+lp-dot1p** }	Required Enter the corresponding priority mapping table view as required.
Configure the mapping relationship	**import** *import-value-list* **export** *export-value*	Required New mapping relationship overwrites the previous one.
Display the configured mapping table	**display qos map-table** [**dot1p-lp** | **dot1p-dp** | **dscp-lp** | **dscp-dp** | **dscp-dot1p** | **dscp-dscp** | **exp-rpr** | **dot1p-rpr** | **ippre-rpr** | **green+dscp-dscp** | **yellow+dscp-dscp** | **red+dscp-dscp** | **green+dscp-dp** | **yellow+dscp-dp** | **red+dscp-dp** | **green+dscp-lp** | **yellow+dscp-lp** | **red+dscp-lp** | **green+dscp-dot1p** | **yellow+dscp-dot1p** | **red+dscp-dot1p** | **green+lp-dot1p** | **yellow+lp-dot1p** | **red+lp-dot1p**]	Optional Available in any view

Configuration Examples

Network requirements

Modify the 802.1p-precedence-to-local-precedence mapping table as follows:

Table 35 The specified 802.1p-precedence-to-local-precedence mapping table

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 802.1p-precedence-to-local-precedence mapping table view.

```
[Sysname] qos map-table dot1p-lp
```

# Modify the 802.1p-precedence-to-local-precedence mapping table 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
```

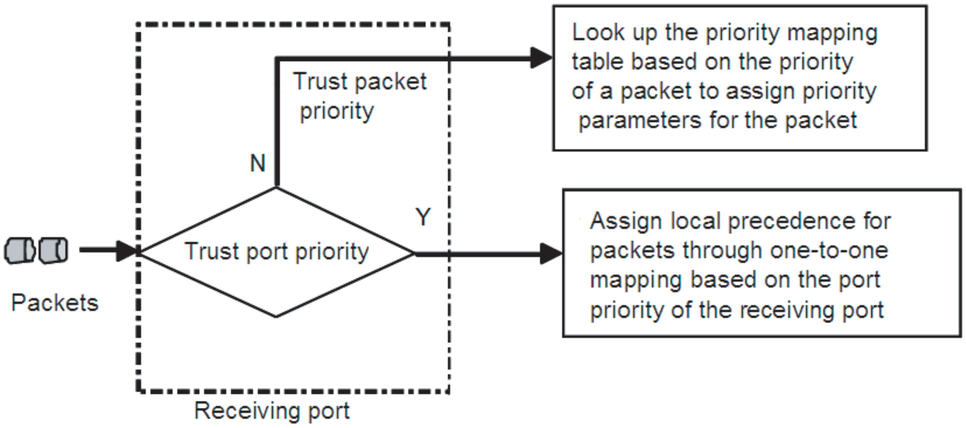
Configuring Port Priority

Switch 8800s provide the following two priority trust modes:

- Trust packet priority: the switch looks up the priority mapping table based on the priority of a packet to assign priority parameters to the packet.
- Trust port priority: the switch assigns local precedence values to all the packets by looking up the priority mapping table based on the pre-defined port priority of the receiving port.

You can configure the priority trust mode for a switch as required. By default, port priority is trusted. The priority mapping process on a Switch 8800 is as shown in Figure 251:

Figure 251 Priority mapping process on a switch supporting the port priority trust mode


Look up the priority mapping table based on the priority of a packet to assign priority parameters for the packet
Trust packet priority
N
Y
Trust port priority
Packets
Assign local precedence for packets through one-to-one mapping based on the port priority of the receiving port
Receiving port

The port priority ranges from 0 to 7. Users can set the port priority as required.

Configuration Prerequisites

The port priority value of the corresponding port is determined.

Configuration Procedure

Follow these steps to configure port priority:

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure port priority		**qos priority** *priority-value*	Optional 0 by default

- *With port priority trust mode configured and the port priority value configured for the port, the switch looks up the 802.1p-precedence-to-local-precedence priority mapping table based on the port priority to assign local precedence to the packets, no matter whether a packet is tagged with a VLAN tag or not.*
- *POS interfaces do not support port priority trust configuration.*

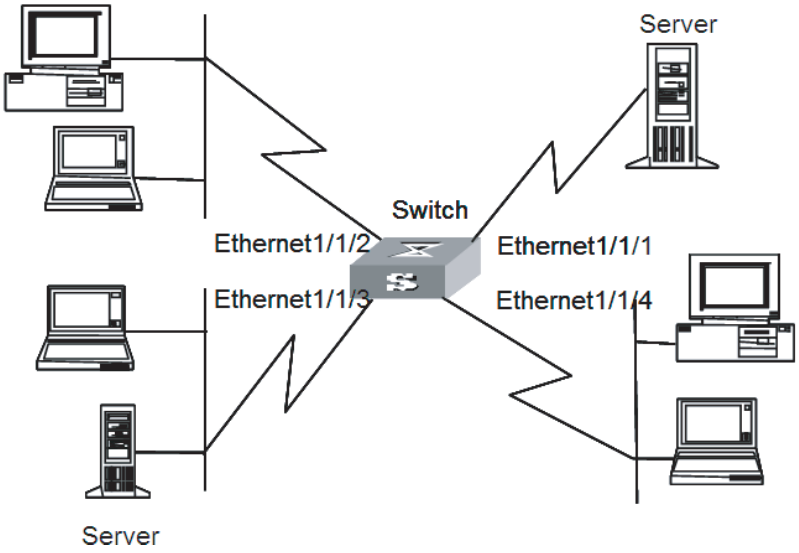
Configuration Examples

Network requirements

- All departments in the enterprise network are interconnected through Switch. The network creates different VLANs for different departments.
- It is required that Switch assign local precedence values to packets through priority mapping on the incoming port.
- The default priority mapping tables of Switch are adopted in priority mapping.

Network diagram

Figure 252 Network diagram for priority trust mode configuration



Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Configure port priority for Ethernet 1/1/1.

```
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] qos priority 1
[Sysname-Ethernet1/1/1] quit
```

# Configure port priority for Ethernet 1/1/2.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] qos priority 3
[Sysname-Ethernet1/1/2] quit
```

# Configure port priority for Ethernet 1/1/3.

```
[Sysname] interface ethernet 1/1/3
[Sysname-Ethernet1/1/3] qos priority 5
[Sysname-Ethernet1/1/3] quit
```

# Configure port priority for Ethernet 1/1/4.

```
[Sysname] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] qos priority 7
[Sysname-Ethernet1/1/4] quit
```

Configuring to Trust Packet Priority

You can configure to trust the 802.1p precedence of a packet. For the packet priority mapping process, refer to “Priority Mapping Overview” on page 851.

Configuration Procedure

Follow these steps to configure to trust packet priority:

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Configure to trust the 802.1p precedence of a packet		**qos trust dot1p**	Optional By default, the 802.1p precedence of a packet is not trusted on a port.
Display the priority trust mode of a port		**display qos trust interface** [*interface-type interface-number*]	Optional Available in any view

- *With the 802.1p precedence values of packets configured to be trusted, priority mapping based on 802.1p precedence is only available to packets with VLAN tags.*
- *POS interfaces do not support packet priority trust configuration.*

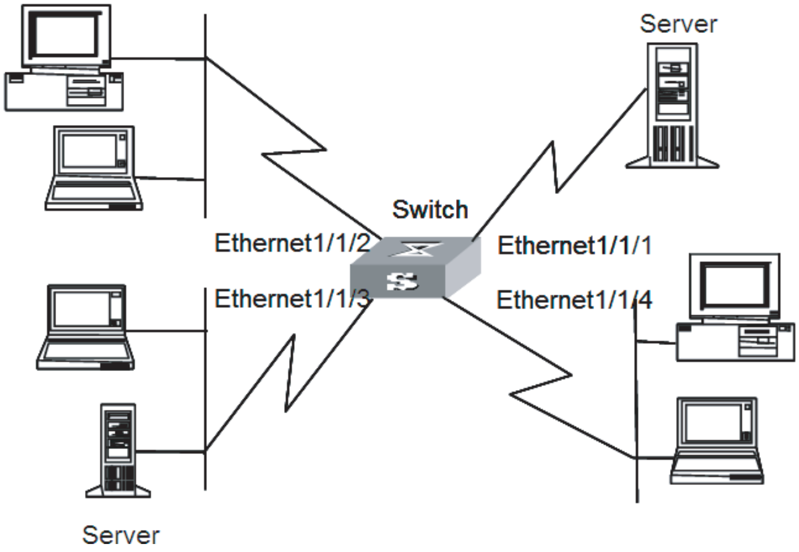
Configuration Examples

Network requirement

- All departments in the enterprise network are interconnected through Switch. The network creates different VLANs for different departments.
- It is required that Switch assign local precedence values to packets through priority mapping on the incoming port.
- The default priority mapping tables of Switch are adopted in priority mapping.

Network diagram

Figure 253 Network diagram for priority trust mode configuration



Configuration procedure

# Enter system view

```
<Sysname> system-view
```

# Enter 802.1p-precedence-to-local-precedence mapping table view to modify the mapping table parameters.

```
[Sysname] qos map-table dot1p-lp
[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
[Sysname-maptbl-dot1p-lp] quit
```

# Configure to trust 802.1p precedence on Ethernet 1/1/1.

```
[Sysname] interface ethernet 1/1/1
[Sysname-Ethernet1/1/1] qos trust dot1p
[Sysname-Ethernet1/1/1] quit
```

# Configure to trust 802.1p precedence on Ethernet 1/1/2.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] qos trust dot1p
[Sysname-Ethernet1/1/2] quit
```

# Configure to trust 802.1p precedence on Ethernet 1/1/3.

```
[Sysname] interface ethernet 1/1/3
[Sysname-Ethernet1/1/3] qos trust dot1p
[Sysname-Ethernet1/1/3] quit
```

# Configure to trust 802.1p precedence on Ethernet 1/1/4.

```
[Sysname] interface ethernet 1/1/4
[Sysname-Ethernet1/1/4] qos trust dot1p
```

65 CONGESTION AVOIDANCE

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

Congestion Avoidance Overview

Excessive congestion can endanger network resources greatly, so some congestion avoidance measures must be taken. Congestion avoidance refers to a traffic control mechanism that can monitor the occupancy status of network resources (such as the queues or buffer). As congestion becomes worse, the system actively drops packets and tries to avoid the network overload through adjusting the network traffics.

Comparing with the end-to-end traffic control, this traffic control herein is of broader significance, which affects more loads of application streams through a device. Of course, while dropping packets, the device may cooperate with traffic control actions (such as TCP traffic control) on the source end to adjust the network's traffic to a reasonable load level. A good combination of packet-drop policies with traffic control mechanisms can maximize the throughput and utilization of network and minimize the packet drop and delay.

Traditional packet-drop policy

The traditional packet-drop policy is tail-drop, that is, when the amount of packets in a queue reaches the maximum value, all newly arrived packets are dropped.

This drop policy leads to global TCP synchronization, that is, when queues drop packets of several TCP links at the same time, these TCP links enter congestion avoidance and slow start status to adjust traffics simultaneously, and then reach traffic peak simultaneously. In this way, network traffic keeps in frequent rises and decreases.

RED and WRED

To avoid global TCP synchronization, random early detection (RED) or weighted random early detection (WRED) can be used.

In RED algorithm, a maximum threshold and a minimum threshold are set for each queue. The packets in the queue are processed as follows:

- When the queue length is smaller than the minimum threshold, no packet is dropped.

- When the queue length exceeds the maximum threshold, all the incoming packets are dropped.
- When the queue length is between the maximum threshold and the minimum threshold, the packets are dropped randomly. The longer the queue is, the higher the drop probability is, but a maximum drop probability exists.

Unlike RED, the random numbers of WRED is generated based on priority. It uses IP precedence to determine the drop policy, and thus the drop probability of packets with high priority is relatively low.

RED and WRED employ the random packet drop policy to avoid global TCP synchronization. When packets of a TCP link are dropped and sent at a low rate, the other TCP links still send packets at high rates. There are always some TCP links sending packets at high rates, thus improving link bandwidth utilization.

Average queue length

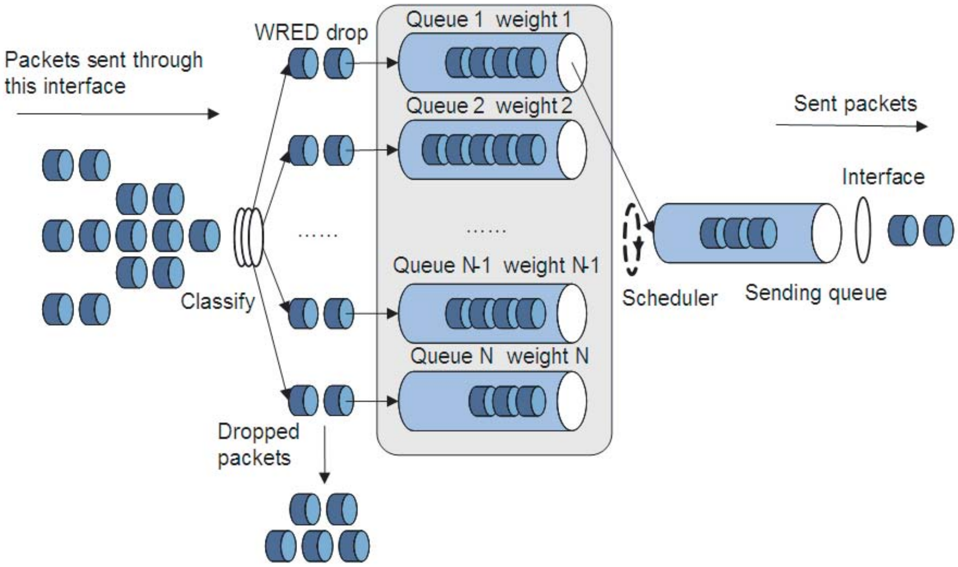
Dropping packets through comparing the queue length with the maximum/minimum threshold treat the burst traffic unfairly and affect traffic transmission. WRED compares the average queue length with the maximum/minimum threshold to determine the drop probability.

The average queue length reflects the queue change tendency and is insensitive to bursting change of the queue length, thus preventing the unfair treatment for the bursting traffic. The average queue length is calculated using the following formula: average queue length = (Previous average queue length × (1-2^{-n})) + (Current queue length ×2^{-n}), where n can be configured through the **qos wred weighting-constant** command.

Relationship between WRED and queuing mechanisms

The relationship between WRED and queuing mechanisms is shown as in the following figure:

Figure 254 Relationship between WRED and queuing mechanisms



Through associating WRED with WFQ, the flow-based WRED can be realized. Because different flow has its own queue during packet classification, the flow with small traffic always has a small queue length, so the packet drop probability is low. The flow with big traffic has bigger queue length, so more packets are dropped. In this way, the benefits of the flow with small traffic are protected.

Configuring WRED

Description on WRED Parameters

Determine the following parameters before configuring WRED:

- The maximum threshold and minimum threshold: when the average queue length is smaller than the minimum threshold, no packet is dropped. When the average queue length is between the maximum threshold and the minimum threshold, the packets are dropped randomly. The longer the queue is, the higher the drop probability is. When the average queue length exceeds the average queue length, all the incoming packets are dropped.
- The exponent used for calculating average queue length: the bigger the exponent is, the more sensitive to real-time queue change the average queue length is.
- The denominator used for calculating the drop probability: this argument functions as the denominator when the drop probability is calculated. The bigger the denominator is, the smaller the calculated drop probability is.

- *3C17532 and 3C17538 modules do not support WRED.*
- *WRED cannot be configured on ports with even port numbers on 3C17526 modules, and WRED configured on a port with an odd port number takes effect on the incoming packets of both the port with the odd port number and the port with the even port number (the odd port number plus one).*

Configuration Procedure

Configuring and applying a queue-based WRED table

Follow these steps to configure and apply a queue-based WRED table:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure a WRED table	**qos wred queue table** *table-name*	Required
Configure the exponent used for calculating average queue length	**queue** *queue-value* **weighting-constant** *exponent*	Optional 9 by default
Configure the other parameters for the WRED table	**queue** *queue-value* [**drop-level** *drop-level*] **low-limit** *low-limit* **high-limit** *high-limit* [**discard-probability** *discard-prob*]	Optional By default, the *low-limit* argument is 10, the *high-limit* argument is 30, and the *discard-prob* argument is 10.

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 the command Configured in Ethernet interface view, the setting is effective on the current interface only; configured in port group view, the setting is effective on all the ports in the port group.
	Enter port group view	**port-group** { **manual** *port-group-name* | **aggregation** *agg-id* }	
Apply the WRED table to the port		**qos wred apply** *table-name*	Required

- *POS interfaces do not support WRED configuration.*
- *In the above table, when the exponent for calculating average queue length and other parameters for the WRED table are configured, the outgoing ports of 3C17526, 3C17532, and 3C17538 modules support four queues, that is, the queue-value argument is in the range of 0 to 3; this argument is in the range of 0 to 7 for the other modules.*

Displaying and Maintaining WRED

Follow these steps to display and maintain WRED:

To do...	Use the command...	Remarks
Display WRED configuration and statistics information on the interface	**display qos wred interface** [*interface-type interface-number*]	Available in any view
Display WRED table configuration status	**display qos wred table** [*table-name*]	

WRED Configuration Examples

Network requirement

Apply a queue-based WRED table to Ethernet 1/1/1.

Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Configure a queue-based WRED table.

```
[Sysname] qos wred queue table queue-table1
[Sysname-wred-table-queue-table1] quit
```

# Enter Ethernet interface view.

```
[Sysname] interface ethernet 1/1/1
```

# Apply the WRED table to Ethernet 1/1/1.

```
[Sysname-Ethernet1/1/1] qos wred apply queue-table1
```

66 AGGREGATION CAR CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "Aggregation CAR Overview" on page 863
- "Referencing Aggregation CAR in Traffic Behaviors" on page 863

Aggregation CAR Overview

Aggregation CAR means to use the same CAR for traffics of multiple traffic behaviors. If aggregation CAR is enabled for multiple traffic behaviors, the total traffics of these traffic behaviors must conform to the TP parameters set in the aggregation CAR.

Referencing Aggregation CAR in Traffic Behaviors

Configuration Prerequisites

- Parameters of the CAR used for aggregation CAR are specified.
- Traffic behaviors where the aggregation CAR is referenced are specified.

Configuration Procedure

Follow these steps to reference an aggregation CAR in traffic behaviors:

Operation	Command	Description
Enter system view	**system-view**	-
Configure parameters for CAR	**qos car** *car-name* **aggregative cir** *committed-information-rate* [**cbs** *committed-burst-size* [**ebs** *excess-burst-size*]] [**pir** *peek-information-rate*] [**red** *action*]	Required By default: CBS is CIR*62.5 EBS is 0 The red packets are dropped
Enter traffic behavior view	**traffic behavior** *behavior-name*	Required
Reference the aggregation CAR in the traffic behavior	**car name** *car-name*	Required

Operation	Command	Description
Display the information about the configured traffic behavior	**display traffic behavior user-defined** [*behavior-name*]	Optional Available in any view
Display the CAR configuration information and statistics information about the specified aggregation CAR	**display qos car name** [*car-name*]	

- *For the description on the default value of CBS, refer to the Switch 8800 Command Reference Guide.*
- *If the ports to which aggregation CAR is applied are not in the same chip, the aggregation CAR is effective for only the ports in the same chip.*

Configuration Examples

Specify the aggregation CAR **aggcar-1** to adopt the following CAR parameters: CIR is 200, CBS is 2,000, and red packets are dropped.

Reference aggregation CAR **aggcar-1** in traffic behavior **be1**.

The configuration procedure is as follows:

```
<Sysname> system-view
[Sysname] qos car aggcar-1 aggregative cir 200 cbs 2000 red discard
[Sysname] traffic behavior be1
[Sysname-behavior-be1] car name aggcar-1
```

67 VLAN POLICY CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "VLAN Policy Overview" on page 865
- "Applying VLAN Policy" on page 866
- "Displaying and Maintaining VLAN Policy" on page 866
- "VLAN Policy Configuration Examples" on page 866

VLAN Policy Overview

QoS policies can be applied in one of the following two modes:

- Interface-based application: a QoS policy is applied to the incoming packets or outgoing packets of an interface.
- VLAN-based application: a QoS policy is applied to all the traffic of a VLAN.

A QoS applied in the interface-based mode is known as an interface policy, and a QoS policy applied in the VLAN-based mode is known as a VLAN policy. With the VLAN policy, you can apply QoS policies to a device and manage these policies conveniently.

On C-type and D-type modules, VLAN policies take effect preferentially. That is, if packets match the VLAN policy, they do not try to match an interface policy; otherwise, they search an interface policy. On B-type modules, interface policies take effect preferentially. With an interface policy applied to an interface, the packets do not search a VLAN policy no matter whether they match the interface policy or not.

VLAN policies are invalid on user authentication interfaces. A user authentication interface joins and exits a VLAN dynamically, and the corresponding VLAN policy is not applied to the interface.

VLAN policies are invalid on dynamic VLANs. VLAN policies cannot be applied to dynamic VLANs. For example, with GARP VLAN registration protocol (GVRP) running, the switch may create a VLAN dynamically, and the corresponding VLAN policy does not take effect on the dynamic VLANs.

As flow templates cannot be configured for a VLAN, the fields of the ACL rules in the QoS policy applied to a VLAN are the fields of the default template.

Applying VLAN Policy

Configuration Prerequisites

- The VLAN poly to be applied is defined. Refer to "Configuring QoS Policy" on page 840 "Configuring QoS Policy" on page 840 for details.
- VLANs where the VLAN policy is to be applied is specified.

Configuration Procedure

Follow these steps to apply the VLAN policy to the specific VLAN:

Operation	Command	Description
Enter system view	**system-view**	-
Apply the VLAN policy to the specific VLAN	**qos vlan-policy** *policy-name* **vlan** *vlan-id-list* { **inbound** | **outbound** }	Required

Displaying and Maintaining VLAN Policy

Follow these steps to display and maintain VLAN policy:

To do...	Use the command...	Remarks
Display the information about a VLAN policy	**display qos vlan-policy** { **name** *policy-name* | **vlan** *vlan-id* } [**slot** *slot-id*]	Available in user view
Clear the statistics information about VLAN policies applied to a VLAN	**reset qos vlan-policy** [**vlan** *vlan-id*]	

VLAN Policy Configuration Examples

Network Requirements

- The VLAN policy **test** is defined to perform TP for packets matching ACL 2000, with the CIR parameter being 8.
- Apply the VLAN policy to the inbound direction of VLAN 200, VLAN 300, VLAN 400, VLAN 500, VLAN 600, VLAN 700, and VLAN 800.

Configuration Procedure

```
<Sysname> system-view
[Sysname] traffic classifier cl1 operator or
[Sysname-classifier-cl1] if-match acl 2000
[Sysname-classifier-cl1] quit
[Sysname] traffic behavior be1
[Sysname-behavior-be1] car cir 8
[Sysname-behavior-be1] quit
[Sysname] qos policy test
[Sysname-qospolicy-test] classifier cl1 behavior be1
[Sysname-qospolicy-test] quit
[Sysname] qos vlan-policy test vlan 200 300 400 500 600 700 800 inbound
```

68 TRAFFIC MIRRORING CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "Traffic Mirroring Overview" on page 867
- "Configuring Traffic Mirroring" on page 867
- "Displaying and Maintaining Traffic Mirroring" on page 868
- "Traffic Mirroring Configuration Examples" on page 868

Traffic Mirroring Overview

Traffic mirroring means to copy packets matching specific traffic classification rules to the specified destination for packet analysis and monitoring.

You can configure to mirror traffic to a port or the CPU.

- Mirroring traffic to a port: copies the matching packets on a port and forwards these packets to the destination port.
- Mirroring traffic to the CPU: copies the matching packets on an interface and forwards these packets to the CPU (the CPU of the module where the traffic mirroring-enabled interface resides).

Configuring Traffic Mirroring

To configure traffic mirroring, you must enter the view of an existing traffic behavior.

- *In a traffic behavior, the action of mirroring traffic to a port is mutually exclusive with the action of mirroring traffic to the CPU.*
- *C-type modules support only the action of mirroring traffic to the CPU.*
- *Traffic cannot be mirrored to the ports of the 3C17526, 3C17532, and 3C17538 modules.*

Mirroring Traffic to a Port

Follow these steps to mirror traffic to a port:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter traffic behavior view	**traffic behavior** *behavior-name*	-
Configure the destination port for traffic mirroring	**mirror-to interface** *interface-type interface-number*	Required

- *If the* ***mirror-to interface*** *command is executed multiple times, the new command overwrites the previous command.*

- *After configuring the action of mirroring traffic to a port in traffic behavior view, configure a policy in policy view to associate the traffic behavior with a traffic class, and then apply the policy to an interface.*

Mirroring Traffic to the CPU

Follow these steps to mirror traffic to the CPU

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter traffic behavior view	**traffic behavior** *behavior-name*	-
Configure the action of mirroring traffic to the traffic for the traffic behavior	**mirror-to cpu**	Required The CPU here refers to the CPU of the module where the interface resides.

After configuring the action of mirroring traffic to the CPU in traffic behavior view, configure a policy in policy view to associate the traffic behavior with a traffic class, and then apply the policy to an interface.

Displaying and Maintaining Traffic Mirroring

Follow these steps to display and maintain 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* [**classifier** *tcl-name*]]	

Traffic Mirroring Configuration Examples

Traffic Mirroring Configuration Examples (To a Port)

Network requirements

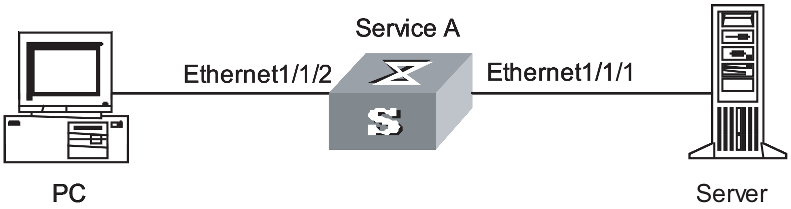
The user’s network is as described below:

- PC is connected to Service A through Ethernet1/1/2.
- Server with two network adapters installed is connected to Ethernet1/1/1 of Switch A.

It is required that Server analyze and monitor all the packets with the source IP address 1.1.1.1 that PC A forwards.

Network diagram

Figure 255 Network diagram for mirroring traffic to a port



Configuration procedure

Configure Service A:

# Enter system view.

```
<Sysname> system-view
```

# Configure ACL 2000 to permit all packets with the source address 1.1.1.1.

```
[Sysname] acl number 2000
[Sysname-acl-basic-2000] rule 1 permit source 1.1.1.1  0
[Sysname-acl-basic-2000] quit
```

# Configure a traffic classifier and use ACL 2000 for traffic classification.

```
[Sysname] traffic classfier 1
[Sysname-classifier-1] if-match acl 2000
[Sysname-classifier-1] quit
```

# Configure a traffic behavior with the action of mirroring traffic to Ethernet 1/1/1.

```
[Sysname] traffic behavior 1
[Sysname-behavior-1] mirror-to interface ethernet 1/1/1
[Sysname-behavior-1] quit
```

# Configure a QoS policy and associate traffic behavior 1 with classifier 1.

```
[Sysname] qos policy 1
[Sysname-qospolicy-1] classifier 1 behavior 1
[Sysname-qospolicy-1] quit
```

# Apply the policy to the inbound direction of Ethernet 1/1/2.

```
[Sysname] interface ethernet 1/1/2
[Sysname-Ethernet1/1/2] qos apply policy 1 inbound
```

After the configuration above, you can analyze and monitor all the packets with the source address 1.1.1.1 that PC A forwards on Server.

69 OUTBOUND TRAFFIC STATISTICS CONFIGURATION

When configuring traffic classification and traffic shaping, go to these sections for information you are interested in:

- "Outbound Traffic Statistics Overview" on page 871
- "Configuring Outbound Traffic Statistics" on page 871
- "Displaying and Maintaining Outbound Traffic Statistics" on page 871

Outbound Traffic Statistics Overview

A Switch 8800 provides two counters for each module to collect statistics on the outbound traffic. You can specify the traffic type for each counter. The traffic type can be all the outbound traffic of the module or the outbound traffic of the combination of the elements (including interface, VLAN, local precedence and drop precedence) on a module.

You can enable the two counters at the same time to collect statistics on the same type of traffic or different types of traffic.

Configuring Outbound Traffic Statistics

Follow these steps to configure outbound traffic statistics:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the outbound traffic statistics function and specify the type of outbound traffic	**qos traffic-counter outbound** { **counter0** | **counter1** } **slot** *slot-num* [**interface** *interface-type interface-number*] [**vlan** *vlan-id*] [**local-precedence** *lp-value*] [**drop-priority** *dp-value*]	By default, the outbound traffic statistics function is disabled.

For the outbound traffic statistics function configured on 3C17526 D-type modules, the monitored object of the counter cannot be an interface.

Displaying and Maintaining Outbound Traffic Statistics

Follow these steps to display and maintain outbound traffic statistics:

To do...	Use the command...	Remarks
Display the outbound traffic statistics information	**display qos traffic-counter outbound** { **counter0** | **counter1** } **slot** *slot-num*	Available in any view

To do...	Use the command...	Remarks
Clear the outbound traffic statistics information collected by a counter	**reset qos traffic-counter outbound** { **counter0** | **counter1** } **slot** *slot-num*	Available in user view

70 AAA, RADIUS AND HWTACACS CONFIGURATION

When configuring AAA, RADIUS and HWTACACS, go to these sections for information you are interested in:

AAA, RADIUS and HWTACACS Configuration Overview

This section covers these topics:

Introduction to AAA

Authentication, authorization, and accounting (AAA) provides a uniform framework for configuring these three security functions to implement the network security management.

The network security mentioned here refers to access control and includes these problems:

- Which users can access the network servers?
- Which services can the authorized users enjoy?
- How to keep accounts for users using the network resources?

Accordingly, AAA provides the following services:

Authentication

AAA supports the following authentication methods:

- None authentication: All users are trusted and no authentication is performed. Generally, this method is not recommended.
- Local authentication: User information (including username, password, and attributes) is configured on the device. Local authentication features high

speed and low cost, but the amount of information that can be stored is limited by the hardware.

- Remote authentication: Both RADIUS and HWTACACS protocols are supported. In this approach, the device acts as the client to communicate with the RADIUS or HWTACACS server. With respect to RADIUS, you can use the standard RADIUS protocol or extended RADIUS protocol to complete authentication in collaboration with systems like CAMS.

Authorization

AAA supports the following authorization methods:

- Direct authorization: All users are trusted and authorized. A user gets the default rights of the system.
- Local authorization: Users are authorized according to the attributes configured for them on the device.
- HWTACACS authorization: Users are authorized using a HWTACACS server.
- RADIUS authorization: RADIUS authorization is a special process in that users are authorized only after they pass authentication. In other words, authorization is bound with authentication. When applying RADIUS scheme, you must specify the same scheme as the authentication scheme and the authorization scheme. It is only in this case that the RADIUS authorization process works. The authentication information is carried in the RADIUS authentication response.

Accounting

AAA supports the following accounting methods:

- None accounting: The system does not keep accounts on the users.
- Local accounting: Local accounting is for controlling the number of local user connections and collecting statistics on number of users.
- Remote accounting: Accounting is implemented by a RADIUS server or HWTACACS server remotely.

AAA usually uses a client/server model, where the client runs on the device that controls user access and the server stores user information. The framework of AAA thus allows for excellent scalability and centralized user information management. Being a management framework, AAA can be implemented through multiple protocols. Currently, AAA is implemented based on RADIUS or HWTACACS.

Introduction to ISP Domain

An Internet service provider (ISP) domain is a group of users that belong to the same ISP. For a username in the *userid@isp-name* format, the *isp-name* following the @ sign is the ISP domain name. 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. Since users of different ISPs may have different user attributes (such as username and password structure, service type, and rights), it is required to configure ISP domains for them and to configure different attribute sets including the AAA policies (such as the RADIUS schemes) for the ISP domains.

Introduction to RADIUS

As described previously, AAA is a management framework and can be implemented through multiple protocols. However, RADIUS is usually used in practice.

What is RADIUS

Remote authentication dial-in user service (RADIUS) is a distributed information interaction protocol in the client/server model. RADIUS can prevent the network from interruption of unauthorized access and is often used in network environments where both high security and remote user access are required. For example, it is often used for managing a large number of geographically dispersed dial-in users that use Modems.

The RADIUS service involves three components:

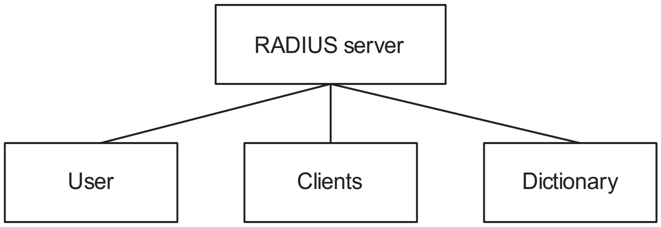
- Protocol: Based on the UDP, RFC 2865 and RFC 2866 define the RADIUS frame format and the message transfer mechanism, and use 1812 as the authentication port and 1813 as the accounting port.
- Server: The RADIUS server runs on the computer or workstation at the center, and maintains information for user authentication and network service access.
- Client: The RADIUS client runs on the NASs located throughout the network.

In the client/server model of RADIUS, the client, a router or a switch, passes user information to the designated RADIUS server and acts on the response of the server (such as connecting/disconnecting users). The RADIUS server receives user connection requests, authenticates users, and returns the required information to the client.

In general, the RADIUS server maintains three databases, namely, Users, Clients, and Dictionary, as shown in Figure 256:

- Users: Stores user information such as the username, password, applied protocols, and IP address.
- Clients: Stores information about RADIUS clients such as the shared key.
- Dictionary: Stores the information for interpreting RADIUS protocol attributes and their values.

Figure 256 Components of the RADIUS server

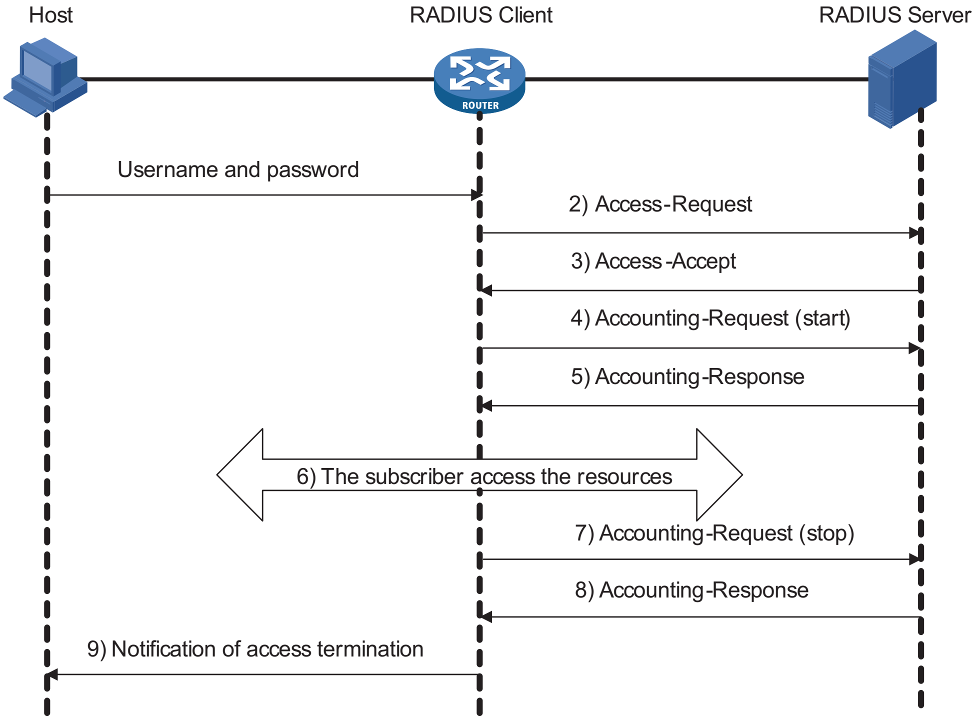


In addition, a RADIUS server can act as the client of another AAA server to provide the proxy authentication or accounting service. A RADIUS server supports multiple user authentication methods, such as PPP-based PAP, CHAP, and UNIX-based login.

Basic message exchange process of RADIUS

In most cases, the user authentication process of a RADIUS server involves a device that can provide the proxy function, such as the NAS. Information exchanged between the RADIUS client and the RADIUS server is authenticated through a shared key for security. The RADIUS protocol combines the authentication and authorization processes by sending authorization information in the authentication response message. See Figure 257.

Figure 257 Basic message exchange process of RADIUS



The following is how RADIUS operates:

1. The user enters the username and password.
2. Having received the username and password, the RADIUS client sends an authentication request (Access-Request) to the RADIUS server.
3. The RADIUS server compares the received user information with that in the Users database. If the authentication succeeds, it sends back an Access-Accept message containing the information of user's right. If the authentication fails, it returns an Access-Reject message.
4. The RADIUS client accepts or denies the user according to the returned authentication result. If it accepts the user, it sends an accounting start request (Accounting-Request) to the RADIUS server, with the value of Status-Type being "start".
5. The RADIUS server returns a start-accounting response (Accounting-Response).
6. The subscriber accesses the network resources.
7. The RADIUS client sends a stop-accounting request (Accounting-Request) to the RADIUS server, with the value of Status-Type being "stop".
8. The RADIUS server returns a stop-accounting response (Accounting-Response).

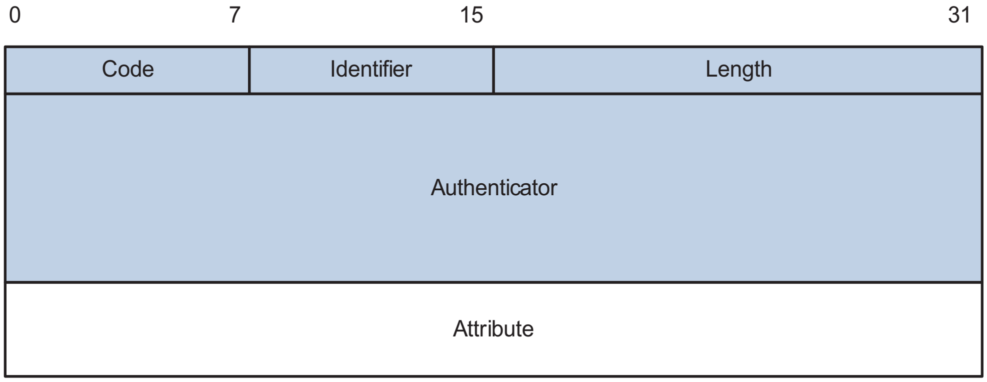
9 The subscriber stops network resource accessing.

RADIUS packet structure

RADIUS resides at the application layer in TCP/IP protocol suite. It defines the way to exchange user information between the device and the ISP RADIUS server.

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 258 shows the RADIUS packet structure.

Figure 258 RADIUS packet structure



Descriptions of fields are as follows:

1 The Code field (1-byte long) is for indicating the type of the RADIUS packet. Table 36 gives the possible values and their meanings.

Table 36 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.
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 accounting on 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.

Table 36 Main values of the Code field

Code	Packet type	Description
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 recorded the accounting information.

2 The Identifier field (1-byte long) is for matching request packets and response packets. It varies with the Attribute field and the received valid response packets, but keeps unchanged during retransmission.

3 The Length field (2-byte long) indicates the length of the entire packet, including the Code, Identifier, Length, Authenticator, and Attribute fields. Bytes beyond the length are considered the padding and are neglected at receipt. If the length of a 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 the reply from the RADIUS server, and is also used in the password hiding algorithm. There are two kinds of authenticators: Request and Response.

5 The Attribute field carries information about the configuration details of a request or response. This field is represented in triplets of Type, Length, and Value.

- Type: One byte, in the range 1 to 255. It is for indicating the type of the attribute. Commonly used attributes for RADIUS authentication and authorization are listed in Table 37.
- 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 37 RADIUS attributes

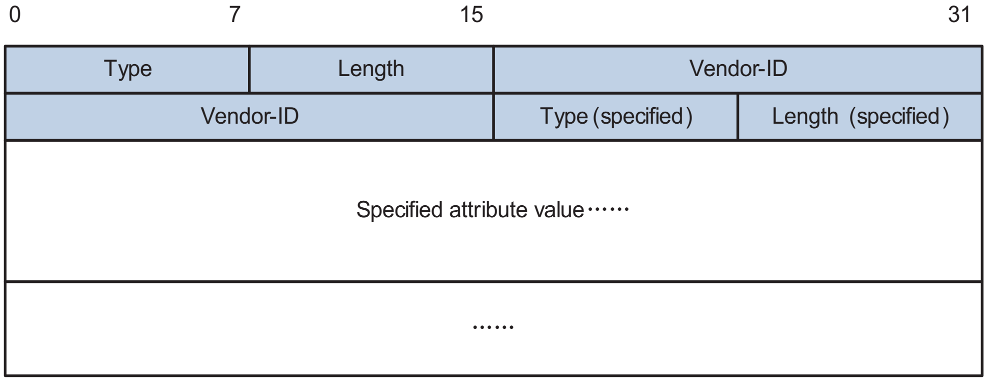
Type	Attribute type	Type	Attribute type
1	User-Name	23	Framed-IPX-Network
2	User-Password	24	State
3	CHAP-Password	25	Class
4	NAS-IP-Address	26	Vendor-Specific
5	NAS-Port	27	Session-Timeout
6	Service-Type	28	Idle-Timeout
7	Framed-Protocol	29	Termination-Action
8	Framed-IP-Address	30	Called-Station-Id
9	Framed-IP-Netmask	31	Calling-Station-Id
10	Framed-Routing	32	NAS-Identifier
11	Filter-ID	33	Proxy-State
12	Framed-MTU	34	Login-LAT-Service
13	Framed-Compression	35	Login-LAT-Node
14	Login-IP-Host	36	Login-LAT-Group
15	Login-Service	37	Framed-AppleTalk-Link
16	Login-TCP-Port	38	Framed-AppleTalk-Network
17	(unassigned)	39	Framed-AppleTalk-Zone

Table 37 RADIUS attributes

Type	Attribute type	Type	Attribute type
18	Reply_Message	40-59	(reserved for accounting)
19	Callback-Number	60	CHAP-Challenge
20	Callback-ID	61	NAS-Port-Type
21	(unassigned)	62	Port-Limit
22	Framed-Route	63	Login-LAT-Port

The RADIUS protocol features excellent extensibility. Attribute 26 (Vender-Specific) allows a vender to define extended attributes to implement functions that the standard RADIUS protocol does not provide. Figure 259 illustrates a segment of a RADIUS packet containing an extended attribute.

Figure 259 Segment of a RADIUS packet containing an extended attribute



Introduction to HWTACACS

What is HWTACACS

3Com terminal access controller access control system (HWTACACS) is an enhanced security protocol based on TACACS (RFC 1492). Similar to RADIUS, it uses the server/client model to implement AAA for the accessing of different types of users, such as point-to-point protocol (PPP), virtual private dial-up network (VPDN), and login users.

Compared with RADIUS, HWTACACS provides more reliable transmission and encryption, and therefore is more suitable for security control. Table 38 lists the primary differences between HWTACACS and RADIUS.

Table 38 Primary differences between HWTACACS and RADIUS

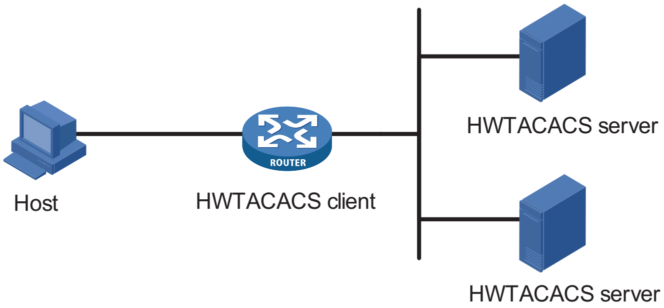
HWTACACS	RADIUS
Uses TCP, providing more reliable network transmission	Uses UDP
Encrypts the entire packet except for the HWTACACS header	Encrypts only the password field in an authentication packet
Separates authentication from authorization. Authentication and authorization can be deployed on different TACACS servers.	Performs authentication and authorization in combination

Table 38 Primary differences between HWTACACS and RADIUS

HWTACACS	RADIUS
Suitable for security control	Suitable for accounting
Supports authorized use of configuration commands	Does not support authorized use of configuration commands

In a typical HWTACACS application, a terminal user needs to log onto the device for operations. Working as the HWTACACS client, the device sends the username and password to the HWTACACS server for authentication. After passing authentication and being authorized, the user can log onto the device to perform operations, as shown in Figure 260.

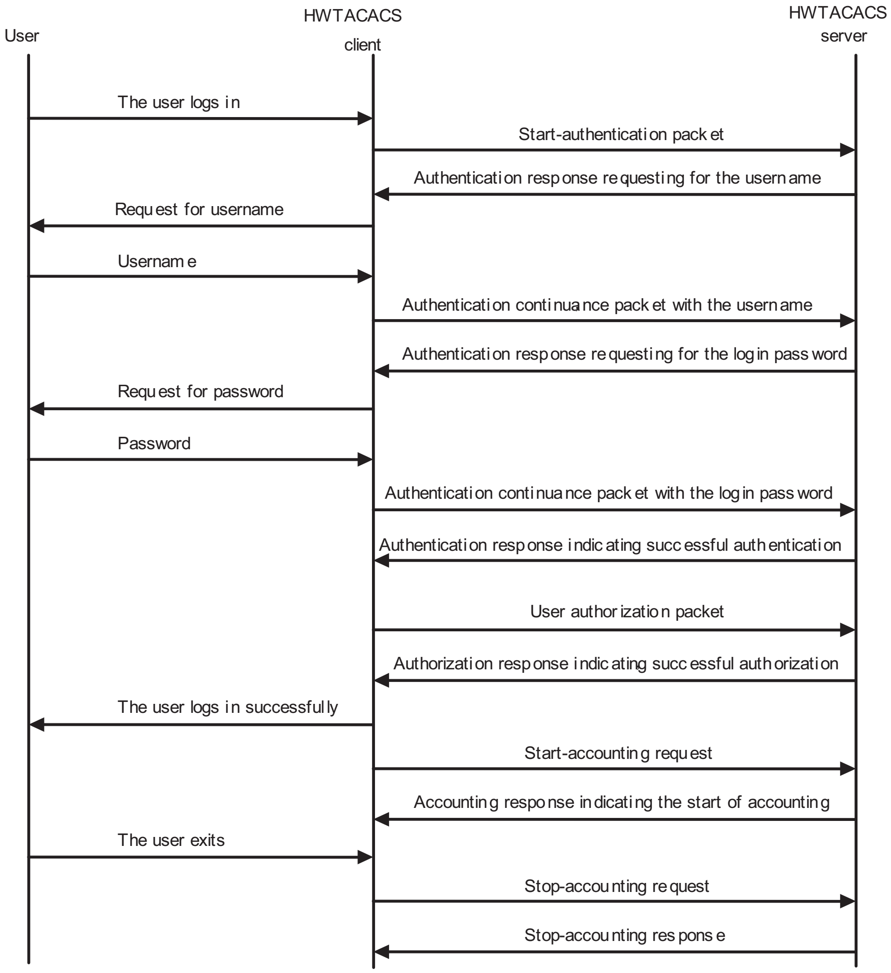
Figure 260 Network diagram for a typical HWTACACS application



Basic message exchange process of HWTACACS

The following takes Telnet user as an example to describe how HWTACACS performs user authentication, authorization, and accounting. Figure 261 illustrates the basic message exchange process of HWTACACS.

Figure 261 Basic message exchange process of HWTACACS for a Telnet user


User
HWTACACS client
HWTACACS server
The user logs in
Start-authentication packet
Authentication response requesting for the username
Request for username
Username
Authentication continuance packet with the username
Authentication response requesting for the login password
Request for password
Password
Authentication continuance packet with the login password
Authentication response indicating successful authentication
User authorization packet
Authorization response indicating successful authorization
The user logs in successfully
Start-accounting request
Accounting response indicating the start of accounting
The user exits
Stop-accounting request
Stop-accounting response

1 A user requests to access the NAS. Upon receiving the request, the HWTACACS client sends a start-authentication packet to the TACACS server.

2 The HWTACACS server sends back an authentication response requesting for the username. Upon receiving the request, the HWTACACS client asks the user for the username.

3 After receiving the username from the user, the HWTACACS client sends to the server an authentication continuance packet carrying the username.

4 The HWTACACS server sends back an authentication response, requesting for the login password. Upon receipt of the response, the HWTACACS client requests the user for the login password.

5 After receiving the login password, the HWTACACS client sends to the HWTACACS server an authentication continuance packet carrying the login password.

6 The HWTACACS server sends back an authentication response indicating that the user has passed authentication.

7 The HWTACACS client sends the user authorization packet to the HWTACACS server.

8 The HWTACACS server sends back the authorization response, indicating that the user is authorized now.

9 Knowing that the user is now authorized, the HWTACACS client pushes the configuration interface of the router or switch to the user.

10 The HWTACACS client sends a start-accounting request to the HWTACACS server.

11 The HWTACACS server sends back an accounting response, indicating that it has received the start-accounting request.

12 When the user logs off, the HWTACACS client sends a stop-accounting request to the HWTACACS server.

13 The HWTACACS server sends back a stop-accounting packet, indicating that the stop-accounting request has been received.

Configuration Task List

AAA configuration task list

Task	Remarks
“Creating an ISP Domain” on page 883	Required
“Configuring ISP Domain Attributes“ on page 884	Optional
“Configuring an AAA Authentication Scheme for an ISP Domain” on page 884	Required For local authentication, refer to “Configuring Local User Attributes” on page 889. For RADIUS authentication, refer to “Configuring RADIUS” on page 891. For HWTACACS authentication, refer to “Configuring HWTACACS” on page 898
“Configuring an AAA Authorization Scheme for an ISP Domain” on page 886	Optional
“Configuring an AAA Accounting Scheme for an ISP Domain” on page 887	Optional
“Configuring Local User Attributes” on page 889	Optional
“Tearing down User Connections Forcibly” on page 891	Optional

RADIUS configuration task list

Task	Remarks
“Creating a RADIUS Scheme” on page 892	Required
“Specifying the RADIUS Authentication/Authorization Servers” on page 892	Required
“Configuring the RADIUS Accounting Servers and Relevant Parameters” on page 893	Optional
“Setting the Shared Key for RADIUS Packets” on page 894	Required

Task	Remarks
“Setting the Maximum Number of RADIUS Request Retransmission Attempts” on page 894	Optional
“Setting the Supported RADIUS Server Type” on page 894	Optional
“Setting the Status of RADIUS Servers” on page 895	Optional
“Configuring Attributes Related to the Data Sent to the RADIUS Server” on page 896	Optional
“Configuring Local RADIUS Server” on page 897	Optional
“Setting Timers Regarding RADIUS Servers” on page 897	Optional

HWTACACS configuration task list

Task	Remarks
“Creating a HWTACACS scheme” on page 898	Required
“Specifying the HWTACACS Authentication Servers” on page 898	Required
“Specifying the HWTACACS Authorization Servers” on page 899	Optional
“Specifying the HWTACACS Accounting Servers” on page 899	Optional
“Setting the Shared Key for HWTACACS Packets” on page 900	Required
“Configuring Attributes Related to the Data Sent to the HWTACACS Server” on page 900	Optional
“Setting Timers Regarding HWTACACS Servers” on page 901	Optional

Configuring AAA

By configuring AAA, you can provide network access service for legal users, protect the networking devices, and avoid unauthorized access and bilking. In addition, you can configure ISP domains to perform AAA on accessing users.

In AAA, users are divided into lan-access users, login users, PPP users, command line users. Except for command line users, you can configure separate authentication/authorization/accounting policies for all the other type 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.

- RADIUS scheme: Reference a configured RADIUS scheme to implement authentication/authorization and accounting. For RADIUS scheme configuration, refer to “Configuring RADIUS” on page 891.
- HWTACACS scheme: Reference a configured HWTACACS scheme to implement authentication/authorization and accounting. For HWTACACS scheme configuration, refer to “Configuring HWTACACS” on page 898.

Creating an ISP Domain

For the NAS, each accessing 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	**domain** *isp-name*	Required
Return to system view	**quit**	-
Specify the default ISP domain	**domain default** { **disable** | **enable** *isp-name* }	Optional The system-default ISP domain named system by default

*You cannot delete the default ISP domain unless you change it to a non-default ISP domain (with the **domain default disable** command) first.*

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 or 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 is in the state of active by default, and users in the domain can request network services.
Specify the maximum number of accessing users in the ISP domain	**access-limit** { **disable** | **enable** *max-user-number* }	Optional No limit by default
Configure the idle cut function	**idle-cut** { **disable** | **enable** *minute* }	Optional Disabled by default
Enable the self-service server localization function and specify the URL of the self-service server for changing user password	**self-service-url** { **disable** | **enable** *url-string* }	Optional Disabled by default
Define an IP address pool for allocating addresses to PPP users	**ip pool** *pool-number low-ip-address* [*high-ip-address*]	Optional No IP address pool is configured by default.

A self-service RADIUS server, for example, CAMS, is required for the self-service server localization function. 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 an AAA Authentication Scheme for an ISP Domain

In AAA, authentication, authorization, and accounting are three 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. You can configure AAA to use only authentication. If you do not perform any authentication configuration, the system-default ISP domain uses the local authentication scheme.

Before configuring an authentication scheme, complete these three tasks:

- For RADIUS or HWTACACS authentication, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication modes do not require any scheme.
- Determine the access mode or service type to be configured. With AAA, you can configure an authentication scheme specifically for each access mode and service type, limiting the authentication protocols that can be used for access.
- Determine whether to configure an authentication scheme for all access modes or service types.

Follow these steps to configure an AAA authentication scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create an ISP domain or enter ISP domain view	**domain** *isp-name*	Required
Specify an authentication scheme 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 scheme for LAN access users	**authentication lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional
Specify the authentication scheme for login users	**authentication login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional
Specify the authentication scheme for PPP users	**authentication ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional

- *The authentication scheme 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 a RADIUS authentication scheme configured, AAA accepts only the authentication result from the RADIUS server. The response from the RADIUS server does include the authorization information when the authentication is successful, 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, the local scheme is the backup scheme when the RADIUS server or HWTACACS server does not make normal response. That is, when the RADIUS server or*

HWTACACS server is available, local authentication is not used. Otherwise, local authentication is used.

- *If the primary authentication scheme is* ***local*** *or* ***none****, the system performs local authentication or does not perform any authentication, rather than uses the RADIUS or HWTACACS scheme.*

Configuring an AAA Authorization Scheme 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 authorized. Authorization is not required. Authorization scheme configuration is optional in AAA configuration.

If you do not perform any authorization configuration, the system-default domain uses the local authorization scheme. With the authorization scheme of **none**, 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 such as users using Telnet or SSH. The default right for FTP users is to use the root directory of the device.

To configure an authorization scheme, follow the steps below:

1 For HWTACACS authorization, configure the HWTACACS scheme to be referenced first. For RADIUS authorization, the RADIUS authorization scheme must be 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 scheme for all access modes or service types.

Follow these steps to configure an AAA authorization scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create an ISP domain or enter ISP domain view	**domain** *isp-name*	Required
Specify the authorization scheme 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 scheme for command line users	**authorization command hwtacacs-scheme** *hwtacacs-scheme-name*	Optional
Specify the authorization scheme for LAN access users	**authorization lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional

To do...	Use the command...	Remarks
Specify the authorization scheme for login users	**authorization login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional
Specify the authorization scheme for PPP users	**authorization ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional

- *The authorization scheme 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*** *keyword and argument combination configured, the local scheme is the backup scheme and is used only when the RADIUS server or HWTACACS server is not available.*
- *If the primary authentication scheme is* ***local*** *or* ***none****, the system performs local authorization or does not perform any authorization, rather than uses the RADIUS or HWTACACS scheme.*
- *Authorization information of the RADIUS server is sent to the RADIUS client along with the authorization response message; therefore, you cannot specify a separate RADIUS 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 an AAA Accounting Scheme 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 scheme configuration is optional. If you do not perform any accounting configuration, the system-default domain uses the local accounting scheme.

To configure an authorization scheme, follow the steps below:

1. For RADIUS or HWTACACS accounting, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication modes do not require any scheme.
2. Determine the access mode or service type to be configured. With AAA, you can configure an accounting scheme 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 scheme for all access modes or service types.

Follow these steps to configure an AAA accounting scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create an ISP domain or enter ISP domain view	**domain** *isp-name*	Required
Enable the accounting optional feature	**accounting optional**	Optional By default, accounting-optional is disabled when an ISP domain is created.
Specify the accounting scheme 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 scheme for LAN access users	**accounting lan-access** { **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional
Specify the accounting scheme for login users	**accounting login** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional
Specify the accounting scheme for PPP users	**accounting ppp** { **hwtacacs-scheme** *hwtacacs-scheme-name* [**local**] | **local** | **none** | **radius-scheme** *radius-scheme-name* [**local**] }	Optional

- *With the **accounting optional** command configured, a user will not be disconnected even if accounting cannot be performed in case no accounting server is available or the communication with the current accounting server fails.*
- *The accounting scheme specified with the **accounting default** command is for all types of users and has a priority lower than that for a specific access mode.*
- *Local accounting only aims to manage the number of local user connections, but provides no statistics function. This management function is only available for local accounting, but is not available for local authorization or local authentication.*
- *With the **radius-scheme*** radius-scheme-name ***local** or **hwtacacs-scheme*** hwtacacs-scheme-name ***local** keyword and argument combination configured, the local scheme is the backup scheme when the RADIUS server or HWTACACS server does not make normal response. That is, when the RADIUS server or HWTACACS server is available, local accounting is not used. Otherwise, local accounting is used.*
- *If the primary accounting scheme is **local** or **none**, the system performs local accounting or does not perform any accounting, rather than uses the RADIUS or HWTACACS scheme.*

- *With the access mode of login, accounting is not supported for FTP services.*

Configuring Local User Attributes

For local authentication, you must create a local user and configure the attributes.

A local user represents a set of users configured on a device, which are 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.

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
Add a local user and enter local user view	**local-user** *user-name*	Required No local user is configured 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.

To do...		Use the command...	Remarks
Specify the service types for the user	Specify the service types for the user	**service-type** { **lan-access** | { **ssh** | **telnet** | **terminal** }* [**level** *level*] }	Required No service is authorized to a user by default
	Authorize the user to use the FTP service	**service-type ftp**	Optional By default, no service is authorized to a user and anonymous access to FTP service is not allowed. If you authorize a user to use the FTP service but do not specify a directory that the user can access, the user can access the root directory of the device by default.
	Set the directory accessible to FTP/SFTP users	**work-directory** *directory-name*	Optional By default, FTP/SFTP users can access the root directory.
	Authorize the user to use the PPP service and configure the callback attribute and caller number	**service-type ppp** [**call-number** *call-number* [**:** *subcall-number*] | **callback-nocheck** | **callback-number** *callback-number*]	Optional By default, no service is authorized to a user and, if the PPP service is authorized, callback without authentication is enabled, no callback number is specified, and the system does not authenticate the caller number of ISDN users.
	Set the callback attributes and calling number attributes for PPP users	**service-type ppp** [**call-number** *call-number* [**:** *subcall-number*] | **callback-nocheck** | **callback-number** *callback-number*]	Optional By default, the system does not authorize users to use any service. By default, no authentication will be performed for callback, no callback number will be set, and no calling number will be authenticated for ISDN users if users are authorized to use the PPP service.'
Set the priority level of the user		**level** *level*	Optional 0 by default
Set attributes for a LAN access user		**attribute** { **access-limit** *max-user-number* | **idle-cut** *minute* | **ip** *ip-address* | **location** { **nas-ip** *ip-address* **port** *slot-number subslot-number port-number* | **port** *slot-number subslot-number port-number* } | **mac** *mac-address* | **vlan** *vlanid* } *	Optional If the specified user is bound to a remote port, you must specify the **nas-ip** (127.0.0.1 by default, indicating the local device) keyword for the user. If the user is bound to a local port, you need not specify the **nas-ip** keyword.

- *With the **local-user password-display-mode cipher-force** command configured, the password is always displayed in cipher text, regardless of the configuration of the **password** command.*

- *Local authentication checks the service types of a local user. If the service types are not available, the user cannot pass authentication. During authorization, a user with no service type configured is authorized with no service by default.*
- *If you specify an authentication method that requires the username and password, including local authentication, RADIUS authentication and HWTACACS authentication, the level of the commands that a user can use after logging in depends on the priority of the user, or the priority of user interface level as with other authentication methods. For an SSH user using RSA public key authentication, the commands that can be used depend on the level configured on the user interface. For details regarding authentication method and command level, refer to "User Interface Configuration" on page 43.*
- *Both the **service-type** and **level** commands can be used to specify user priority. The one used later has the final effect.*
- *The **attribute ip** command only applies to authentications that support IP address passing, such as 802.1x. If you configure the command to authentications that do not support IP address passing, such as MAC address authentication, the local authentication will fail.*
- *The **attribute port** command binds a port by its number only, regardless of the port type.*
- *The **idle-cut** command configured in user view applies to lan-access users only.*
- *In active/standby mode, if the directory specified by the active card does not exist on the standby card, you may fail to log into the system or cannot perform normal operation subsequent to successful login after active/standby switchover occurs.*
- *If the current working directory specified by FTP/SFTP contains the slot number of the standby card, you will fail to log into the system after active/standby switchover occurs. Therefore, it is recommended that the specified working directory should contain no slot number information.*

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 user connections forcibly	**cut connection** { **access-type** { **dot1x** | **mac-authentication** | **portal** } | **all** | **domain** *isp-name* | **interface** *interface-type interface-number* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name* | **vlan** *vlan-id* } [**slot** *slot-number*]	Required Applies to only LAN access user connections

Configuring RADIUS

The RADIUS protocol is configured scheme by scheme. 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 from another point of view, primary servers and secondary servers. In another 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" on page 883.

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 By default, the system has created a RADIUS scheme named "system".

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 By default, the system has created a RADIUS scheme named "system".
Configure the IP address and UDP port of the primary RADIUS authentication/authorization server	**primary authentication** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 1812 for the port.
Configure the IP address and UDP port of the secondary RADIUS authentication/authorization server	**secondary authentication** *ip-address* [*port-number*]	Optional The defaults are as follows: 0.0.0.0 for the IP address, and 1812 for the port.

- *In practice, you may specify two RADIUS servers as the primary and secondary authentication/authorization servers respectively. At a moment, 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.*
- *In the default RADIUS scheme **system**, the IP address and the port number of the primary authentication server are 127.0.0.1 and 1645 respectively.*

Configuring 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 By default, the system has created a RADIUS scheme named "system".
Configure the IP address and UDP port of the primary RADIUS accounting server	**primary accounting** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 1813 for the port.
Configure the IP address and UDP port of the secondary RADIUS accounting server	**secondary accounting** *ip-address* [*port-number*]	Optional The defaults are as follows: 0.0.0.0 for the IP address, and 1813 for the port.
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

- *In practice, you can specify two RADIUS servers as the primary and secondary accounting servers respectively; or specify one server to function as both. Besides, since 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.*
- *In the default RADIUS scheme* ***system****, the IP address and port of the primary accounting server are respectively 127.0.0.1 and 1646.*
- *Currently, neither RADIUS nor HWTACACS supports 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 By default, a RADIUS scheme named "system" has been created in the system.
Set the shared key for RADIUS authentication/authorization or accounting packets	**key** { **accounting** | **authentication** } *string*	Required No key by default

CAUTION: *The shared key configured on the device must be the same as that configured on the RADIUS server.*

Setting the Maximum Number of RADIUS Request Retransmission Attempts

Since 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 the authentication a failure.

Follow these steps to set the maximum number 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 By default, a RADIUS scheme named "system" has been created in the system.
Set the number of retransmission attempts of RADIUS packets	**retry** *retry-times*	Optional 3 by default

- *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 Switch 8800 Command Reference Guide 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**	-

To do...	Use the command...	Remarks
Create a RADIUS scheme and enter RADIUS scheme view	**radius scheme** *radius-scheme-name*	Required By default, a RADIUS scheme named "system" has been created in the system.
Specify the RADIUS server type supported by the device	**server-type** { **extended** | **standard** }	Optional By default, the supported RADIUS server type is **standard**. In the default system scheme, the default RADIUS server type is **extended** .

- *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 server is used, you must configure the RADIUS server to **extended.***

Setting the Status of RADIUS Servers

When a primary server, authentication/authorization server or accounting server, fails, the device automatically turns to the secondary server.

After the status of a primary server stays blocked for a period specified by the **timer quiet** command, the device tries to communicate with the primary server. If the primary server has resumed, the device turns to use the primary server and stops communicating with the secondary server. In this case, the status of the primary server is active again and the status of the secondary server remains the same. After accounting starts, the communication between the client and the secondary server remains unchanged.

If both the primary server and the secondary server are in the state of active or blocked, the device sends the packets only to the primary server.

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 By default, a RADIUS scheme named "system" has been created in the system.

To do...	Use the command...	Remarks
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** }	

Configuring Attributes Related to the Data Sent to the RADIUS Server

Follow these steps to configure the attributes related to the data sent to the RADIUS server:

<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><b>system-view</b></td><td>-</td></tr>
<tr><td colspan="2">Enable the RADIUS trap function</td><td><b>radius trap</b> { <b>accounting-server-down</b> | <b>authentication-server-down</b> }</td><td>Optional
Disabled by default</td></tr>
<tr><td colspan="2">Create a RADIUS scheme and enter RADIUS scheme view</td><td><b>radius scheme</b> <i>radius-scheme-name</i></td><td>Required
By default, a RADIUS scheme named "system" has been created in the system.</td></tr>
<tr><td colspan="2">Specify the format of the username to be sent to a RADIUS server</td><td><b>user-name-format</b> { <b>with-domain</b> | <b>without-domain</b> }</td><td>Optional
By default, the ISP domain name is included in the username.</td></tr>
<tr><td colspan="2">Specify the unit for data flows or packets to be sent to a RADIUS server</td><td><b>data-flow-format data</b> { { <b>byte</b> | <b>giga-byte</b> | <b>kilo-byte</b> | <b>mega-byte</b> } | <b>packet</b> { <b>giga-packet</b> | <b>kilo-packet</b> | <b>mega-packet</b> | <b>one-packet</b> } }*</td><td>Optional
The defaults are as follows:
<b>byte</b> for data flows, and <b>one-packet</b> for data packets.</td></tr>
<tr><td rowspan="3">Set the source IP address of the device to send RADIUS packets</td><td>In RADIUS scheme view</td><td><b>nas-ip</b> <i>ip-address</i></td><td rowspan="3">Use either command
By default, the outbound port serves as the source IP address to send RADIUS packets</td></tr>
<tr><td rowspan="2">In system view</td><td><b>quit</b></td></tr>
<tr><td><b>radius nas-ip</b> <i>ip-address</i></td></tr>
</table>

- *Some earlier RADIUS servers cannot recognize usernames that contain an ISP domain name, therefore before sending a username including a domain name to such a RADIUS server, the device must remove the domain name. This command is thus provided for you to decide whether to include a domain name in a username to be sent to a RADIUS server.*
- *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.*

- *For the default scheme named "system", the username contains no domain name.*
- *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 overwrites the configuration of the* ***radius nas-ip*** *command.*

Configuring Local RADIUS Server

The device, as a RADIUS client, supports the traditional service: perform user authentication using an authentication/authorization server and accounting server respectively. Furthermore, it provides local simple RADIUS server functions (including authentication, authorization and accounting).You can execute the following commands to configure the parameters of the local RADIUS server.

Follow the steps below to configure the local RADIUS server.

To do...	Use the Command...	Remarks
Enter system view	**system-view**	-
Configure local RADIUS server	**local-server nas-ip** *ip-address* **key** *password*	Required By default, no parameters are configured for the local RADIUS server.

- *When the local RADIUS authentication server function is used, the number of the UDP port for authentication/authorization must be 1645, the number of the UDP port for accounting must be 1646, and the IP address of the server is that of the local server.*
- *The shared key configured using this command must be consistent with that for authentication/authorization or accounting packets configured using the* ***key*** *{* ***accounting*** *|* ***authentication*** *} command in RADIUS scheme view.*
- *The device supports a maximum of 16 local RADIUS servers including the default local RADIUS authentication server.*

Setting Timers Regarding RADIUS Servers

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.

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 By default, a RADIUS scheme named "system" has been created in the system.
Set the RADIUS server response timeout timer	**timer response-timeout** *seconds*	Optional 3 seconds by default

To do...	Use the command...	Remarks
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

- *The product of the maximum number of retransmission attempts of RADIUS packets and the RADIUS server response timeout period cannot be greater than 75.*
- *To configure the maximum number of retransmission attempts of RADIUS packets, refer to the command* ***retry*** *in the Switch 8800 Command Reference Guide.*

Configuring HWTACACS

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 No HWTACACS scheme exists by default.

- *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 No HWTACACS scheme exists by default.
Configure the IP address and port of the primary HWTACACS authentication server	**primary authentication** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS authentication server	**secondary authentication** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.

- *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 No HWTACACS scheme exists by default.
Configure the IP address and port of the primary HWTACACS authorization server	**primary authorization** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS authorization server	**secondary authorization** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.

CAUTION:

- *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 No HWTACACS scheme exists by default.
Configure the IP address and port of the primary HWTACACS accounting server	**primary accounting** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS accounting server	**secondary accounting** *ip-address* [*port-number*]	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
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

- *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, neither RADIUS nor HWTACACS supports 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 No HWTACACS scheme exists by default.
Set the shared keys for HWTACACS authentication, authorization, and accounting packets	**key** { **accounting** | **authorization** | **authentication** } *string*	Required No shared key exists by default.

Configuring Attributes Related to the Data Sent to the 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 No HWTACACS scheme exists by default.
Specify the format of the username to be sent to a HWTACACS server		**user-name-format** { **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*	

CAUTION:

- *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 No HWTACACS scheme exists 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

- *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 AAA, RADIUS and HWTACACS

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

To do...	Use the command...	Remarks
Display information about specified or all user connections	**display connection** [**access-type** { **dot1x** | **mac-authentication** | **portal** } | **domain** *isp-name* | **interface** *interface-type interface-number* | **ip** *ip-address* | **mac** *mac-address* | **ucibindex** *ucib-index* | **user-name** *user-name* | **vlan** *vlan-id*] [**slot** *slot-number*]	Available in any view
Display information about specified or all local users	**display local-user** [**domain** *isp-name* | **idle-cut** { **disable** | **enable** } | **service-type** { **ftp** | **lan-access** | **ppp** | **ssh** | **telnet** | **terminal** } | **state** { **active** | **block** } | **user-name** *user-name* | **vlan** *vlan-id*] [**slot** *slot-number*]	

Displaying and Maintaining RADIUS

To do...	Use the command...	Remarks
Display the statistics of the local RADIUS authentication server	**display local-server statistics**	Available in any view
Display the configuration information of a specified RADIUS scheme or all RADIUS schemes	**display radius** [*radius-server-name*] [**slot** *slot-number*]	
Display statistics about RADIUS packets	**display radius statistics** [**slot** *slot-number*]	
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*]	
Clear the statistics of RADIUS	**reset radius statistics** [**slot** *slot-number*]	Available in user view
Delete the buffered stop-accounting packets that are not responded	**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*]	
Clear the statistics of the local server.	**reset local-server statistics**	

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
Display information about buffered stop-accounting requests that get no responses	**display stop-accounting-buffer** { **hwtacacs-scheme** *hwtacacs-scheme-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]	
Clear the statistics of HWTACACS	**reset hwtacacs statistics** { **accounting** | **all** | **authentication** | **authorization** } [**slot** *slot-number*]	Available in user view
Clear the buffered stop-accounting packets that are not responded	**reset stop-accounting-buffer** { **hwtacacs-scheme** *hwtacacs-scheme-name* | **session-id** *session-id* | **time-range** *start-time stop-time* | **user-name** *user-name* } [**slot** *slot-number*]	

AAA, RADIUS and HWTACACS Configuration Examples

AAA for Telnet/SSH Users by a RADIUS Server

- *Configuration of RADIUS authentication, authorization, and accounting for SSH users is similar to that for Telnet users. The following takes Telnet users as an example.*
- *Currently, keeping accounts on FTP users is not supported.*

Network requirements

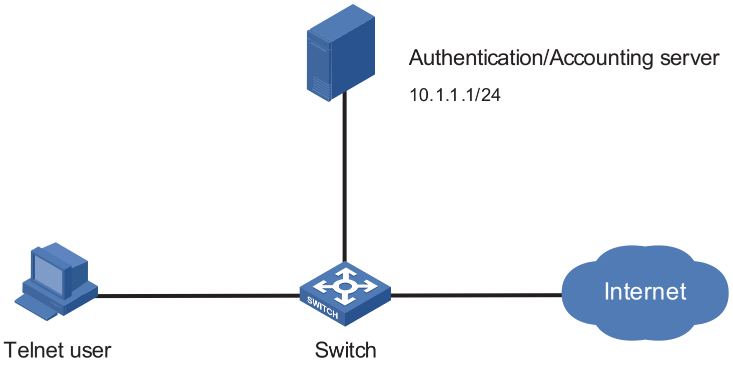
- Configure the switch so that the RADIUS server can perform authentication, authorization and accounting to Telnet users, as shown in Figure 262.
- Connect the RADIUS server of CAMS (functioning as an authentication/accounting RADIUS server) to the switch. The IP address of the server is 10.1.1.1.
- Configure the shared key whereby the switch and authentication RADIUS server exchange packets as "expert", configure the shared key whereby the switch and accounting RADIUS server exchange packets as "expert", and

configure the username sent to the RADIUS server to contain domain name information.

- Configure the shared key whereby to exchange packets with the switch to "expert" on the RADIUS server, set the number of the port for authentication and accounting, and add a Telnet username and login password (the format of the username is "*userid@isp-name*").

Network diagram

Figure 262 Configure AAA for Telnet users by a RADIUS server



Configuration procedure

# Enable the Telnet server on the device.

```
<Sysname> system-view
[Sysname] telnet server enable
```

# Configure the switch to use AAA for authenticating Telnet users.

```
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode scheme
[Sysname-ui-vty0-4] quit
```

# Create ISP domain.

```
[Sysname] domain 1
```

# Configure the accounting to be optional. As a CAMS server does not respond to any accounting packets, this is required for a CAMS server.

```
[Sysname-isp-1] accounting optional
[Sysname-isp-1] quit
```

# Configure the RADIUS scheme.

```
<Sysname> system-view
[Sysname] radius scheme rad
[Sysname-radius-rad] primary authentication 10.1.1.1 1812
[Sysname-radius-rad] primary accounting 10.1.1.1 1813
[Sysname-radius-rad] key authentication expert
[Sysname-radius-rad] key accounting expert
[Sysname-radius-rad] server-type extended
[Sysname-radius-rad] user-name-format with-domain
[Sysname-radius-rad] quit
```

# Apply the AAA schemes to the domain. Here all the three schemes of authentication, authorization, and accounting schemes are configured.

```
<Sysname> system-view
[Sysname] domain 1
[Sysname-isp-1] authentication login radius-scheme rad
[Sysname-isp-1] authorization login radius-scheme rad
[Sysname-isp-1] accounting login radius-scheme rad
[Sysname-isp-1] quit
```

# You can achieve the same purpose by setting default AAA schemes for all types of users.

```
[Sysname] domain 1
[Sysname-isp-1] authentication default radius-scheme rad
[Sysname-isp-1] authorization default radius-scheme rad
[Sysname-isp-1] accounting default radius-scheme rad
```

AAA for FTP/Telnet Users by the Device Itself

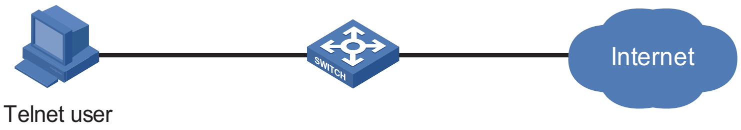
- *Configuration of local authentication and authorization for FTP users is similar to that for Telnet users. The following takes Telnet users as an example.*
- *Currently, keeping accounts on FTP users is not supported.*

Network requirements

As shown in Figure 263, configure the switch to perform local authentication, authorization, and accounting of Telnet users.

Network diagram

Figure 263 Configure local authentication/authorization/accounting for Telnet users



Configuration procedure

1 Solution 1: Use local authentication, authorization, and accounting

# Enable the Telnet server on the device.

```
<Sysname> system-view
[Sysname] telnet server enable
```

# Configure the switch to use AAA for Telnet users.

```
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode scheme
[Sysname-ui-vty0-4] quit
```

# Create local user named telnet.

```
<Sysname> system-view
[Sysname] local-user telnet
[Sysname-luser-telnet] service-type telnet
[Sysname-luser-telnet] password simple aabbccddeeff
[Sysname-luser-telnet] quit
```

# Configure the AAA schemes the ISP domain as local authentication, authorization and accounting.

```
[Sysname] domain system
[Sysname-isp-system] authentication login local
[Sysname-isp-system] authorization login local
[Sysname-isp-system] accounting login local
[Sysname-isp-system] quit
```

# You can achieve the same purpose by setting the default AAA schemes for all types of users.

```
[Sysname] domain system
[Sysname-isp-system] authentication default local
[Sysname-isp-system] authorization default local
[Sysname-isp-system] accounting default local
```

When a user is telneting into the router, the user can use the user name of *userid* @system for local authentication.

2 Solution 2: Use the local RADIUS server

This solution is similar to that given in "AAA for Telnet/SSH Users by a RADIUS Server" on page 903. But you only need to do the following:

- Configuring the local user;
- Configuring the authentication/authorization server, with IP address 127.0.0.1, shared secret key aabbcc, UDP port for authentication/authorization 1645, and UDP port for accounting 1646.
- Configuring the local RADIUS server, with IP address 127.0.0.1, shared secret key aabbcc.

The detailed configuration is as follows:

# Enable the Telnet server on the device.

```
<Sysname> system-view
[Sysname] telnet server enable
```

# Configure the switch to use AAA for Telnet users.

```
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode scheme
[Sysname-ui-vty0-4] quit
```

# Create telnet for the local user.

```
[Sysname] local-user telnet
[Sysname-luser-telnet] service-type telnet
```

```
[Sysname-luser-telnet] password simple aabbccddeeff
[Sysname-luser-telnet] quit
```

# Configure the RADIUS scheme.

```
[Sysname] radius scheme rad
[Sysname-radius-rad] primary authentication 127.0.0.1 1645
[Sysname-radius-rad] primary accounting 127.0.0.1 1646
[Sysname-radius-rad] key authentication aabbcc
[Sysname-radius-rad] key accounting aabbcc
[Sysname-radius-rad] server-type extended
```

# Configure the AAA scheme for the domain.

```
[Sysname] domain 1
[Sysname-isp-1] authentication login radius-scheme rad
[Sysname-isp-1] authorization login radius-scheme rad
[Sysname-isp-1] accounting login radius-scheme rad
[Sysname-isp-cams] quit
```

# Configure the local RADIUS server.

```
[Sysname] local-server nas-ip 127.0.0.1 key aabbcc
```

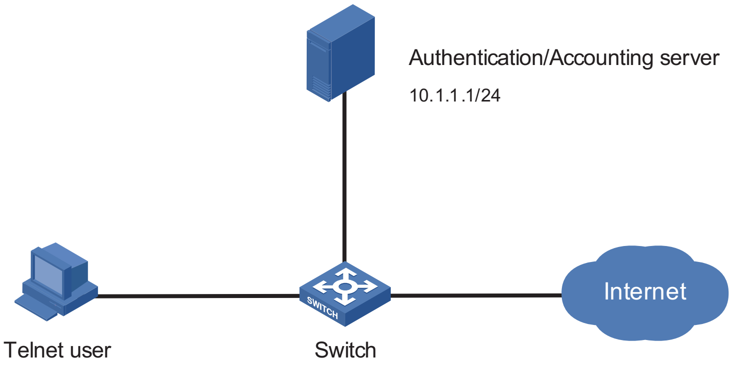
AAA for Telnet Users by a HWTACACS Server

Network requirements

- As shown in Figure 264, configure the switch to use the HWTACACS server to provide authentication, authorization, and accounting services to Telnet users.
- The HWTACACS server is used for authentication, authentication, and accounting, and is connected to the switch. Its IP address is 10.1.1.1.
- On the switch, set the shared keys for authentication, authorization, and accounting packets to **expert**. The username that the switch sends to the HWTACACS server contains no domain name.
- On the HWTACACS server, set the shared key for packets exchanged with the switch to **expert**.

Network diagram

Figure 264 Configure AAA for Telnet users by a HWTACACS Server



Configuration procedure

# Enable the Telnet server function.

```
<Sysname> system-view
[Sysname] telnet server enable
```

# Configure AAA for Telnet users.

```
[Sysname] user-interface vty 0 4
[Sysname-ui-vty0-4] authentication-mode scheme
[Sysname-ui-vty0-4] quit
```

# Configure the HWTACACS scheme.

```
<Sysname> system-view
[Sysname] hwtacacs scheme hwtac
[Sysname-hwtacacs-hwtac] primary authentication 10.1.1.1 49
[Sysname-hwtacacs-hwtac] primary authorization 10.1.1.1 49
[Sysname-hwtacacs-hwtac] primary accounting 10.1.1.1 49
[Sysname-hwtacacs-hwtac] key authentication expert
[Sysname-hwtacacs-hwtac] key authorization expert
[Sysname-hwtacacs-hwtac] key accounting expert
[Sysname-hwtacacs-hwtac] user-name-format without-domain
[Sysname-hwtacacs-hwtac] quit
```

# Apply the AAA schemes to the domain.

```
<Sysname> system-view
[Sysname] domain 1
[Sysname-isp-1] authentication login hwtacacs-scheme hwtac
[Sysname-isp-1] authorization login hwtacacs-scheme hwtac
[Sysname-isp-1] accounting login hwtacacs-scheme hwtac
[Sysname-isp-1] quit
```

# Configure the default AAA schemes for all types of users.

```
[Sysname] domain 1
[Sysname-isp-1] authentication default hwtacacs-scheme hwtac
[Sysname-isp-1] authorization default hwtacacs-scheme hwtac
[Sysname-isp-1] accounting default hwtacacs-scheme hwtac
```

Troubleshooting AAA, RADIUS, and HWTACACS

Troubleshooting RADIUS

Symptom 1: User authentication/authorization always fails.

Analysis:

- The username is not in the format of "*userid@isp-name*", or no default ISP domain is specified for the device.
- This user is not available in the database of the RADIUS server.
- The user does not enter a correct password.
- The shared key on the RADIUS server is different from that on the device.
- The device cannot communicate with the RADIUS server (you can check the communication by pinging the RADIUS server on the device).

Symptom 2: RADIUS packets cannot reach the RADIUS server.

Analysis:

- The device fails to communicate with the RADIUS server (on physical layer or link layer).
- No IP address is assigned to the RADIUS server on the device.
- The UDP ports for authentication/authorization and accounting are not configured correctly.

Symptom 3: A user is authenticated and authorized, but accounting for the user is not normal.

Analysis:

- The accounting port is not correctly configured.
- The accounting server and authentication/authorization server are not the same equipment. However, the device requires that authentication/authorization and accounting should be performed on the same server (they should have the same IP address).

Troubleshooting HWTACACS

Refer to "Troubleshooting RADIUS" on page 908 if you encounter a HWTACACS fault.

71 802.1X CONFIGURATION

When configuring 802.1x, go to these sections for information you are interested in:

- "802.1x Overview" on page 911
- "Configuring 802.1x" on page 920
- "Configuring a Guest VLAN" on page 922
- "Displaying and Maintaining 802.1x" on page 923
- "802.1x Configuration Example" on page 923
- "Guest VLAN Configuration Example" on page 926

802.1x Overview

The 802.1x protocol was proposed by IEEE802 LAN/WAN committee for security problems on wireless LANs (WLAN). Currently, it is widely used on Ethernet as a common port access control mechanism.

As a port-based network access control protocol, 802.1x authenticates and controls accessing devices at the level of port. 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. A device failing the authentication is logically disconnected.

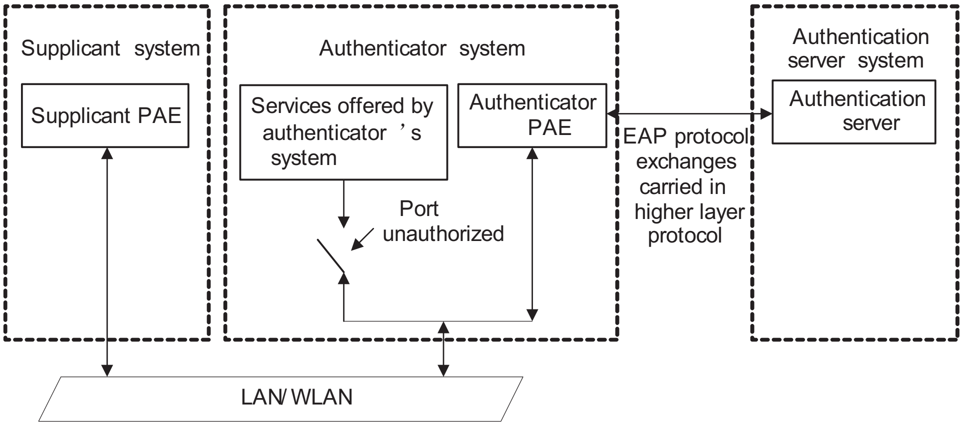
To get more information about 802.1x, go to these topics:

- "Architecture of 802.1x" on page 911
- "Operation of 802.1x" on page 913
- "EAP Encapsulation over LANs" on page 913
- "EAP Encapsulation over RADIUS" on page 915
- "Authentication Process of 802.1x" on page 915
- "802.1x Timers" on page 918
- "Implementation of 802.1x in the Devices" on page 919
- "Features Working Together with 802.1x" on page 919

Architecture of 802.1x

802.1x operates in the typical client/server model and defines three entities: supplicant system, authenticator system, and authentication server system, as shown in Figure 265.

Figure 265 Architecture of 802.1x


Supplicant system
Supplicant PAE
Authenticator system
Services offered by authenticator ' s system
Authenticator PAE
Port unauthorized
EAP protocol exchanges carried in higher layer protocol
Authentication server system
Authentication server
LAN/ WLAN

- Supplicant system: A system at one end of the LAN segment, which is authenticated by the system at the other end. A supplicant system is usually a user-end device and initiates 802.1x authentication through 802.1x client software supporting the EAP over LANs (EAPOL) protocol.
- Authenticator system: A system at one end of the LAN segment, which authenticates the system at the other end. An authenticator system is usually an 802.1x-enabled network device and provides ports (physical or logical) for supplicants to access the LAN.
- Authentication server system: The system providing authentication, authorization, and accounting services for the authenticator system.

The above systems involve three basic concepts: PAE, Controlled port, Control direction.

PAE

Port access entity (PAE) refers to the entity on a given port of a device that performs the 802.1x algorithm and protocol operations. The authenticator PAE uses the authentication server to authenticate a supplicant trying to access the LAN and controls the status of the controlled port according to the authentication result, putting the controlled port in the state of authorized or unauthorized. The supplicant PAE responds to the authentication request of the authenticator PAE and provides authentication information. The supplicant PAE can also send authentication requests and logoff requests to the authenticator.

Controlled port

An authenticator provides ports for supplicants to access the LAN. Each of the ports can be regarded as 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 supplicant can always send and receive authentication frames.
- The controlled port is open to allow normal 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.

Control direction

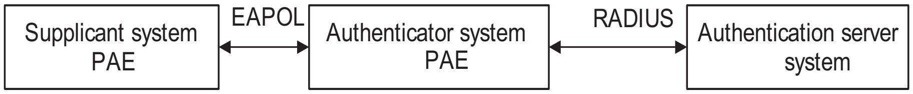
In the unauthorized state, the controlled port can be set to deny traffic to and from the supplicant or just the traffic from the supplicant.

Currently, the Switch 8800 supports only denying the traffic from the supplicant.

Operation of 802.1x

The 802.1x authentication system employs the extensible authentication protocol (EAP) to support authentication information exchange between the supplicant PAE, authenticator PAE, and authentication server.

Figure 266 Operation of 802.1x



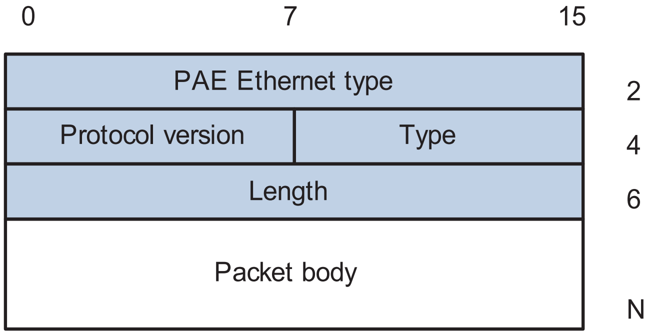
- Between the supplicant PAE and authenticator PAE, EAP protocol packets are encapsulated using EAPOL and transferred over the LAN.
- Between the authenticator PAE and authentication server, EAP protocol packets can be handled in two modes: EAP relay and EAP termination. In EAP relay mode, EAP protocol packets are encapsulated using the EAP attributes of RADIUS (remote authentication dial-in user service) and then relayed to the RADIUS server. In EAP termination mode, EAP protocol packets are terminated at the authenticator PAE, repackaged in the password authentication protocol (PAP) or challenge handshake authentication protocol (CHAP) attributes of RADIUS packets, and then transferred to the RADIUS server.
- The authentication server is usually a RADIUS server. It maintains information about users, such as the username, password, VLAN to which the user belongs, CAR parameters, priority level, and ACL.
- After a user passes the authentication, the authentication server passes information about the user to the authenticator, which controls the status of the controlled port according to the instruction of the authentication server.

EAP Encapsulation over LANs

EAPOL frame format

EAPOL, defined by 802.1x, is intended to carry EAP protocol packets between supplicants and authenticators over LANs. Figure 267 shows the EAPOL frame format.

Figure 267 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 packet. The following types are defined:

- EAP-Packet (a value of 0x00), frame for carrying authentication information.
- EAPOL-Start (a value of 0x01), frame for initiating authentication.
- EAPOL-Logoff (a value of 0x02), frame for logoff request.
- EAPOL-Key (a value of 0x03), frame for carrying key information.
- EAPOL-Encapsulated-ASF-Alert (a value of 0x04), frame for carrying alerting information compliant to Alert Standard Forum (ASF).

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.

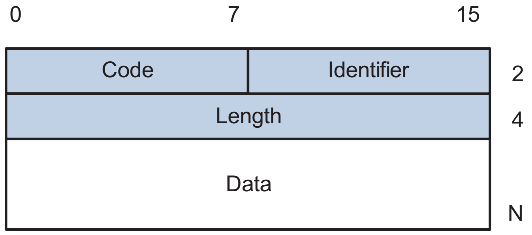
Packet body: The format of this field varies with the value of the Type field.

A frame of the type of EAPOL-Start, EAPOL-Logoff, or EAPOL-Key exists between a supplicant and an authenticator. A frame of the type of EAP-Packet is repackaged and transferred over RADIUS to get through complex networks to reach the authentication server. A frame of the type of EAPOL-Encapsulated-ASF-Alert carries network management-related information (for example, various warning messages) and is terminated at the authenticator.

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 268.

Figure 268 EAP packet format



Code: Type of the EAP packet, which can be Request, Response, Success, or Failure.

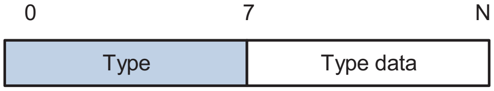
Identifier: Allows matching of responses with requests.

Length: Length of the EAP packet, including the Code, Identifier, Length, and Data fields, in bytes.

Data: This field is zero or more bytes and its format is determined by the Code field.

An EAP packet of the type of Success or Failure has no Data field, and has a length of 4. The Data field in an EAP packet of the type of Request or Response is in the format shown in Figure 269.

Figure 269 Format of the Data field in an EAP request/response packet



Type: EAP authentication type. A value of 1 represents Identity, indicating that the packet is for querying the identity of the supplicant. A value of 4 represents MD5 Challenge, which corresponds closely to the PPP CHAP protocol.

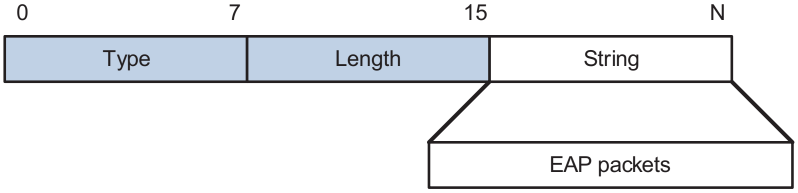
EAP Encapsulation over RADIUS

Two attributes of RADIUS are intended for supporting EAP authentication: EAP-Message and Message-Authenticator. For information about RADIUS packet format, refer to *"Configuring RADIUS" on page 891*.

EAP-Message

The EAP-Message attribute is used to encapsulate EAP packets. Figure 270 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.

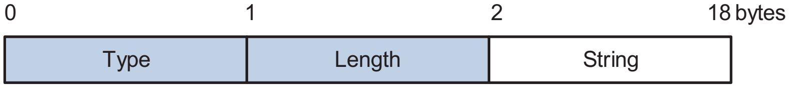
Figure 270 Encapsulation format of the EAP-Message attribute



Message-Authenticator

The Message-Authenticator attribute is used to prevent access requests from being snooped during EAP authentication. It must be included in any packet with the EAP-Message attribute; otherwise, the packet will be considered invalid and get discarded. Figure 271 shows the encapsulation format of the Message-Authenticator attribute. The type field is 80 and the total length is 18 bytes.

Figure 271 Encapsulation format of the Message-Authenticator attribute



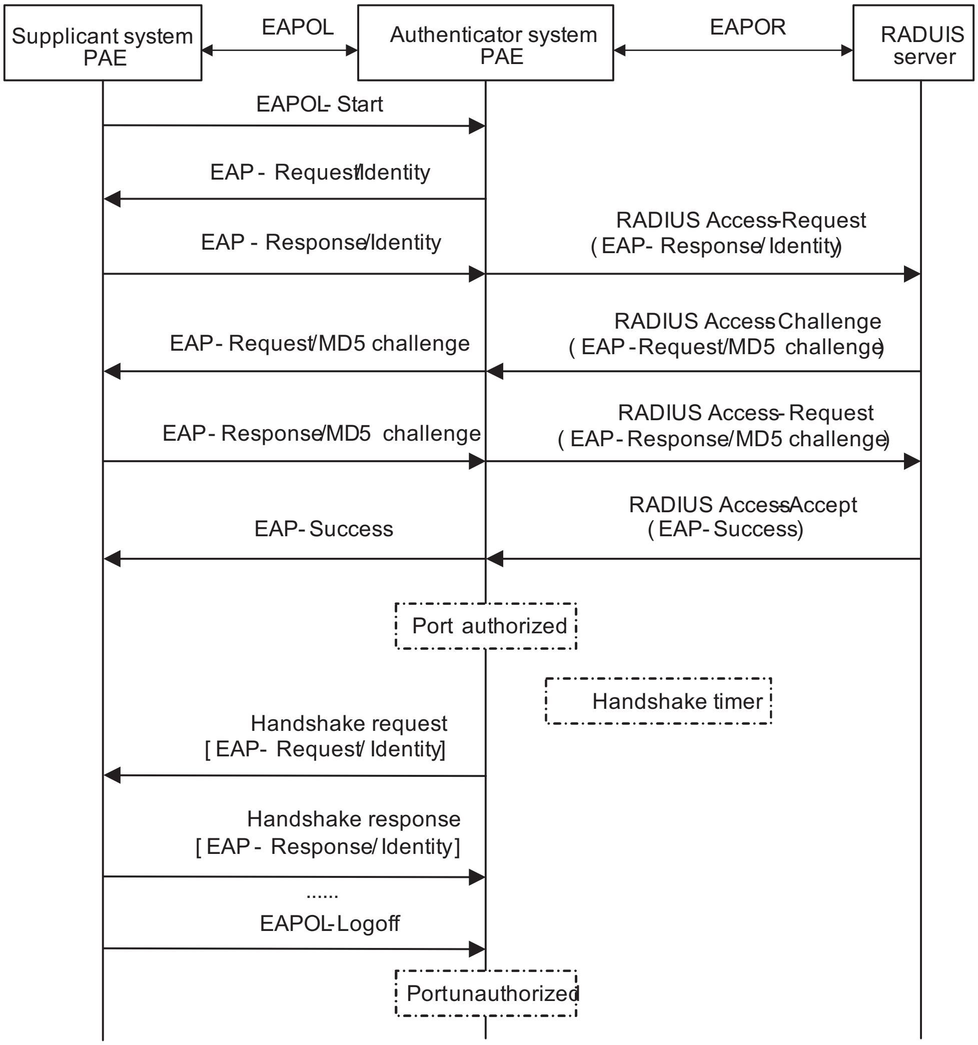
Authentication Process of 802.1x

802.1x authentication can be initiated by either a user or the authenticator system. A user initiates authentication by launching the 802.1x client software to send an EAPOL-Start frame to the authenticator system, while the authenticator system sends an EAP-Request/Identity packet to an unauthenticated user when detecting that the user is trying to login. An 802.1x authenticator system communicates with a remotely located RADIUS server in two modes: EAP relay and EAP termination. The following description takes the first case 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 a high 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. See Figure 272 for the message exchange procedure.

Figure 272 Message exchange in EAP relay mode


Supplicant system PAE
EAPOL
Authenticator system PAE
EAPOR
RADUIS server
EAPOL- Start
EAP - Request/Identity
EAP - Response/Identity
RADIUS Access-Request (EAP- Response/ Identity)
EAP- Request/MD5 challenge
RADIUS Access-Challenge (EAP -Request/MD5 challenge)
EAP- Response/MD5 challenge
RADIUS Access- Request (EAP- Response/MD5 challenge)
EAP- Success
RADIUS Access-Accept (EAP- Success)
Port authorized
Handshake timer
Handshake request [EAP- Request/ Identity]
Handshake response [EAP - Response/ Identity]
......
EAPOL-Logoff
Portunauthorized

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 authenticator to initiate an authentication process.

2 Upon receiving the EAPOL-Start frame, the authenticator responds with an EAP-Request/Identity packet for the username of the supplicant.

3 When the supplicant receives the EAP-Request/Identity packet, it encapsulates the username in an EAP-Response/Identity packet and sends the packet to the authenticator.

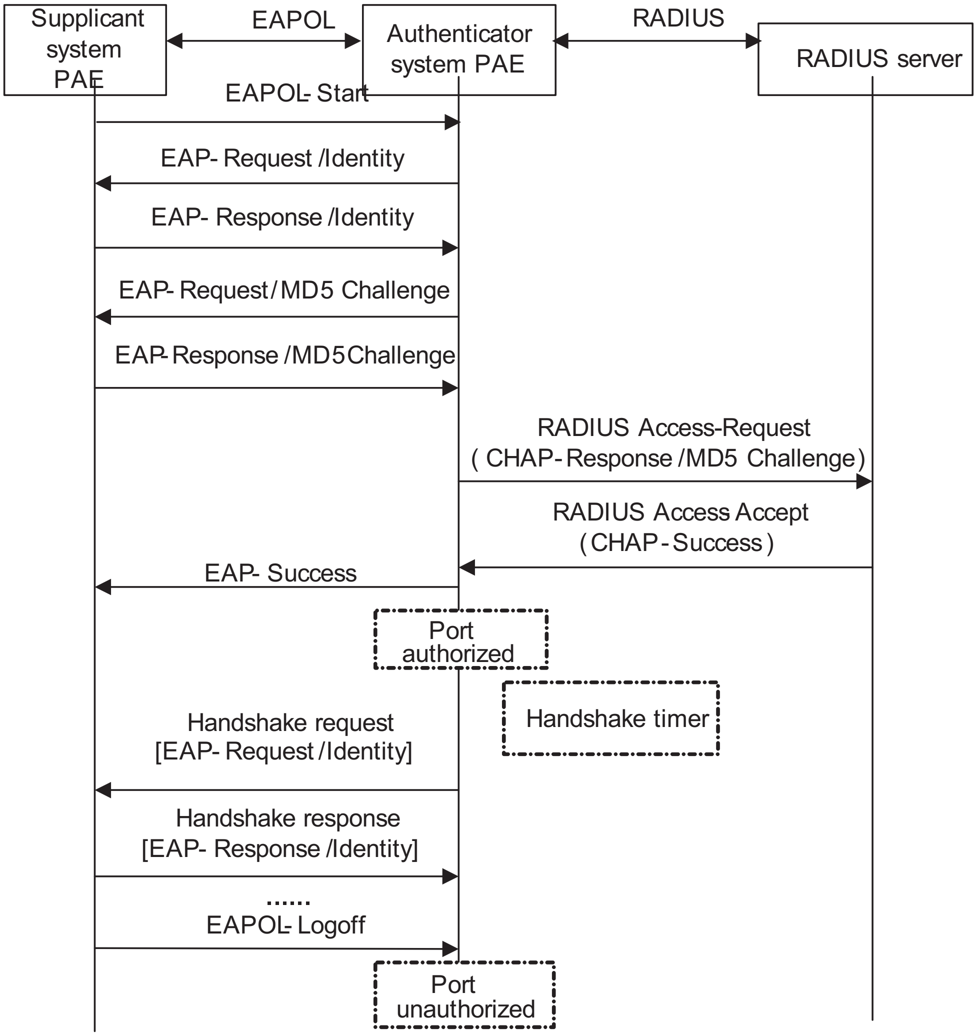
4 Upon receiving the EAP-Response/Identity packet, the authenticator relays the packet in a RADIUS Access-Request packet to the authentication server.

5. When receiving the RADIUS Access-Request packet, the authentication 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 authenticator.
6. After receiving the RADIUS Access-Challenge packet, the authenticator relays the contained EAP-Request/MD5 Challenge packet to the supplicant.
7. When receiving the EAP-Request/MD5 Challenge packet, the supplicant 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 authenticator.
8. After receiving the EAP-Response/MD5 Challenge packet, the authenticator relays the packet in a RADIUS Access-Request packet to the authentication server.
9. When receiving the RADIUS Access-Request packet, the authentication 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 authenticator a RADIUS Access-Accept packet, instructing the authenticator to open the port to permit the access request of the supplicant.
10. After the supplicant gets online, the authenticator periodically sends handshake requests to the supplicant to check whether the supplicant is still online. By default, if two consecutive handshake attempts end up with failure, the authenticator concludes that the supplicant has gone offline and performs the necessary operations, guaranteeing that the authenticator always knows when a supplicant goes offline.
11. The supplicant can also sends an EAPOL-Logoff frame to the authenticator to terminate the authenticated status. In this case, the authenticator changes the status of the port from authorized to unauthorized.

EAP termination

In EAP termination mode, EAP packets are terminated at the authenticator and then repackaged into the PAP or CHAP attributes of RADIUS and transferred to the RADIUS server for authentication, authorization, and accounting. See Figure 273 for the message exchange procedure.

Figure 273 Message exchange in EAP termination mode



Different from the authentication process in EAP relay mode, it is the authenticator that generates the random challenge for encrypting the user password information in EAP termination authentication process. Consequently, the authenticator sends the challenge together with the username and encrypted password information from the supplicant to the authentication server for authentication.

802.1x Timers

Several timers are used in the 802.1x authentication process to guarantee that the supplicants, the authenticators, and the RADIUS server interact with each other in a reasonable manner. The following are the major 802.1x timers:

- Username request timeout timer (tx-period): Once an authenticator sends an EAP-Request/Identity frame to a supplicant, it starts this timer. If this timer expires but it receives no response from the supplicant, it retransmits the request. In addition, to be compatible with clients that do not send EAPOL-Start requests unsolicitedly, the Switch 8800 multicasts EAP-Request/Identity frame periodically to detect the clients, with the multicast interval defined by tx-period.

- Supplicant timeout timer (supp-timeout): Once an authenticator sends an EAP-Request/MD5 Challenge frame to a supplicant, it starts this timer. If this timer expires but it receives no response from the supplicant, it retransmits the request.
- Server timeout timer (server-timeout): Once an authenticator 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 supplicant passes authentication, the authenticator sends to the supplicant handshake requests at this interval to check whether the supplicant is online. If the authenticator receives no response after sending the allowed maximum number of handshake requests, it considers that the supplicant is offline.
- Quiet timer (quiet-period): When a supplicant fails the authentication, the authenticator refuses further authentication requests from the supplicant in this period of time.

Implementation of 802.1x in the Devices

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.

These extensions can help improve network security and manageability dramatically.

After an 802.1x supplicant passes authentication, the authentication server sends authorization information to the authenticator. If the authorization information contains VLAN authorization information, the authenticator adds the port connecting the supplicant 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 supplicant goes offline, the configured one takes effect.

Features Working Together with 802.1x

VLAN Assigning

After an 802.1x user passes the authentication, the server will send an authorization message to the switch. If the authorization message includes the assigned VLAN information, the switch adds the port that the user uses for 802.1x authentication to the assigned VLAN.

The assigned VLAN neither changes nor affects the configuration of a port. However, since the assigned VLAN has higher priority than the user-configured VLAN, it is the assigned VLAN that takes effect after a user passes authentication. After the user goes offline, the port returns to its original VLAN.

- *If the port link type is Access, the authentication server will assign a VLAN successfully.*
- *If the port link type is Hybrid or Trunk, the authentication server will fail to assign a VLAN.*

Guest VLAN

Guest VLAN is the default VLAN that a supplicant can access without authentication. After the supplicant passes 802.1x authentication, s/he can access other network resources. A user of the guest VLAN can perform operations such as downloading and upgrading the authentication client software. If a supplicant does not have the required authentication client software or the version of the client software is lower, the supplicant will fail the authentication and the port the supplicant uses to access the authenticator will be added into the guest VLAN.

If a device with 802.1x enabled and the guest VLAN correctly configured sends an EAP-Request/Identity packet for the allowed maximum number of times but gets no response, it adds the port into the guest VLAN.

When a supplicant added into the guest VLAN initiates another authentication process, if the authentication is not successful, the supplicant stays in the guest VLAN; otherwise, two cases may occur:

- The authentication server assigns a VLAN: The port leaves the guest VLAN and joins the assigned VLAN. If the supplicant goes offline, the port returns to its original VLAN, that is, the VLAN to which it is configured to belong and it belongs before joining the guest VLAN.
- The authentication server does not assign any VLAN: The port leaves the guest VLAN and returns to its original VLAN. If the supplicant goes offline, the port just stays in its original VLAN.

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:

- For remote RADIUS authentication, the username and password information must be configured on the RADIUS server and the RADIUS client-related configurations must be performed on the authenticator.
- For local authentication, the username and password information must be configured on the authenticator and the service type must be set to **lan-access**.

For details about these configuration tasks, refer to *"AAA, RADIUS and HWTACACS Configuration Overview" on page 873*.

Configuration Procedure

Follow these steps to configure 802.1x:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enable 802.1x globally	**dot1x**	Required Disabled by default
Enable 802.1x for one or more ports	**dot1x interface** *interface-list*	Required
	interface *interface-type interface-number* **dot1x**	Disabled for any port by default
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
Enable detection and control of users logging in through proxies globally	**dot1x supp-proxy-check** { **logoff** | **trap** }	Optional Disabled by default
Set the maximum number of users to be supported simultaneously for specified or all ports	**dot1x max-user** *user-number* [**interface** *interface-list*]	Optional 1024 by default
Set the 802.1x authentication method	**dot1x authentication-method** { **chap** | **eap** | **pap** }	Optional CHAP by default
Set the maximum number of attempts to send an authentication request to a supplicant	**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, 30 seconds for the username request timeout timer, 30 seconds for the supplicant timeout timer, and 100 seconds for the server timeout timer.
Enable the quiet-period timer	**dot1x quiet-period**	Optional Disabled by default
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Enable detection and control of users logging in through proxies for the port	**dot1x supp-proxy-check** { **logoff** | **trap** }	Optional Disabled by default
Enable online user handshake	**dot1x handshake**	Optional Enabled by default

Note that:

- 802.1x must be enabled both globally in system view and for the intended ports in system view or Ethernet interface view. Otherwise, it does not function.
- Generally, it is unnecessary to change 802.1x timers unless in some special or extreme network environments.
- The 802.1x proxy detection function must be enabled both globally in system view and for intended ports in system view or Ethernet interface view. Otherwise, it does not function.
- The 802.1x proxy detection function depends on the online user handshake function. Be sure to enable handshake before enabling proxy detection and to disable proxy detection before disabling handshake.
- You can neither add an 802.1x-enabled port into an aggregation group nor enable 802.1x on a port being a member of an aggregation group.
- In EAP relay authentication mode, the authenticator 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 *the Switch 8800 Command Reference Guide.*
- If the username of a supplicant contains the version number or one or more blank spaces, you can neither retrieve information nor disconnect the supplicant by using the username. However, you can use items such as IP address and connection index number to do so.

Configuring a Guest VLAN

Configuration Prerequisites

- Enable 802.1x
- Set the port access control method to **portbased** for the port
- Set the port access control mode to **auto** for the port
- Set the port link type to **access**.
- Create the VLAN to be specified as the guest VLAN

Configuration Procedure

Follow these steps to configure Guest VLAN:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the guest VLAN for specified or all ports	**dot1x guest-vlan** *vlan-id* [**interface** *interface-list*] Or in Ethernet interface view **interface** *interface-type interface-number* **dot1x guest-vlan** *vlan-id*	Required By default, a port is configured with no guest VLAN.

- *A super VLAN cannot be set as the guest VLAN. Similarly, a guest VLAN cannot be set as the super VLAN. For information about super VLAN, refer to "Super VLAN Configuration" on page 167.*
- *The guest VLAN function does not apply to non-access ports.*
- *Configurations in system view are effective to all ports while configurations in interface view are effective to the current port only.*

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

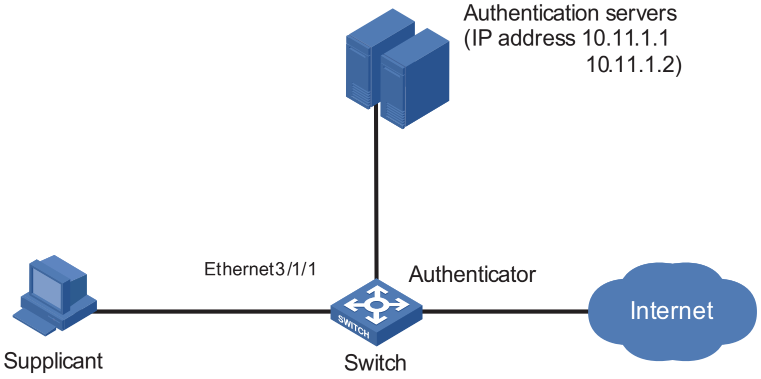
802.1x Configuration Example

Network requirements

- As shown in Figure 274, a host is connected to port Ethernet 3/1/1 on the switch.
- The access control method of **macbased** is required on the port to control supplicants.
- All AAA supplicants 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 authenticator gets users offline.
- A server group with two RADIUS servers is connected to the switch. The IP addresses of the servers are 10.11.1.1 and 10.11.1.2 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 switch to exchange packets with the authentication server as **name**, and that for the switch to exchange packets with the accounting server as **money**.
- Specify the switch 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 switch 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 **localpassword** 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.

Network diagram

Figure 274 Network diagram for 802.1x configuration



Configuration procedure

The following configuration procedure covers most AAA/RADIUS configuration commands for the authenticator, while configuration on the supplicant and RADIUS server are omitted. For information about AAA/RADIUS configuration commands, refer to "AAA, RADIUS and HWTACACS Configuration" on page 873.

# Add local access user **localuser**, enable the idle cut function, and set the idle cut interval.

```
<Sysname> system-view
[Sysname] local-user localuser
[Sysname-luser-localuser] service-type lan-access
[Sysname-luser-localuser] password simple localpassword
[Sysname-luser-localuser] attribute idle-cut 20
[Sysname-luser-localuser] quit
```

# Create RADIUS scheme **radius1** and enter its view.

```
[Sysname] radius scheme radius1
```

# Configure the IP addresses of the primary authentication and accounting RADIUS servers.

```
[Sysname-radius-radius1] primary authentication 10.11.1.1
[Sysname-radius-radius1] primary accounting 10.11.1.2
```

# Configure the IP addresses of the secondary authentication and accounting RADIUS servers.

```
[Sysname-radius-radius1] secondary authentication 10.11.1.2
[Sysname-radius-radius1] secondary accounting 10.11.1.1
```

# Specify the shared key for the switch to exchange packets with the authentication server.

```
[Sysname-radius-radius1] key authentication name
```

# Specify the shared key for the switch to exchange packets with the accounting server.

```
[Sysname-radius-radius1] key accounting money
```

# Set the interval for the switch to retransmit packets to the RADIUS server and the maximum number of transmission attempts.

```
[Sysname-radius-radius1] timer response-timeout 5
[Sysname-radius-radius1] retry 5
```

# Set the interval for the switch to send real time accounting packets to the RADIUS server.

```
[Sysname-radius-radius1] timer realtime-accounting 15
```

# Specify the switch to remove the domain name of any username before passing the username to the RADIUS server.

```
[Sysname-radius-radius1] user-name-format without-domain
[Sysname-radius-radius1] quit
```

# Create default user domain aabbcc.net and enter its view.

```
[Sysname] domain aabbcc.net
[Sysname-isp-aabbcc.net] quit
[Sysname] domain default enable aabbcc.net
[Sysname] domain aabbcc.net
```

# Set radius1 as the RADIUS scheme for users of the domain and specify to use local authentication as the secondary scheme.

```
[Sysname-isp-aabbcc.net] authentication lan-access radius-scheme radius1 local
[Sysname-isp-aabbcc.net] authorization lan-access radius-scheme radius1 local
[Sysname-isp-aabbcc.net] accounting lan-access radius-scheme radius1 local
```

# Set the maximum number of users for the domain as 30.

```
[Sysname-isp-aabbcc.net] access-limit enable 30
```

# Enable the idle cut function and set the idle cut interval.

```
[Sysname-isp-aabbcc.net] idle-cut enable 20
[Sysname-isp-aabbcc.net] quit
```

# Enable 802.1x globally.

```
[Sysname] dot1x
```

# Enable 802.1x for port Ethernet 3/1/1.

```
[Sysname] interface ethernet 3/1/1
[Sysname-Ethernet3/1/1] dot1x
[Sysname-Ethernet3/1/1] quit
```

# Set the port access control method. (Optional. The default answers the requirement.)

```
[Sysname] dot1x port-method macbased interface ethernet 3/1/1
```

Guest VLAN Configuration Example

Network requirements

As shown in Figure 275:

- A host is connected to port Ethernet 1/1/3 of the switch and must pass 802.1x authentication to access the Internet.
- The authentication server run RADIUS and is in VLAN 2.
- The update server, which is in VLAN 10, is for client software download and upgrade.
- Port Ethernet 1/1/8 of the switch, which is in VLAN 5, is for accessing the Internet.

As shown in Figure 276:

- On port Ethernet 1/1/3, enable 802.1x and set VLAN 10 as the guest VLAN.

As shown in Figure 277:

- Authenticated supplicants are assigned to VLAN 5 and permitted to access the Internet.

Network diagrams

Figure 275 Network diagram for guest VLAN configuration

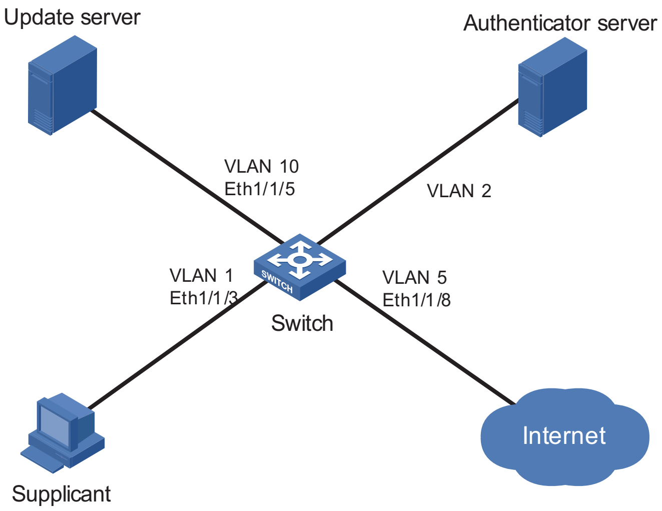


Figure 276 Network diagram with VLAN 10 as the guest VLAN

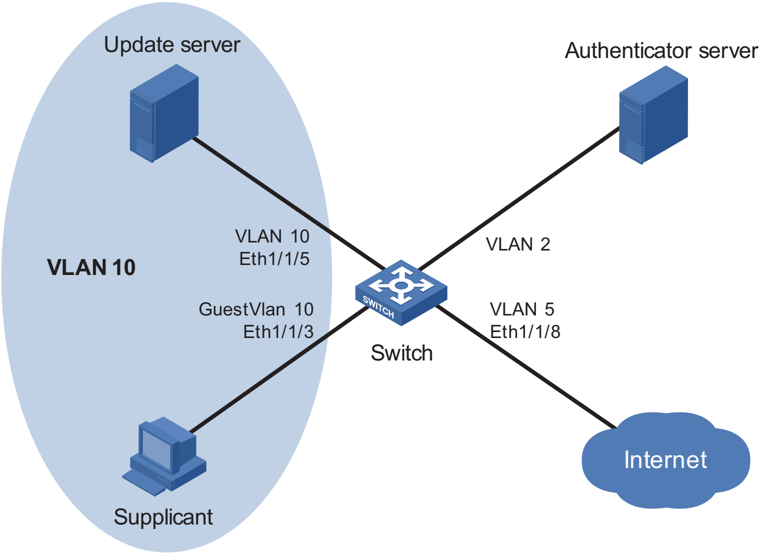
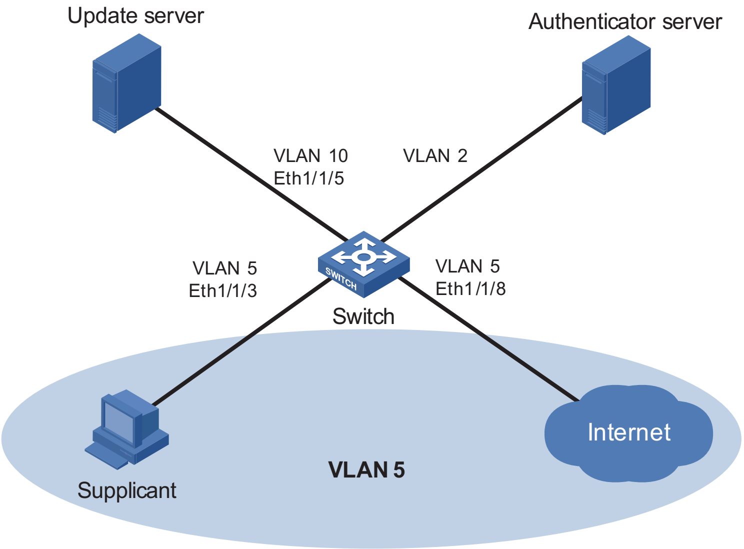


Figure 277 Network diagram when the supplicant passes authentication



Configuration procedure

# Configure RADIUS scheme 2000.

```
<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.11.1.1 1812
[Sysname-radius-2000] primary accounting 10.11.1.1 1813
[Sysname-radius-2000] key authentication nec
[Sysname-radius-2000] key accounting nec
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit
```

# Configure domain system and specify to use RADIUS scheme 2000 for users of the domain.

```
[Sysname] domain system
[Sysname-isp-system] authentication lan-access radius-scheme 2000
[Sysname-isp-system] authorization lan-access radius-scheme 2000
[Sysname-isp-system] accounting lan-access radius-scheme 2000
[Sysname-isp-system] quit
```

# Enable 802.1x globally.

```
[Sysname] dot1x
```

# Enable 802.1x for port Ethernet 1/1/3.

```
[Sysname] interface ethernet 1/1/3
[Sysname-ethernet1/1/3] dot1x
```

# Set the port access control method to **portbased**.

```
[Sysname-ethernet1/1/3] dot1x port-method portbased
```

# Set the port access control mode to **auto**.

```
[Sysname-ethernet1/1/3] dot1x port-control auto
```

# Set the port link type to **access**.

```
[Sysname-ethernet1/1/3] quit/3] port link-type access
[Sysname-ethernet1/1/3] quit
```

# Create VLAN 10.

```
[Sysname] vlan 10
[Sysname-vlan10] quit
```

# Specify port Ethernet 1/1/3 to use VLAN 10 as its guest VLAN.

```
[Sysname] dot1x guest-vlan 10 interface ethernet1/1/3
```

You can use the **display current-configuration** or **display interface ethernet1/1/3** 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.

72 CONFIGURING SSH VERSION 2.0

When configuring SSH2.0, go to these sections for information you are interested in:

SSH2.0 Overview

Secure shell (SSH) offers an approach to securely logging into a remote device. It can protect 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.

An SSH channel can be established through a local connection or WAN, as shown in Figure 278 and Figure 279.

Figure 278 Establish an SSH channel through local connection

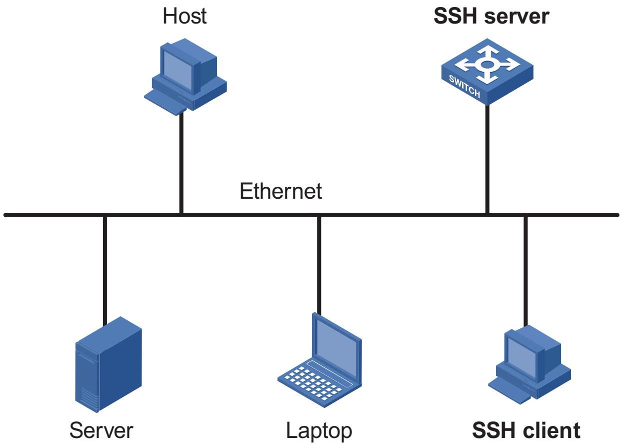
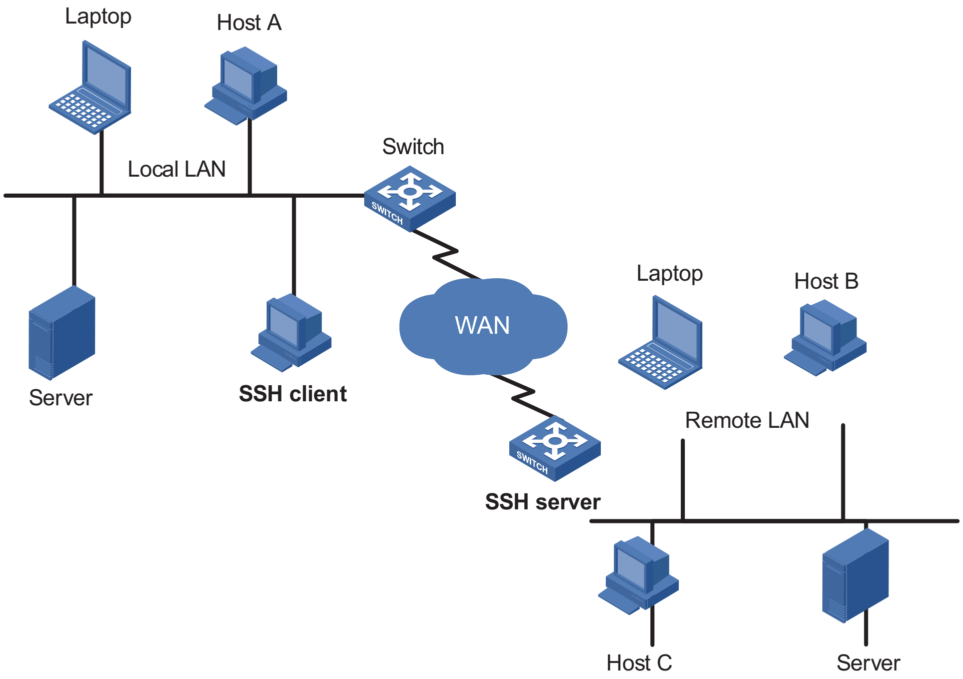


Figure 279 Establish an SSH channel through WAN

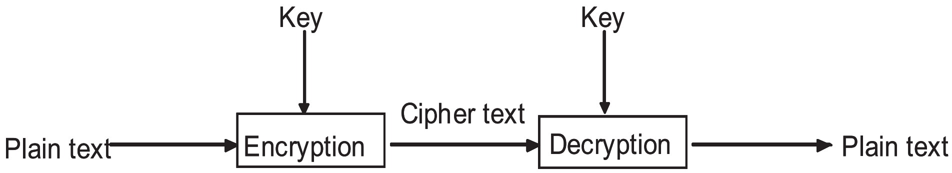


- *Currently, when acting as an SSH server, the device supports two SSH versions: SSH2 and SSH1. When acting as an SSH client, the device supports SSH2 only.*
- *Unless otherwise noted, the "SSH" term in this document refers to SSH2.*

Algorithm and Key

Algorithm is a set of transformation rules for encryption and decryption. Information without being encrypted is known as plain text, while information that is encrypted is known as cipher text. Encryption and decryption are performed using a string of characters called a key, which controls the transformation between plain text and cipher text, for example, changing the plain text into cipher text or cipher text into plain text.

Figure 280 Encryption and decryption



Key-based algorithm is usually classified into symmetric key algorithm and asymmetric key algorithm.

Asymmetric Key Algorithm

Asymmetric key algorithm means that a key pair exists at both ends. The key pair consists of a private key and a public key. The public key is effective for both ends, while the private key is effective only for the local end.

Asymmetric key algorithm encrypts data using the public key and decrypts the data using the private key, thus ensuring data security.

You can also use the asymmetric key algorithm for digital signature. For example, user 1 adds his 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, this means that the data originates from user 1.

Revest Shamir and Adleman (RSA) is an asymmetric key algorithms. RSA can be used for both data encryption and signature.

SSH Operating Process

The session establishment between an SSH client and the SSH server involves the following five stages:

Table 39 Stages in establishing a session between the SSH client and the server

Stages	Description
"Version negotiation" on page 931	SSH1 and SSH2 are supported. The two parties negotiate a version to use.
"Key and algorithm negotiation" on page 931	SSH supports multiple algorithms. The two parties negotiate an algorithm for communication.
"Authentication" on page 932	The SSH server authenticates the client in response to the client's authentication request.
"Session request" on page 933	This client sends a session request to the server.
"Interactive session" on page 933	The client and the server start to communicate with each other.

Version negotiation

- The server opens port 22 to listen to connection requests from clients.
- 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.
- 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.
- 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 to determine whether it can cooperate with the client.
- If the negotiation is successful, the server and the client go on to key and algorithm negotiation; otherwise, the server breaks the TCP connection.

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, 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.

- 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.

Through the above steps, the server and the client get the same session key, which is to be used to encrypt and decrypt data exchanged between the server and the client later. The server and the client use session ID in the authentication stage.

CAUTION: *Before the phase of negotiation, the system has generated a server key pair and host key pair on the server. They are used for generating session keys. The server key pair is only available for SSH1.*

Authentication

- The client sends to the server an authentication request, which includes the username, authentication method and information related to the authentication method.
- 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.
- The client selects a method from the list to initiate another authentication.
- The above process repeats until the authentication succeeds or the authentication times timeout and the session is torn down.

SSH provides two authentication methods: password authentication and public key authentication.

In password authentication:

- The client encrypts the 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, compares them against those it maintains, and then informs the client of the authentication result.

In RSA authentication:

- The client sends an RSA authentication request (containing its public key) to the server. Upon receiving the request, the server checks its validity. If the request is not valid, the server directly sends a failure message. Otherwise, the server generates a 32-byte random number, arranges the random number into a multiple-precision (MP) integer according to the most significant bit (MSB), encrypts the MP integer using the public key of the client, and initiates an authentication challenge to the client.
- Upon receiving the challenge message, the client decrypts the MP integer using its own private key, generates a message abstract MD5 using the integer and session ID (an intermediate result generated in the key and algorithm negotiation phase), encrypts the 16-byte MD5 value, and then sends the encrypted MD5 value to the server.
- Upon receiving the MD5 value, the server reverts it to the original value, and compares the reverted MD5 value with the MD5 value calculated by itself. If the two MD5 values are the same, the server sends an authentication success message. Otherwise, the server sends an authentication failure message.

Besides password authentication and RSA authentication, SSH2.0 provides another two authentication methods:

- **password-publickey**: Performs both password authentication and publickey authentication of the client. A client running SSH1 client only needs to pass either type of the two, while a client running SSH2 client must pass both of them to login.
- **all**: Set the authentication mode to either "password" or "RSA". Clients will attempt to log in through RSA first.

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. If the client passes authentication, 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.

Interactive session

After a session is assigned successfully, the connection enters the interactive session mode. In this stage, the server and the client exchanges data in this 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.

- *During interactive session, the client can send the commands to be performed by pasting the text, which must be within 2000 bytes (including spaces). It is recommended that the text pasted be commands in the same view; otherwise, the server may not be able to perform the commands.*
- *If the text exceeds 2000 bytes, you can upload the configuration file to the server and use the configuration file to restart the server so that the server executes the commands.*

Introduction to SSH Configuration Tasks

Table 40 List of SSH configuration tasks

Configuration tasks		Remarks
Configuring the SSH server	Enabling SSH server	Required
	Configuring the protocol support for a user interface	Required
	Creating/Destroying/Exporting RSA Keys	Required
	Configuring authentication mode for SSH user	Optional
	Configuring service type for SSH users	Optional
	Setting the SSH Management Parameters	Optional
	Configuring RSA public key for the client	Required for the SSH users that use the RSA authentication mode
	Assigning RSA public keys to SSH users	Required for the SSH users that use the RSA authentication mode
Configuring the SSH client		Optional
Configuring the device as an SSH client		Optional

Configuring the SSH Server

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 Protocol Support for a User Interface

After enabling the SSH server, you must configure the protocol support for the involved interface(s). Note that the configuration takes effect at the next login.

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 single-user interface view or multi-user interface view	**user-interface** [*type-keyword*] *number* [*ending-number*]	Required
Set the login authentication method	**authentication-mode scheme** [**command-authorization**]	Required
Specify the protocols for the user interfaces to support	**protocol inbound** { **all** | **pad** | **ssh** | **telnet** }	Optional All protocols are supported by default.

CAUTION:

- *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 configure the* ***authentication-mode password*** *command and the* ***authentication-mode none*** *command.*

Creating/Destroying/Exporting RSA Keys

For successful SSH login, you must create the RSA key pairs first.

With SSH enabled, users still cannot log into the server through SSH if neither RSA host key pair nor server key pair is generated.

You can display the created RSA host public key on the screen in a specified format, or export it to a specified file for use when configuring the key at a remote site.

Follow these steps to create, destroy, or export the host key pair and server key pair:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Generate an RSA host key pair and server key pair	**rsa local-key-pair create**	Required
Destroy an RSA host key pair and server key pair	**rsa local-key-pair destroy**	Required
Display RSA host public keys in the screen in a specified format or export RSA host public keys to a specified file	**rsa local-key-pair export** { **ssh1** | **ssh2** | **openssh** } [*filename*]	Required Available in any view

CAUTION:

- *The configuration of the* ***rsa local-key-pair create*** *command can survive a reboot. You only need to configure it once.*
- *For a server key and host key, the minimum length is 512 bits, and the maximum length is 2,048 bits. In SSH2, some clients require that the keys generated on the server should be at least 768 bits in length.*
- *If you have configured a key pair, the system prompts whether you want to overwrite this key pair when you try to configure another key pair.*

Configuring Authentication Mode for SSH Users

A newly configured authentication mode will take effect when users log in next time.

Follow these steps to configure the authentication mode for SSH users.

To do...	Use the command...	Remarks
Enter system	**system-view**	-

To do...	Use the command...	Remarks
Configure an authentication mode for SSH users	**ssh user** *username* **authentication-type** { **password** | **rsa** | **password-publickey** | **all** }	Optional By default, the system specifies the authentication mode as "RSA".

CAUTION: *If a user uses the RSA authentication mode, this user and its public key must be configured on a switch. If a user uses the password authentication mode, his/her account information can be configured on a switch or remote authentication server (for example, a RADIUS authentication server).*

Configuring Service Type for SSH Users

Follow these steps to configure the service type for SSH users:

To do...	Use the command...	Remarks
Enter system view	**System-view**	-
Specify a service type for a specific user	**ssh user** *username* **service-type** { **stelnet** | **sftp** | **all** }	Required By default, the service type is Stelnet.

CAUTION:

- ***stelnet*** *(Secure Telnet) refers to the traditional SSH service. For details, refer to "SSH2.0 Overview" on page 929. For details about* ***sftp*** *(Secure FTP), refer to "SFTP Overview" on page 959.*
- *To log into the server through SFTP, you must set the service type to* ***sftp*** *or* ***all****. If the SFTP service is not used, you must set the service type to* ***stelnet*** *or* ***all****.*
- *SSH1 does not support the service type of* ***sftp****. If clients log into the server using SSH1, you must set the service type to* ***stelnet*** *or* ***all*** *on the server. Otherwise, clients cannot log into the server successfully.*

Setting the SSH Management Parameters

SSH management includes:

- Enabling the SSH server to be compatible with SSH1
- 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**	-
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.

To do...	Use the command...	Remarks
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

Authentication will fail if the number of authentication attempts (including both RSA and password authentication) exceeds that specified in the ***ssh server authentication-retries*** *command.*

Configuring RSA Public Key for the Client

This configuration is applicable when the RSA authentication mode is used for SSH users. If the password authentication mode is configured for SSH users, this configuration is not required.

The RSA public key configured on the device is for the SSH user on the client. On the client, you need to specify an RSA private key corresponding to the RSA public key for the SSH user. The key pair on the client is generated at random by the client software that supports SSH.

You can configure an RSA public key of the client manually or by importing from a public key file.

- For the first method, you can configure the host public key of the client to the server using Copy plus Paste.
- For the second method, the system automatically converts the public key file generated by the client software to PKCS codes, and configures the public key of the client. The public key file of the RSA key must be FTPed/TFTPed to the server in advance.

CAUTION: *When acting as an SSH server, the device cannot FTP the public key of the client to the server through Secure CRT 4.07.*

Follow these steps to configure the RSA public key of the client manually.

To do...	Use the Command...	Remarks
Enter system view	**system-view**	-
Enter public key view	**rsa peer-public-key** *keyname*	-
Enter public key editing view	**public-key-code begin**	-
Configure the public key of the client	Enter public key data directly	Required When you enter public key data, there can be spaces between characters, you can also press Enter to enter data continuously, and the configured public key must be a hexadecimal string of characters in the public key format.

Exit public key editing view to public key view	**public-key-code end**	- Save the entered public key data when exiting the view
Exit public key view to system view	**peer-public-key end**	-

Follow these steps to import RSA public key of the client from a public key file.

To do...	Use the Command...	Remarks
Enter system view	**system-view**	-
Import RSA public key of the SSH user from a public key file	**rsa peer-public-key** *keyname* **import sshkey** *filename*	Required

Assigning RSA Public Keys to SSH Users

If the SSH user uses the RSA authentication mode, you need to specify a public key of the client on the server. When the SSH client logs into the server, the server will authenticate the SSH client using the public key.

If the SSH user uses the password authentication mode, this configuration is not required.

Follow these steps to assign an RSA public key to the SSH user.

To do...	Use the Command...	Remarks
Enter system view	**system-view**	-
Assign an RSA public key to the SSH user	**ssh user** *username* **assign rsa-key** *keyname*	Required *keyname* indicates the name of an existing public key. When you execute this command, the last assigned public key will prevail if public keys are already assigned to the user.

- *An SSH user is created on the SSH server to specify an authentication mode, SSH service type, and public key for the user. You can create an SSH user by configuring any one among the* ***ssh user assign rsa-key****,* ***ssh user authentication-type****, and* ***ssh user service-type*** *commands. For the Switch 8800 Families, up to 1,024 SSH users can be created, the default authentication mode is RSA, and the default service type is stelnet.*
- *If no SSH user is created, the SSH client can still log into the SSH server through a local user (if created using the* ***local-user*** *command) so long as the service type is set to SSH. In this case, the default authentication mode is "password", and the default SSH service type is Stelnet.*

Configuring the SSH Client

There is a wide range of SSH client software, including PuTTY, and OpenSSH. To establish a connection between the SSH client and the server, you need to configure the SSH client as follows:

- Assign an IP address to the server.
- Set the remote connection protocol to SSH. Usually, the client can support a great variety of remote connection protocols, like Telnet, Rlogin, and SSH. To establish an SSH connection, you must set the remote connection protocol to SSH.
- Select an SSH version. The device currently supports SSH2.0, so you can select 2.0 or earlier versions.
- Specify an RSA private key file. If you configure the SSH user to use RSA authentication and specify an RSA public key for the SSH user on the server, you must specify a corresponding RSA private key file on the client. RSA key pairs are generated by the tools attached to the client software.

Using PuTTY, PuTTYGen and SSHKEY as an example, the section below describes how to configure the SSH client.

Generating a Client Key

On the SSH client (Windows XP was used in this example), two applications must be installed in addition to the Putty SSH, Putty, and Puttygen programs. You can download them from http://www.putty.nl/download.html.

You use Puttygen to create public and private keys. Save the keys to the local users' hard drive. There are two ways to import the client's public key on the Switch 8800. Only one method is needed to import the key, not both. Both methods need an external application to generate the public key. Puttygen is used in the examples.

The first method is to transfer the public key to the Switch 8800 flash:/ file system using FTP. Import the key using the **rsa peer-public-key (n*ame of key to be stored on the switch*) import sshkey (*the client's public key*)** command.

The second method is th use the **Sshconvert** command to convert the public key to the RSA key code (the **sshkey.exe** application). You use the output from this command a as input to the switch when prompted after entering the following commands:

rsa peer-public-key (string chosen as the public key)

public-key-code begin

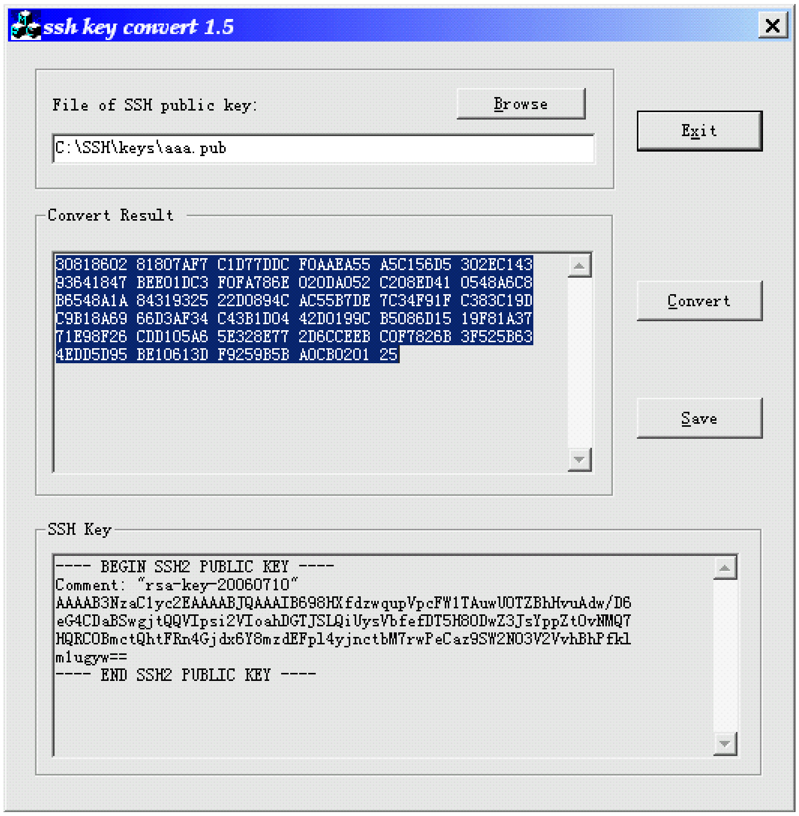
Then when prompted for **rsa-key-code**, enter the large hex string highlighted in Figure 281 by copying and pasting the code.

public-key-code end

peer-public-key end

Both methods are described in more detail below.

Figure 281 Generate a client key (1)



File of SSH public key: C:\SSH\keys\aaa.pub

Convert Result

```
30818602 81807AF7 C1D77DDC F0AAEA55 A5C156D5 302EC143
93641847 BEE01DC3 F0FA786E 020DA052 C208ED41 0548A6C8
B6548A1A 84319325 22D0894C AC55B7DE 7C34F91F C383C19D
C9B18A69 66D3AF34 C43B1D04 42D0199C B5086D15 19F81A37
71E98F26 CDD105A6 5E328E77 2D6CCEEB C0F7826B 3F525B63
4EDD5D95 BE10613D F9259B5B A0CB0201 25
```

SSH Key

```
---- BEGIN SSH2 PUBLIC KEY ----
Comment: "rsa-key-20060710"
AAAAB3NzaC1yc2EAAAABJQAAAIB698HXfdzwqupVpcFW1TAuwUOTZBhHvuAdw/D6
eG4CDaBSwgjtQQVIpsi2VIoahDGTJSLQiUysVbfefDT5H8ODwZ3JsYppZtOvNMQ7
HQRCOBmctQhtFRn4Gjdx6Y8mzdEFpl4yjnctbM7rwPeCaz9SW2NO3V2VvhBhPfkl
m1ugyw==
---- END SSH2 PUBLIC KEY ----
```

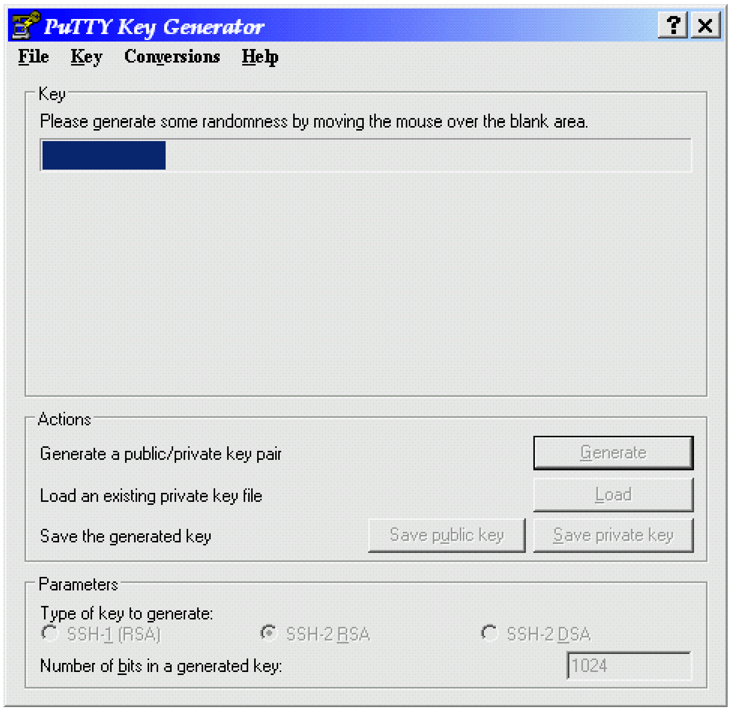
You need issue the **rsa local-key-pair create** command on the switch once as shown below.

1 Enter the system view and issue the following bold commands.

```
[SW8800] rsa local-key-pair create
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
       It will take a few minutes.
Input the bits in the modulus[default = 1024]:1024
Generating keys...
...................++++++
.......++++++
...................++++++++
.....++++++++
......Done!
[SW8800]
[SW8800]user-interface vty 0 4
[SW8800-ui-vty0-4]authen scheme
 Notice: Exec(Telnet) user must be added, otherwise operator can't
login!

( Telnet users are already configured on this switch)

[SW8800-ui-vty0-4]proto inbound all
[SW8800-ui-vty0-4]q
```

2 To generate the RSA keys, on the SSH client, generate the public and private keys using the puttygen application by clicking the **Generate** button on the PuTTY Key Generator dialog box (see Figure 282).

Figure 282 Generate a client key (2)



Please generate some randomness by moving the mouse over the blank area.

Generate a public/private key pair

Load an existing private key file

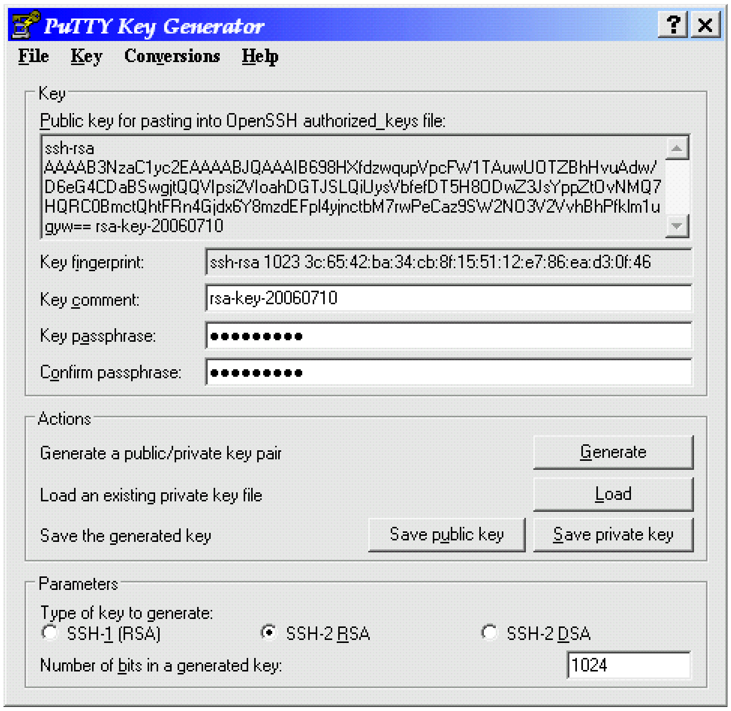
Save the generated key

Type of key to generate:
○ SSH-1 (RSA) ● SSH-2 RSA ○ SSH-2 DSA

Number of bits in a generated key: 1024

3 Move the mouse over the blank area as instructed by the prompt in the **Key** section of the dialog box. When the key is generated, the dialog box in Figure 283 appears.

Figure 283 Generate a client key (3)



Public key for pasting into OpenSSH authorized_keys file:

```
ssh-rsa
AAAAB3NzaC1yc2EAAAABJQAAAIB698HXfdzwqupVpcFW1TAuwUOTZBhHvuAdw/
D6eG4CDaBSwgjtQQVlpsi2VloahDGTJSLQiUysVbfefDT5H8ODwZ3JsYppZtOvNMQ7
HQRC0BmctQhtFRn4Gjdx6Y8mzdEFpl4yjnctbM7rwPeCaz9SW2NO3V2VvhBhPfklm1u
gyw== rsa-key-20060710
```

Key fingerprint: ssh-rsa 1023 3c:65:42:ba:34:cb:8f:15:51:12:e7:86:ea:d3:0f:46

Key comment: rsa-key-20060710

Key passphrase: ●●●●●●●●●●

Confirm passphrase: ●●●●●●●●●●

Generate a public/private key pair

Load an existing private key file

Save the generated key

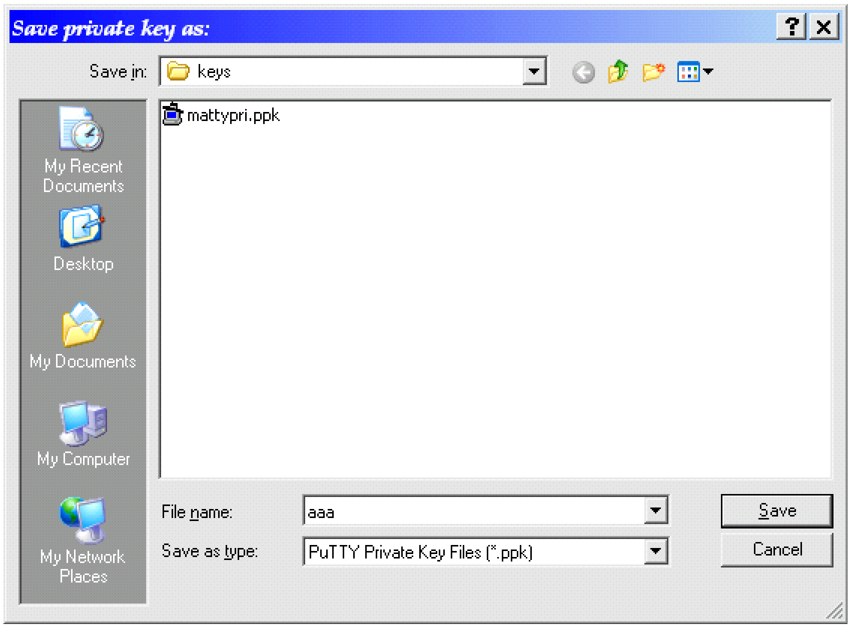
Type of key to generate:
○ SSH-1 (RSA) ● SSH-2 RSA ○ SSH-2 DSA

Number of bits in a generated key: 1024

4 You can optionally specify a passphrase in the **Key passphrase** field and repeat that key in the **Confirm passphrase** field as shown above. Make sure to remember this phrase because you will need it later in this procedure.

5 Click the **Save private key** button. The **Save private key as:** dialog box is displayed as shown in Figure 284. Enter a name for that key and click **Save**.

Figure 284 Saving the private key



Save in: keys

mattypri.ppk

File name: aaa

Save as type: PuTTY Private Key Files (*.ppk)

6 Click **Save public key** and specify a name for that key.

The private key has a ***.ppk*** *extension the public key as a* ***.pub*** *extension.*

Importing the Client Public Key Using Method One

This section describes how to import the client public key to the Switch 8800 file system using FTP.

1 Enable the FTP server on the Switch 8800 from the system view as follows:

```
<SW8800>sys
System View: return to User View with Ctrl+Z.
[SW8800]ftp server enable
% FTP server has been started

[SW8800]
```

2 Create an FTP user on the SW8800 so that the client can FTP to the Switch 8800 and transfer the public key.

```
[SW8800]local-user ftp
New local user added.
[SW8800-luser-ftp]password ?
  cipher  Display password with cipher text
  simple  Display password with plain text

[SW8800-luser-ftp]password simple ftp
[SW8800-luser-ftp]level 3
[SW8800-luser-ftp]service-type ftp
[SW8800-luser-ftp]q
[SW8800]
```

3 From the client PC where the key is stored, FTP to the Switch 8800. This example uses ftp from a command (or MS-DOS) window. Change to the directory where the key resides. The IP address of the SW8800 is 158.101.23.252.

```
C:\SSH\keys>
C:\SSH\keys>ftp 158.101.23.252
Connected to 158.101.23.252.
220 FTP service ready.
User (158.101.23.252:(none)): ftp
331 Password required for ftp.
Password: (the password for ftp user, will not be displayed)
230 User logged in.
ftp> bin
200 Type set to I.
ftp> put aaa.pub
200 Port command okay.
150 Opening BINARY mode data connection for aaa.pub.
226 Transfer complete.
ftp: 294 bytes sent in 0.00Seconds 294000.00Kbytes/sec.
ftp> bye
221 Server closing.
C:\SSH\keys>
```

4 Return to the Switch 8800:

```
[SW8800]rsa peer-public-key SWSW8800002 import sshkey aaa.pub
The public key is successfully imported from the file.
```

*If you display the directory's contents (using the **dir** command it displays the **aaa.pub** file.*

5 To remove the user ftp:

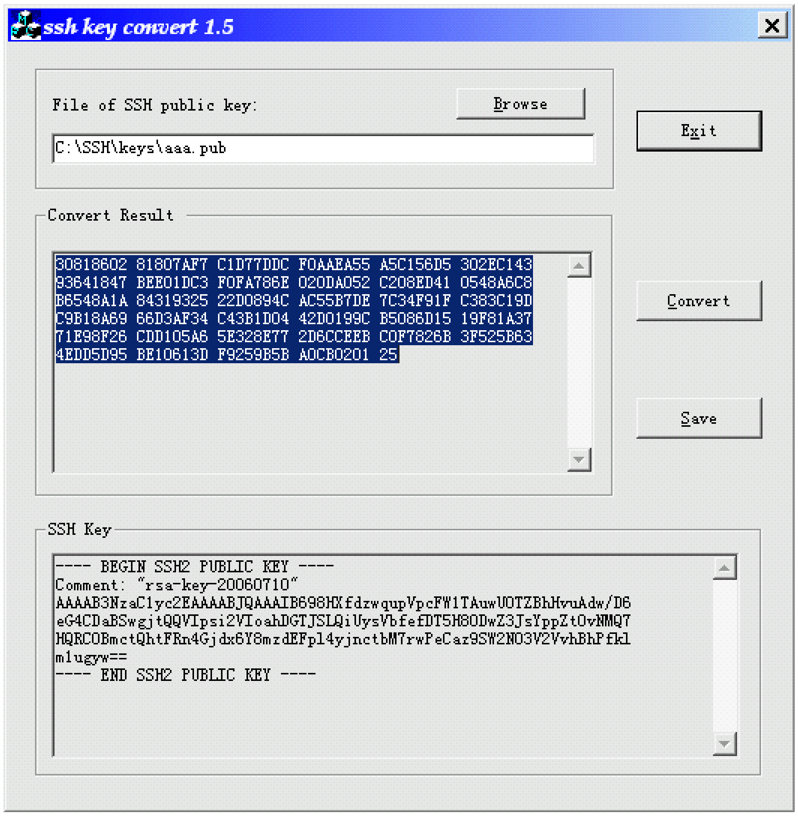
```
[SW8800] undo local-user ftp
Updating user(s) information, please wait.........
[SW8800]
```

Importing the Client Public Key Using Method Two

When using this method, the SSH client uses the application **sshkey.exe** to convert the public key, creating a large hex string to be entered on the Switch 8800. From the **sshkey.exe** application:

1 Browse to where you stored the public key on the client PC, and convert the key. The ssh key convert dialog box displays the large hex string as shown in Figure 285.

Figure 285 Convering the public key



File of SSH public key:

C:\SSH\keys\aaa.pub

Convert Result

```
30818602 81807AF7 C1D77DDC F0AAEA55 A5C156D5 302EC143
93641847 BEE01DC3 F0FA786E 020DA052 C208ED41 0548A6C8
B6548A1A 84319325 22D0894C AC55B7DE 7C34F91F C383C19D
C9B18A69 66D3AF34 C43B1D04 42D0199C B5086D15 19F81A37
71E98F26 CDD105A6 5E328E77 2D6CCEEB C0F7826B 3F525B63
4EDD5D95 BE10613D F9259B5B A0CB0201 25
```

SSH Key

```
---- BEGIN SSH2 PUBLIC KEY ----
Comment: "rsa-key-20060710"
AAAAB3NzaC1yc2EAAAABJQAAAIB698HXfdzwqupVpcFW1TAuwUOTZBhHvuAdw/D6
eG4CDaBSwgjtQQVIpsi2VIoahDGTJSLQiUysVbfefDT5H8ODwZ3JsYppZtOvNMQ7
HQRCOBmctQhtFRn4Gjdx6Y8mzdEFpl4yjnctbM7rwPeCaz9SW2NO3V2VvhBhPfkl
m1ugyw==
---- END SSH2 PUBLIC KEY ----
```

2 If your client is connected to the Switch 8800, you can copy the Convert Result hex string from this screen using the **Ctrl + c,** Then go to the switch's CLI and past the hex string into the switch.

The name of the key is SWSW8800002 in this example.

```
[SW8800] rsa peer-public-key SWSW8800002

[SW8800-rsa-public-key] public-key-code begin
RSA key code view: return to last view with "public-key-code end".

[SW8800-rsa-key-code]30818602 81807AF7 C1D77DDC F0AAEA55 A5C156D5 30
```

```
2EC143
[SW8800-rsa-key-code]93641847 BEE01DC3 F0FA786E 020DA052 C208ED41 05
48A6C8
[SW8800-rsa-key-code]B6548A1A 84319325 22D0894C AC55B7DE 7C34F91F C3
83C19D
[SW8800-rsa-key-code]C9B18A69 66D3AF34 C43B1D04 42D0199C B5086D15 19
F81A37
[SW8800-rsa-key-code]71E98F26 CDD105A6 5E328E77 2D6CCEEB C0F7826B 3F
525B63
[SW8800-rsa-key-code]4EDD5D95 BE10613D F9259B5B A0CB0201 25
```

You may have to hit the "enter" key if the key code does not end at the end of the line.

```
[SW8800-rsa-key-code]public-key-code end
[SW8800-rsa-public-key]peer-public-key end
```

Displaying the Public Key

To display the key that was just created, enter:

```
[SW8800]display rsa peer-public-key

=====================================
    Key name: SWSW8800002
    Key address:
=====================================
Key Code:
308186
  028180
    7AF7C1D7 7DDCF0AA EA55A5C1 56D5302E C1439364 1847BEE0 1DC3F0FA
786E020D
    A052C208 ED410548 A6C8B654 8A1A8431 932522D0 894CAC55 B7DE7C34
F91FC383
    C19DC9B1 8A6966D3 AF34C43B 1D0442D0 199CB508 6D1519F8 1A3771E9
8F26CDD1
    05A65E32 8E772D6C CEEBC0F7 826B3F52 5B634EDD 5D95BE10 613DF925
9B5BA0CB
  0201
    25

[SW8800]
```

Creating the SSH UserID and Asociating it with the Client

Enter the following commands to create the SSH user id (client002 in this example) and associate a public key with that ID (SWSW8800002 in this example)

```
[SW8800]ssh user client002 assign rsa-key SWSW8800002
Info: Successful to create SSH user.
[SW8800]ssh user client002 authentication-type RSA
[SW8800]ssh user client002 service-type ?
  all      Specify service name
  sftp     Specify service name
  stelnet  Specify service name

[SW8800]ssh user client002 service-type all
[SW8800]dis ssh user client002
 Username    Authentication-type  User-public-key-name  Service-type
 client002              rsa                     SWSW8800002
stelnet|sftp
```

Assigning an IP Address to the Server

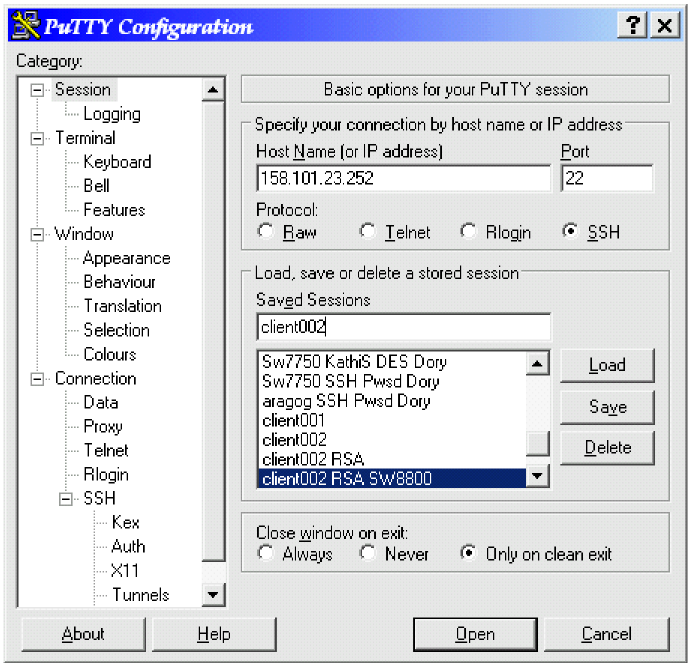
To configure the IP address:

1 Execute PuTTY.exe. The system displays a client configuration interface.

2 Specify the IP address in the **Host Name** field.

3 If you wish to save the configuration, enter a session name in the **Saved Sessions** field as shown in Figure 286.

Figure 286 SSH client configuration (1)



PuTTY Configuration

Basic options for your PuTTY session

Specify your connection by host name or IP address

Host Name (or IP address): 158.101.23.252

Port: 22

Protocol: ○ Raw ○ Telnet ○ Rlogin ● SSH

Load, save or delete a stored session

Saved Sessions: client002

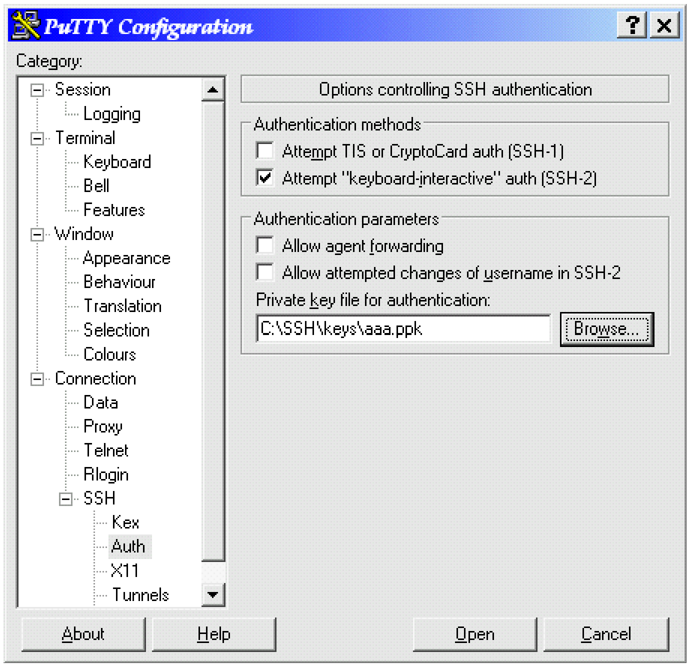
Sw7750 KathiS DES Dory
Sw7750 SSH Pwsd Dory
aragog SSH Pwsd Dory
client001
client002
client002 RSA
client002 RSA SW8800

Close window on exit: ○ Always ○ Never ● Only on clean exit

4 Point the private key by clicking on the **SSH > Auth** in the Category tree. The options for controlling SSH authentication.

5 Browse to the location of the private key you saved earlier in this chapter and choose that file as shown in Figure 287.

Figure 287 SSH client configuration (2)



Options controlling SSH authentication

Authentication methods

☐ Attempt TIS or CryptoCard auth (SSH-1)

☑ Attempt "keyboard-interactive" auth (SSH-2)

Authentication parameters

☐ Allow agent forwarding

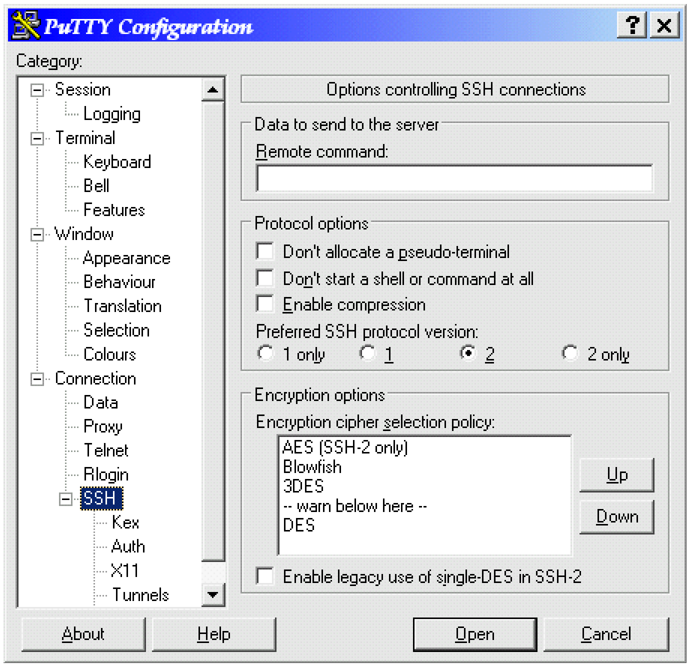
☐ Allow attempted changes of username in SSH-2

Private key file for authentication:

C:\SSH\keys\aaa.ppk Browse...

6 Check the SSH version by clicking on SSH. This example is using Version 2 as shown in Figure 288.

Figure 288 SSH client configuration (3)



Options controlling SSH connections

Data to send to the server

Remote command:

Protocol options

☐ Don't allocate a pseudo-terminal

☐ Don't start a shell or command at all

☐ Enable compression

Preferred SSH protocol version:

○ 1 only ○ 1 ● 2 ○ 2 only

Encryption options

Encryption cipher selection policy:

AES (SSH-2 only)
Blowfish
3DES
-- warn below here --
DES

Up Down

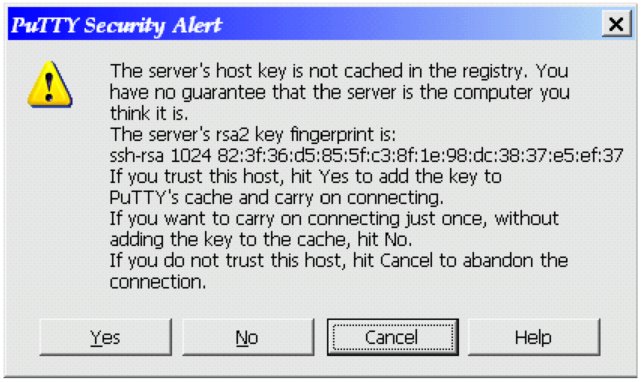
☐ Enable legacy use of single-DES in SSH-2

7 Save by profile by clicking on the session then **Save**.

Communicating with the Switch

ITo communicate with the switch, click **Open**. A dialog box with the switch's IP address in the title bar is displayed. Also, a PuTTy Security Alert is displayed (as shown in Figure 289) after invoking putty when using RSA for the first time. After the first time, this message is not displayed unless you generate new keys. If you are using SSH with a client using only a passwork, this is not displayed

Figure 289 PuTTY Security Alter

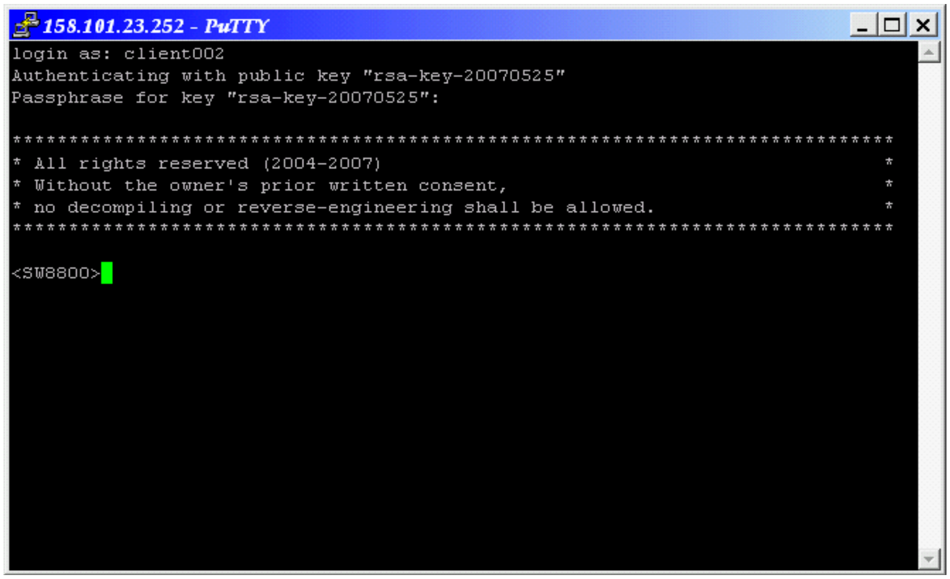


PuTTY Security Alert

The server's host key is not cached in the registry. You have no guarantee that the server is the computer you think it is.
The server's rsa2 key fingerprint is:
ssh-rsa 1024 82:3f:36:d5:85:5f:c3:8f:1e:98:dc:38:37:e5:ef:37
If you trust this host, hit Yes to add the key to PuTTY's cache and carry on connecting.
If you want to carry on connecting just once, without adding the key to the cache, hit No.
If you do not trust this host, hit Cancel to abandon the connection.

Click **Yes** to continue. You are then prompted for login id (client002 in this example) and the passphrase. Specify the passphrase you created earlier in this chapter and you should be successfully connected and logged into the switch as shown in Figure 290.

Figure 290 SSH client



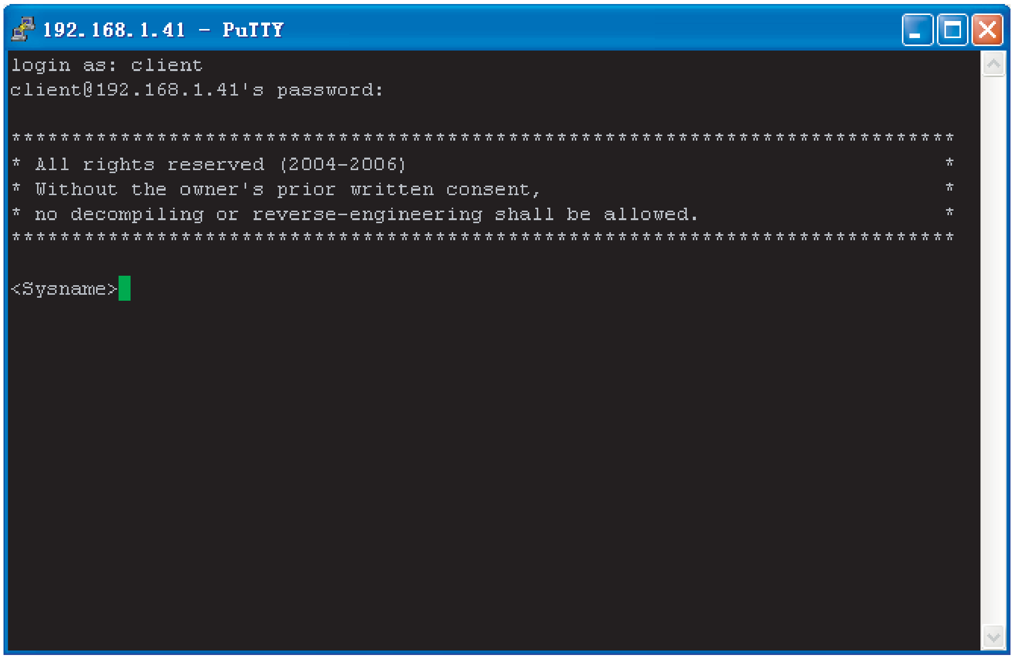
158.101.23.252 - PuTTY

```
login as: client002
Authenticating with public key "rsa-key-20070525"
Passphrase for key "rsa-key-20070525":

******************************************************************************
* All rights reserved (2004-2007)                                            *
* Without the owner's prior written consent,                                 *
* no decompiling or reverse-engineering shall be allowed.                    *
******************************************************************************

<SW8800>
```

Opening an SSH Connection Using a Password

1 Click **Open**. Then the system displays an SSH client interface, as shown in Figure 288. If the connection is normal, the system will prompt you to enter a username and password.

Figure 291 SSH client



```
login as: client
client@192.168.1.41's password:

*****************************************************************************
* All rights reserved (2004-2006)                                           *
* Without the owner's prior written consent,                                *
* no decompiling or reverse-engineering shall be allowed.                   *
*****************************************************************************

<Sysname>
```

2 Enter the correct username and password to log into the server successfully.

3 To log out of the SSH server, execute the **quit** command.

Configuring the Device as an SSH Client

Prerequisite

Configure the SSH server completely. For details, refer to "Configuring the SSH Server" on page 934.

Configuring the Device as an SSH Client

When the device, as an SSH client, is connected to the SSH server, you can configure the SSH client whether to perform first authentication to the accessed SSH server.

- First authentication: When the SSH client accesses the SSH server for the first time but is not configured with the host public key of the server, users can choose to access the server continuously and save the host public key on the client. When users access the server next time, the saved host public key will be used to authenticate the server.
- If first authentication is not supported, the client will refuse to access the server if not configured with the host public key of the server. Users must configure the host public key of the server to be accessed on the local device in advance,

and specify the name of the host public key of the server to be connected, so that the client can authenticate the server to be connected.

In addition, you can configure the client to access the SSH server using a specified IP address or port address.

Configure the SSH client that supports first authentication

Follow these steps to configure the SSH client that supports first authentication.

To do…		Use the command…	Remarks
Enter system view		**system-view**	-
Configure the SSH client to perform first authentication to the accessed SSH server		**ssh client first-time**	Optional By default, first authentication is performed on the client.
Specify an source IP address or source interface for the SSH client	Specify an source IPv4 address or source interface for the SSH client	**ssh client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	Optional By default, the client accesses the SSH server using the interface address specified by the device route.
	Specify an source IPv6 address or source interface for the SSH client	**ssh client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	
Establish a connection between the SSH client and server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	Establish a connection between the SSH client and IPv4 server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	**ssh2** { *host-ip* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **aes128** | **des** } | **prefer_stoc_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** }] *	Use one command
	Establish a connection between the SSH client and IPv6 server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	**ssh2 ipv6** { *ipv6-address* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **aes128** | **des** } | **prefer_stoc_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** }] *	

Configure the SSH client that supports first authentication

Follow these steps to configure the SSH client that does not support first authentication.

To do...		Use the command...	Remarks
Enter system view		**system-view**	-
Configure the SSH client not to perform first authentication to the accessed SSH server		**undo ssh client first-time**	Required By default, first authentication is performed on the client.
Enter public key view		**rsa peer-public-key** *keyname*	-
Enter public key editing view		**public-key-code begin**	-
Configure the public key of the server		Enter the public key data directly	- When you enter public key data, there can be spaces between characters, you can also press Enter to enter data continuously, and the configured public key must be a hexadecimal string of characters in the public key format.
Return to public key view		**public-key-code end**	- Save the entered public key data when exiting the view
Return to system view		**peer-public-key end**	-
Specify the name of the host public key of the server to be connected on the client		**ssh client authentication server** { *server-ip* | *server-name* } **assign rsa-key** *keyname*	Required
Specify an source IP address or source interface for the SSH client	Specify an source IPv4 address or source interface for the SSH client	**ssh client source** { **ip** *ip-address* | **interface** *interface-type interface-number* }	Optional By default, the client accesses the SSH server using the interface address specified by the device route.
	Specify an source IPv6 address or source interface for the SSH client	**ssh client ipv6 source** { **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }	

To do...		Use the command...	Remarks
Establish a connection between the SSH client and server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	Establish a connection between the SSH client and IPv4 server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	**ssh2** { *host-ip* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **aes128** | **des** } | **prefer_stoc_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** }] *	Use one command
	Establish a connection between the SSH client and IPv6 server, and specify the preferred key exchange algorithm, preferred encryption algorithm, and preferred HMAC algorithm for the client and server	**ssh2 ipv6** { *ipv6-address* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **aes128** | **des** } | **prefer_stoc_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** }] *	

Displaying and Maintaining SSH

Follow these steps to display and maintain the SSH protocol.

To do...	Use the command...	Remarks
View the public key information of the host key pair and server key pair	**display rsa local-key-pair public**	Available in any view
Display the remote RSA public key	**display rsa peer-public-key** [**brief** | **name** *keyname*]	Available in any view
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 the status information or session information of an SSH server	**display ssh server** { **status** | **session** }	Available in any view
Display the mappings between host public keys and SSH servers saved on a 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

SSH Server Configuration Examples

Network requirements

As shown in Figure 292, establish a local connection between the terminal (SSH client) and the Ethernet switch. The terminal logs into the switch through SSH, so as to ensure security of data exchange. For the SSH client, the username is “client001”, and the password is “aabbccddeeff”.

Network diagram

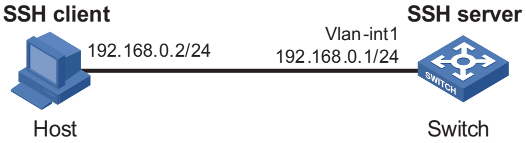


Figure 292 Local configuration of SSH

Configuration procedure

1 Configure the SSH server, Switch

# Generate a key pair and enable the SSH server.

```
<Switch> system-view
[Switch] rsa local-key-pair create
[Switch] ssh server enable
```

# Assign an IP address to the Vlan-interface 1. The client will be connected to the SSH server through this address.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.0.1 255.255.255.0
[Switch-Vlan-interface1] quit
```

The IP address of the client host and the IP address of the VLAN interface on the switch must be in a network segment. It is set to 192.168.0.2.

2 Configure the password authentication mode for the SSH user

# Configure the SSH client to log into the user interface through AAA

```
[Switch] user-interface vty 1
[Switch-ui-vty1] authentication-mode scheme
```

# Set the remote user login protocol on the switch to SSH.

```
[Switch-ui-vty1] protocol inbound ssh
[Switch-ui-vty1] quit
```

# Create a local user named “client001”.

```
[Switch] local-user client001
[Switch-luser-client001] password simple aabbccddeeff
[Switch-luser-client001] service-type ssh level 3
[Switch-luser-client001] quit
[Switch] ssh user client001 authentication-type password
```

Configure the authentication timeout time, number of attempts, and server key update interval as default values.

Then, you need to run the SSH2.0-capable client software on the terminal connected to the switch, configure the IP address of the reachable interface of the SSH server (switch) to 192.168.0.1, configure the protocol type as SSH, and configure the protocol version to 2. Launch the SSH connection, and enter the username “client001” and password “aabbccddeeff” as prompted. Then, you can enter the configuration interface of the switch.

```
login as: client001
client001@192.168.0.1’s password:

*********************************************************
*All rights reserved (2004-2006)                        *
*Without the owner’s prior written consent,             *
*no decompiling or reverse-switch fabricering shall be allowed.*
*********************************************************

<Switch>
```

3 Configure the RSA authentication mode for the SSH user

# Configure AAA on the user interface.

```
[Switch] user-interface vty 1
[Switch-ui-vty1] authentication-mode scheme
```

# Set the remote user login protocol on the switch to SSH.

```
[Switch-ui-vty1] protocol inbound ssh
```

# Set the privilege level to 3 for the user.

```
[Switch-ui-vty1] user privilege level 3
[Switch-ui-vty1] quit
```

# Set the authentication mode to RSA for the remote user "client001" on the switch.

```
[Switch] ssh user client001 authentication-type rsa
```

Then, you need to generate an RSA key pair (including public key and private key) at random on the SSH2.0-capable client software, and configure the RSA public key (the RSA public key is a PKCS-compliant hexadecimal string that is encoded by the SSHKEY.EXE software) to the specified rsa peer-public-key on the SSH server.

# Set an RSA key on the switch.

```
[Switch] rsa peer-public-key Switch001
[Switch-rsa-public-key] public-key-code begin
[Switch-rsa-key-code]30818602 818078C4 32AD7864 BB0137AA 516284BB 3F55F0E3
[Switch-rsa-key-code]F6DD9FC2 4A570215 68D2B3F7 5188A1C3 2B2D40BE D47A08FA
[Switch-rsa-key-code]CF41AF4E 8CCC2ED0 C5F9D1C5 22FC0625 BA54BCB3 D1CBB500
[Switch-rsa-key-code]A177E917 642BE3B5 C683B0EB 1EC041F0 08EF60B7 8B6ED628
[Switch-rsa-key-code]9830ED46 0BA21FDB F55E7C81 5D1A2045 54BFC853 5358E5CF
[Switch-rsa-key-code]7D7DDF25 03C44C00 E2F49539 5C4B0201 25
```

```
[Switch-rsa-key-code] public-key-code end
[Switch-rsa-public-key] peer-public-key end
```

# If the server stores the public key of the client through a file named "Switch001", you can import the public key directly from the file.

```
[Switch] rsa peer-public-key Switch001 import sshkey Switch001
```

# Specify the public key "Switch001" for the user "client001".

```
[Switch] ssh user client001 assign rsa-key Switch001
```

For RSA authentication, you need to configure the IP address, protocol type, and protocol version of the SSH server on the client, and to specify an RSA private key file (generated by the client software at random). Launch the SSH connection, and enter a username and password as prompted. Then, you can enter the configuration interface of the switch.

```
login as: client001
Authenticating with public key "rsa-key-20061023"

*********************************************************
*All rights reserved (2004-2006)                        *
*Without the owner’s prior written consent,             *
*no decompiling or reverse-switch fabricering shall be allowed.*
*********************************************************

<Switch>
```

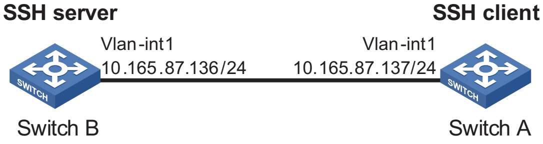
SSH Client Configuration Examples

Network Requirements

As shown in Figure 293, configure Switch A as a client, and configure Switch A to log into Switch B through SSH. For the SSH client, the username is "client001", and the password is "aabbccddeeff".

Network Diagram

Figure 293 SSH client configuration



Configuration

1 Configure Switch B

# Generate an RSA host key pair and server key pair, and enable the SSH server.

```
<SwitchB> system-view
[SwitchB] rsa local-key-pair create
[SwitchB] ssh server enable
```

# Assign an IP address to the Vlan-interface 1. The client will be connected to the SSH server through this address.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 10.165.87.136 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

# Configure the SSH client to log into the user interface through AAA

```
[SwitchB] user-interface vty 1
[SwitchB-ui-vty1] authentication-mode scheme
```

# Set the remote user login protocol on the switch to SSH.

```
[SwitchB-ui-vty1] protocol inbound ssh
[SwitchB-ui-vty1] quit
```

# Create a local user named "client001".

```
[SwitchB] local-user client001
[SwitchB-luser-client001] password simple aabbccddeeff
[SwitchB-luser-client001] service-type ssh level 3
[SwitchB-luser-client001] quit
```

# Configure the password authentication mode for the SSH user. Configure the authentication timeout time, number of attempts, and server key update interval as default values.

```
[SwitchB] ssh user client001 authentication-type password
```

If configuring RSA authentication for the SSH user, you need to configure a host public key for Switch A. For details, refer to related section in "SSH Server Configuration Examples" on page 953.

2 Configure Switch A

# The IP address of the Vlan interface on Switch A and the IP address of the Vlan interface on Switch B must be in the same network segment. It is set to 10.165.87.137.

```
<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
```

# Configure the client not to perform first authentication to the server.

```
[SwitchA] undo ssh client first-time
```

# Configure a host public key for the SSH server on the client.

```
[SwitchA] rsa peer-public-key public
[SwitchA-rsa-public-key]public-key-code begin
[SwitchA-rsa-key-code]308186028180739A291ABDA704F5D93DC8FDF84C427463
```

```
[SwitchA-rsa-key-code]1991C164B0DF178C55FA833591C7D47D5381D09CE82913
[SwitchA-rsa-key-code]D7EDF9C08511D83CA4ED2B30B809808EB0D1F52D045DE4
[SwitchA-rsa-key-code]0861B74A0E135523CCD74CAC61F8E58C452B2F3F2DA0DC
[SwitchA-rsa-key-code]C48E3306367FE187BDD944018B3B69F3CBB0A573202C16
[SwitchA-rsa-key-code]BB2FC1ACF3EC8F828D55A36F1CDDC4BB45504F020125
[SwitchA-rsa-key-code] public-key-code end
[SwitchA-rsa-public-key] peer-public-key end
[SwitchA] ssh client authentication server 10.165.87.136 assign rsa-key public
```

# Establish an SSH connection to the server with the IP address of 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:
*********************************************************
*All rights reserved (2004-2006)                        *
*Without the owner’s prior written consent,             *
*no decompiling or reverse-switch fabricering shall be allowed.*
*********************************************************

<SwitchB>
```

73 SFTP SERVICE

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

- "SFTP Overview" on page 959
- "Configuring an SFTP Server" on page 959
- "Configuring an SFTP Client" on page 960
- "SFTP Configuration Examples" on page 964

SFTP Overview

The secure file transfer protocol (SFTP) is a new feature in SSH 2.0.

SFTP uses the SSH connection to provide secure data transfer. The device can serve as the SFTP server, allowing a remote user to login to 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 "Introduction to SSH Configuration Tasks" on page 934.
- 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 Service Type for SSH Users" on page 936.

Enabling the SFTP Server

This configuration task is to enable the SFTP service so that a client can login to 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

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.

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*	Required 10 minutes by default

Configuring an SFTP Client

Specifying a Source IP Address or Interface for the SFTP Client

The client accesses the SFTP server using the specified IP address and port address.

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* }	Use one command as required. By default, an SFTP client uses the port 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
Enter system view	**system-view**	-

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** { *host-ip* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **aes128** | **des** } | **prefer_stoc_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** }] *	Use one command
	Establish a connection to the remote IPv6 SFTP server, and enter SFTP Client view	**sftp ipv6** { *ipv6-address* | *host-name* } [*port-number*] [**prefer_ctos_cipher** { **3des** | **aes128** | **des** } | **prefer_ctos_hmac** { **md5** | **md5_96** | **sha1** | **sha1_96** } | **prefer_kex** { **dh_exchange_group** | **dh_group1** } | **prefer_stoc_cipher** { **3des** | **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 system view	**system-view**	-
Establish a connection to the remote SFTP server, and enter SFTP Client view	**sftp** { *host-ip* | *host-name* } [*port-number*] [**prefer_kex** { **dh_group1** | **dh_exchange_group** } | **prefer_ctos_cipher** { **des** | **aes128** | **3des** } | **prefer_stoc_cipher** { **des** | **aes128** | **3des** } | **prefer_ctos_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** } | **prefer_stoc_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** }]*	Required
Change the working directory of the remote SFTP server	**cd** [*remote-path*]	Optional
Return to the upper-level directory	**cdup**	Optional

To do...	Use the command...	Remarks
Display the current working directory of the remote SFTP server	**pwd**	Optional
Display files under a specified directory	**dir** [**-a** | **-l**] [*remote-path*]	Optional
	ls [**-a** | **-l**] [*remote-path*]	The **dir** command functions the same as the **ls** command.
Change the name of a specified directory on the SFTP server	**rename** *oldname newname*	Optional
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 system view	**system-view**	-
Establish a connection to the remote SFTP server, and enter SFTP Client view	**sftp** { *host-ip* | *host-name* } [*port-number*] [**prefer_kex** { **dh_group1** | **dh_exchange_group** } | **prefer_ctos_cipher** { **des** | **aes128** | **3des** } | **prefer_stoc_cipher** { **des** | **aes128** | **3des** } | **prefer_ctos_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** } | **prefer_stoc_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** }]*	Required
Change the name of a specified file 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
	ls [**-a** | **-l**] [*remote-path*]	The **dir** command functions the same as the **ls** command.

To do...	Use the command...	Remarks
Delete a file from the SFTP server	**delete** *remote-file*&<1-10>	Optional
	remove *remote-file*&<1-10>	The **delete** command functions the same as the **remove** command.

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 system view	**system-view**	-
Establish a connection to the remote SFTP server, and enter SFTP Client view	**sftp** { *host-ip* | *host-name* } [*port-number*] [**prefer_kex** { **dh_group1** | **dh_exchange_group** } | **prefer_ctos_cipher** { **des** | **aes128** | **3des** } | **prefer_stoc_cipher** { **des** | **aes128** | **3des** } | **prefer_ctos_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** } | **prefer_stoc_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** }]*	Required
Display a list of all commands or the help information of an SFTP client command	**help** [**all** | *command-name*]	Required

Disabling the SFTP Client

This configuration task is to disable the SFTP client.

Follow these steps to disable the SFTP client:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Establish a connection to the remote SFTP server, and enter SFTP Client view	**sftp** { *host-ip* | *host-name* } [*port-number*] [**prefer_kex** { **dh_group1** | **dh_exchange_group** } | **prefer_ctos_cipher** { **des** | **aes128** | **3des** } | **prefer_stoc_cipher** { **des** | **aes128** | **3des** } | **prefer_ctos_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** } | **prefer_stoc_hmac** { **sha1** | **sha1_96** | **md5** | **md5_96** }] *	-
Terminate the connection to the remote SFTP server and return to system view	**bye**	Required. Use any command.
	exit	These three commands have the same function.
	quit	

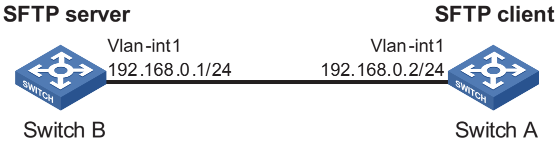
SFTP Configuration Examples

Network requirements

As shown in Figure 294, an SSH connection is established between Switch A and Switch B. Switch A, an SFTP client, uses the username client001 and password aabbcc to login to Switch B for file management and file transfer.

Network diagram

Figure 294 Network diagram for SFTP configuration (on routers)



Configuration procedure

1 Configure the SFTP server (Switch B)

# Generate an RSA host key pair and server key pair, and enable the SSH server.

```
<SwitchB> system-view
[SwitchB] rsa local-key-pair create
[SwitchB] ssh server enable
```

# Assign an IP address to the Vlan-interface 1. The client will be connected to the SSH server through this address.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

# Configure the SSH client to log into the user interface through AAA

```
[SwitchB] user-interface vty 1
[SwitchB-ui-vty1] authentication-mode scheme
```

# Set the remote user login protocol on the switch to SSH.

```
[SwitchB-ui-vty1] protocol inbound ssh
[SwitchB-ui-vty1] quit
```

# Create a local user named "client001".

```
[SwitchB] local-user client001
[SwitchB-luser-client001] password simple aabbccddeeff
[SwitchB-luser-client001] service-type ssh
[SwitchB-luser-client001] quit
```

# Configure the password authentication mode for the SSH user. Configure the authentication timeout time, number of attempts, and server key update interval as default values.

```
[SwitchB] ssh user client001 authentication-type password
```

If configuring RSA authentication for the SSH user, you need to configure a host public key for Switch A. For details, refer to related section in "SSH Server Configuration Examples" on page 953.

# Enable the SFTP server.

```
[SwitchB] sftp server enable
```

# Specify the service type as SFTP for the user.

```
[SwitchB] ssh user client001 service-type sftp
```

2 Configure the client (Switch A)

# The IP address of the Vlan interface on Switch A and the IP address of the Vlan interface on Switch B must be in the same network segment. It is set to 192.168.0.2.

```
<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
```

# Establish a connection to the remote SFTP server, and enter SFTP Client view

```
[SwitchA] sftp 192.168.0.1
Input Username: client001
Trying 192.168.0.1 ...
Press CTRL+K to abort
Connected to 192.168.0.1 ...

The Server is not authenticated. Do you continue access it? [Y/N]:y
Do you want to save the server's public key? [Y/N]:y
Enter password:

sftp-client>
```

# Display the current directory of the server, delete the file named "z", and check whether the file is deleted from the directory 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 pubkey1
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 pubkey1
```

```
drwxrwxrwx   1 noone    nogroup         0 Sep 01 06:22 new
-rwxrwxrwx   1 noone    nogroup       225 Sep 01 06:55 pub
```

# Create a directory named "new1", and check whether the new directory is 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 pubkey1
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 the directory "new1" into "new2", and check whether the directory is 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 pubkey1
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 to the local device, and rename the file into "public".

```
sftp-client> get pubkey2 public
Remote  file:/pubkey2 --->  Local file: public
Downloading file successfully ended
```

# Upload the local file "pu" to the server, rename the file into "puk", and check whether the file "pu" is 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 pubkey1
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>
```

```
# Exit SFTP.
```

```
sftp-client> quit
Bye
[SwitchA]
```

74 PASSWORD CONTROL CONFIGURATION

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

- "Password Control Overview" on page 967
- "Password Control Configuration Task List" on page 969
- "Configuring Password Control" on page 969
- "Displaying and Maintaining Password Control" on page 972
- "Password Control Configuration Example" on page 972

Password Control Overview

Password control refers to a set of functions provided by the local authentication server to achieve password security based on predefined policies. The password control functions include the following nine.

1 Minimum password length

With this function, you can set a minimum password length as required for system security. As such, when a user enters a shorter password, the system considers it invalid and prompts the user to re-enter a password.

A password cannot exceed 63 characters.

2 Password aging

Password aging imposes a lifecycle on a user password. After the password aging time expires, the user needs to change the password.

If a user enters an expired password, the system displays an error message and prompts the user to provide a new password and to confirm it by entering it again. The new password must be a valid one and the user must enter exactly the same password when confirming it. Otherwise, the login will fail.

3 Early notice on pending password expiration

When a user logs in, the system checks whether the password will expire in a time equal to or less than the specified period. If so, the system notifies the user of the expiry time and provides a choice for the user to change the password. If the user provides a new password, the system records the new password and the time. If the user chooses to leave the password or the user fails to change it, the system allows the user to log in using the present password until the password expires.

Telnet, SSH, and terminal users can change their passwords by themselves. FTP users, on the contrary, can only have their passwords changed by the administrator.

4 Password history

With this feature enabled, the system maintains certain entries of passwords that a user has used. When a user changes the password, the system checks the new password against the used ones to see whether it was used before and, if so, displays an error message.

You can set the maximum number of history password records for the system to maintain for each user. When the number of history password records exceeds your setting, the latest record will overwrite the earliest one.

5 Login attempt restriction

Limiting the times of entering wrong passwords can effectively prevent malicious password cracking.

Once a user fails to pass authentication, the system adds the user into a blacklist. When a user tries but fails to login for the allowed maximum number of successive authentication attempts, the system may prohibit or allow the user to login, depending on your choice:

- Prohibiting the user from logging into the system until the user is removed from the blacklist.
- Allowing the user to log in and removing the user from the blacklist when the user logs into the system or the blacklist entry times out (the blacklist entry aging time is 20 minutes).
- Prohibiting the user from logging in for a configurable period of time. After this period, the user will be deleted from the blacklist and can log into the system again.

- *A blacklist can contain up to 1,024 entries. A login attempt using a wrong username will undoubtedly fail but the username is not added into the blacklist.*
- *FTP users and virtual terminal line (VTY) users are blacklisted when they fail the authentication.*
- *Users accessing the system through the Console or AUX interface are never blacklisted. This is because the system is unable to obtain the IP addresses of these users and these users are privileged and therefore relatively secure to the system.*

6 Password composition

A password can be a combination of characters from the following four categories:

- Uppercase letters A to Z
- Lowercase letters a to z
- Digits 0 to 9
- 32 special characters including blank space and ~‘!@#$%^&*()_+-={}|[]:";’<>,./.

Depending on the system security requirements, you can set the minimum number of categories a password must contain and the minimum number of characters of each category.

There are four password combination levels: 1, 2, 3, and 4, each representing the number of categories that a password must at least contain. Level 1 means that a password must contain characters of one category, level 2 at least two categories, and so on.

When a user sets or changes the password, the system checks if the password satisfies the composition requirement. If not, the system displays an error message.

7 Password display in the form of a string of *

For the sake of security, the password a user enters is displayed in the form of a string of *.

8 Authentication timeout management

If a user fails to log in within a configurable period of time, the system tears down the connection.

This function applies to Telnet users only.

9 Logging

The system logs all successful password changing events.

Password Control Configuration Task List

Task	Remarks
"Enabling Password Control" on page 970	Required
"Setting Global Password Control Parameters" on page 970	Optional
"Setting Local User Password Control Parameters" on page 971	Optional
"Setting Super Password Control Parameters" on page 971	Optional
"Setting a Local User Password in Interactive Mode" on page 972	Optional

Configuring Password Control

- *Global settings in system view apply to all local user passwords and super passwords.*
- *Settings in local user view apply to the local user password only.*
- *Settings for super passwords apply to super passwords only.*

The above three types of settings have different priorities:

- For local user passwords, the settings in local user view override those in system view unless the former are not provided.

- For super passwords, the settings for super password override those in system view unless the former are not provided.

Enabling Password Control

Among the nine password control functions, you can enable or disable the following four functions as desired:

- Password aging
- Minimum password length
- Password history
- Password composition

You must enable a function for its relevant configurations to take effect.

Follow these steps to enable a password control function:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable a password control function	**password-control** { **aging** | **length** | **history** | **composition** } **enable**	Optional All of the four password control functions are enabled by default.

Setting Global Password Control Parameters

Follow these steps to set global password control parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Set the password aging time	**password-control aging** *aging-time*	Optional 90 days by default
Set the minimum password length	**password-control length** *length*	Optional 10 characters by default
Configure the password composition policy	**password-control composition type-number** *policy-type* [**type-length** *type-length*]	Optional By default, the minimum number of password composition types is 1 and the minimum number of characters of a password composition type is 1 too.
Set the maximum number of history password records for each user	**password-control history** *max-record-num*	Optional 4 by default
Specify the maximum number of login attempts and the action to be taken when a user fails to login after the specified number of attempts	**password-control login-attempt** *login-times* [**exceed** { **lock** | **unlock** | **lock-time** *time* }]	Optional By default, the maximum number of login attempts is 3 and a user failing to login after the specified number of attempts must wait for 120 minutes before trying again.
Set the number of days during which the user is warned of the pending password expiration	**password-control alert-before-expire** *alert-time*	Optional 7 days by default

To do...	Use the command...	Remarks
Set the authentication timeout time	**password-control authentication-timeout** *authentication-timeout*	Optional 60 seconds by default

CAUTION: *Configuration for the action to be taken when a user fails to login after the specified number of attempts takes effect immediately, and can thus affect the users already in the blacklist.*

Setting Local User Password Control Parameters

Follow these steps to set password control parameters for a local user:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a local user and enter local user view	**local-user** *user-name*	-
Configure the password aging time for the local user	**password-control aging** *aging-time*	Optional 90 days by default
Configure the minimum password length for the local user	**password-control length** *length*	Optional 10 characters by default
Configure the password composition policy for the local user	**password-control composition type-number** *policy-type* [**type-length** *type-length*]	Optional By default, the minimum number of password composition types is 1 and the minimum number of characters of a password composition type is 1 too.

Setting Super Password Control Parameters

CLI commands fall into four levels: visit, monitor, system, and manage. Accordingly, login users fall into four levels, each corresponding to a command level. A user of a certain level can only use the commands at that level or lower levels. To switch from a lower user level to a higher one, a user needs to enter a password for authentication. This password is called a super password. For details on super passwords, refer to "Basic Configurations" on page 27.

Follow these steps to set super password control parameters:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Set the password aging time for super passwords	**password-control super aging** *aging-time*	Optional 90 days by default
Configure the minimum length for super passwords	**password-control super length** *length*	Optional 10 characters by default

To do...	Use the command...	Remarks
Configure the password composition policy for super passwords	**password-control super composition type-number** *policy-type* [**type-length** *type-length*]	Optional By default, the minimum number of password composition types is 1 and the minimum number of characters of a password composition type is 1 too.

Setting a Local User Password in Interactive Mode

Follow these steps to set the password for a local user in interactive mode:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create a local user and enter local user view	**local-user** *user-name*	-
Set the password for the local user	**password**	Required By default, no password is set for a local user in interactive mode

Displaying and Maintaining Password Control

To do...	Use the command...	Remarks
Display password control configuration information	**display password-control** [**super**]	Available in any view
Display information about users blacklisted due to authentication failure	**display password-control blacklist** [**user-name** *name* | **ip** *ip-address*]	Available in any view
Delete users from the blacklist	**reset password-control blacklist** [**user-name** *name*]	Available in user view
Clear history password records	**reset password-control history-record** [**user-name** *name* | **super** [**level** *level*]]	Available in user view

The ***reset password-control history-record*** *command can delete the history password records of one or all users even when the password history function is disabled.*

Password Control Configuration Example

Network requirements

The following password control functions are required:

- A user is prohibited from logging in after two successive login failures; the password aging time is 30 days.
- A super password must contain at least three types of the valid characters and the valid characters of each type must not be less than five.
- The password of the local user named test must not be less than six characters and must consist of at least two types of the valid characters, with at least five characters of each type. The password aging time is 20 days.

Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Prohibit the user from logging in after two successive login failures.

```
[Sysname] password-control login-attempt 2 exceed lock
```

# Set the password aging time to 30 days for all passwords.

```
[Sysname] password-control aging 30
```

# Set the minimum number of composition types for super passwords to 3 and the minimum number of characters of each composition type to 5.

```
[Sysname] password-control super composition type-number 3 type-length 5
```

# Configure a super password.

```
[Sysname] super password level 3 simple 11111AAAAAaaaaa
```

# Create a local user named test.

```
[Sysname] local-user test
```

# Set the minimum password length to 6 for the local user.

```
[Sysname-luser-test] password-control length 6
```

# Set the minimum number of password composition types to 2 and the minimum number of characters of each password composition type to 5 for the local user.

```
[Sysname-luser-test] password-control composition type-number 2 type-length 5
```

# Set the password aging time to 20 days for the local user.

```
[Sysname-luser-test] password-control aging 20
```

# Configure the password of the local user.

```
[Sysname-luser-test] password simple 11111#####
```

75 MAC AUTHENTICATION CONFIGURATION

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

- "MAC Authentication Overview" on page 975
- "Related Concepts" on page 975
- "Configuring MAC Authentication" on page 976
- "Displaying and Maintaining MAC Authentication" on page 977
- "MAC Authentication Configuration Example" on page 977

MAC Authentication Overview

MAC authentication provides a way for authenticating users based on ports and MAC addresses, without requiring any client software to be installed on the hosts. MAC authentication uses the MAC address of the user's access device as the authentication user name and password. Once detecting a new MAC address, it initiates the authentication process.

MAC authentication can be performed on a RADIUS (remote authentication dial-In user service) server or locally:

- In RADIUS authentication, the device serves as an RADIUS client. It forwards the detected user MAC address as the user name and password to the RADIUS server for authentication. If the authentication succeeds, the user is allowed to access the network resources.
- In local authentication, the user MAC address must be manually configured on the device as the user name and password.

For details about RADIUS and local authentication, refer to "AAA, RADIUS and HWTACACS Configuration" on page 873.

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 an online user has gone offline. Once detecting that a user becomes offline, the device sends to the RADIUS server a stop accounting notice.
- Quiet timer: Whenever a user fails MAC authentication, the device does not initiate any 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 from accessing 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 simply by the device until the quiet timer expires. This prevents an invalid user from being authenticated repeatedly in a short time.

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.

CAUTION: *For local authentication:*

- *The user name and password of a local user must be the MAC address of the user.*
- *The MAC address to be used as the user name and password of a local user must be in the HHH format and contain only lower-case letters and no* "-".
- *The service type of the local user must be configured as* ***lan-access****.*

Configuration Procedure

Follow these steps to configure centralized 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* **interface** *interface-type interface-number* **mac-authentication**	Required Disabled by default
Specify the ISP domain for MAC authentication	**mac-authentication domain** *isp-name*	Optional The default ISP domain (system) is used by default.
Set the offline detect timer	**mac-authentication timer offline-detect** *offline-detect-value*	Optional Interval of detecting whether the user is offline 300 seconds by default
Set the quiet timer	**mac-authentication timer quiet** *quiet-value*	Optional When user authentication fails, the device will be quiet for a period of time before reinitiating the authentication. One minute by default

To do...	Use the command...	Remarks
Set the server timeout timer	**mac-authentication timer server-timeout** *server-timeout-value*	Optional Timeout timer for the connection to the RADIUS server 100 seconds by default

CAUTION:

- *You can configure MAC authentication on a specific port before global MAC authentication is enabled, but the configuration will not take effect until global MAC authentication is enabled.*
- *MAC authentication and 802.1x cannot be enabled on the same port.*
- *You can neither add a MAC authentication enabled port into an aggregation group, nor enable MAC authentication on a port added into an aggregation group.*

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
Display user connection information	**display connection**	Available in any view
Clear the MAC authentication statistics	**reset mac-authentication statistics**	Available in user view

MAC Authentication Configuration Example

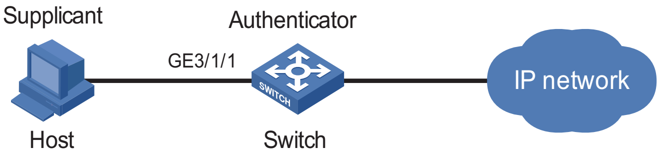
Network requirements

As illustrated in Figure 295, a supplicant is connected to the Ethernet switch through port GigabitEthernet 3/1/1.

- 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 domain aabbcc.net. 00e0fc010101 is used as the user name and password for local authentication.

Network Diagram

Figure 295 Network diagram for MAC authentication



Configuration Procedure

# Add a local user.

```
<Sysname> system-view
[Sysname] local-user 00e0fc010101
[Sysname-luser-00e0fc010101] password simple 00e0fc010101
[Sysname-luser-00e0fc010101] service-type lan-access
[Sysname-luser-00e0fc010101] quit
```

# Configure ISP domain aabbcc.net, and specify to perform local authentication.

```
[Sysname] domain aabbcc.net
[Sysname-isp-aabbcc.net] authentication lan-access local
[Sysname-isp-aabbcc.net] quit
```

# Enable centralized MAC authentication globally.

```
[Sysname] mac-authentication
```

# Enable centralized MAC authentication for port GigabitEthernet 3/1/1.

```
[Sysname] mac-authentication interface GigabitEthernet 3/1/1
```

# Specify the ISP domain for MAC authentication.

```
[Sysname] mac-authentication domain aabbcc.net
```

# Set the centralized MAC authentication timers.

```
[Sysname] mac-authentication timer offline-detect 180
[Sysname] mac-authentication timer quiet 3
```

Verify your configuration

# Display global MAC authentication information.

```
<Sysname> display mac-authentication
MAC address authentication is Enabled.
         Offline detect period is 180s
         Quiet period is 3 minute(s).
         Server response timeout value is 100s
         The max allowed user number is 1024 per slot
         Current user number amounts to 0
         Current domain is aabbcc.net
Silent Mac User info:
         MAC ADDR               From Port           Port Index
Gigabitethernet3/1/1 is link-up
```

76 NAT CONFIGURATION

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

NAT Overview

Introduction to NAT

Network Address Translation (NAT for short) provides a way of translating the IP address in an IP packet header to another IP address. In practice, NAT is primarily designed for private network users to access public networks. This way of using a smaller number of public IP addresses to represent a larger number of private IP addresses can effectively alleviate the depletion of IP addresses.

Private or internal IP addresses refer to IP addresses used in an internal network whereas public or external IP addresses refer to the globally unique IP addresses used on the Internet.

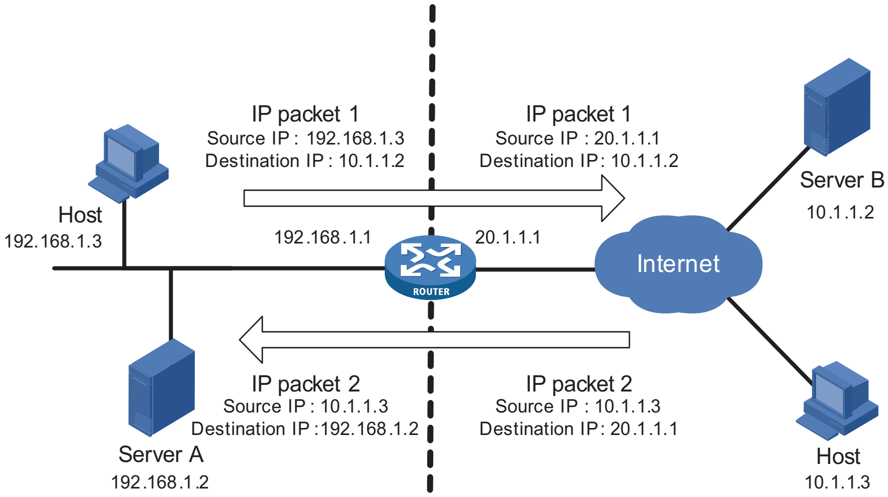
According to RFC 1918, three blocks of IP addresses are reserved for private networks:

- *In Class A: 10.0.0.0 to 10.255.255.255;*
- *In Class B: 172.16.0.0 to 172.31.255.255;*
- *In Class C: 192.168.0.0 to 192.168.255.255;*

The above three ranges of IP addresses are not assigned over the Internet. You can use these IP addresses in enterprises freely without the need for applying them from the ISPs or the registration center.

Figure 296 depicts a basic NAT operation:

Figure 296 A basic NAT operation



- NAT gateway lies between the private network and the public network.
- The internal PC (with source IP address 192.168.1.3) sends an IP packet (IP packet 1) to the external server (with source IP address 10.1.1.2) through the NAT gateway.
- Upon receipt of the packet, the NAT gateway checks the packet header and translates the original private address 192.168.1.3 to a globally unique IP address 20.1.1.1 for routing over the Internet. After that, the gateway forwards the packet and records the mapping between the two addresses in its network address translation table.
- The external server responds the internal PC with an IP packet (IP packet 2 with original destination IP address 20.1.1.1) through the NAT gateway. Upon receipt of the packet, the NAT gateway checks the packet header and looks in its network address translation table for the mapping and replaces the original destination address with the private address 192.168.1.3.

The above NAT operation is transparent to the terminals like the Host and the Server in the above figure. The external server believes that the IP address of the internal PC is 20.1.1.1, and is unaware of the private address 192.168.1.3. As such, NAT hides the private network from the external networks.

Despite the advantage of allowing internal hosts to access external resources and providing privacy, NAT also has the following disadvantages:

- As NAT involves translation of IP addresses, the packet headers that carry these addresses cannot be encrypted. This is also true to the application protocol packets when the contained IP address or port number needs to be translated. For example, you cannot encrypt an FTP connection, or its **port** command cannot work correctly.
- Network debugging becomes more difficult. For example, when a host in a private network tries to attack other networks, it is harder to pinpoint the attacking host as the host IP address has been hidden.
- The influence of NAT on network performance is not obvious when the bandwidth is lower than 1.5 Gbps. The bottleneck in this scenario lies in the

transmission rate. However, when the bandwidth is higher than 1.5 Gbps, NAT could affect the switch performance to a certain extent.

NAT Functionalities

Many-to-many NAT and NAT control

As depicted in Figure 296, when an internal network user accesses an external network, NAT uses an external or public IP address to replace the original internal IP address. In Figure 296, this address is the outbound interface address (a public IP address) of the NAT gateway. This means that all internal hosts use the same external IP address when accessing external networks. In this scenario, only one host is allowed to access external networks at a given time. Hence, it is referred to as "one-to-one NAT".

Another form of NAT solves this problem by allowing the NAT gateway to have multiple public IP addresses. When the first internal host accesses external networks, NAT chooses a public IP address for it, records the mapping between the two addresses and transfers data packets. When the second internal host accesses external networks, a similar process happens, but this time another public IP address is used, and so are the remaining internal hosts. In this way, multiple internal hosts can access the external networks simultaneously. This type of NAT is called "many-to-many NAT".

The number of public IP addresses an NAT gateway has is far less than the number of internal hosts, because not all internal hosts will access the external networks at the same time. The number of necessary public IP addresses should be determined based on the statistics on the number of the hosts that might access external networks during peak time.

In practice, an enterprise may need to allow some internal hosts to access external networks while prohibiting others. This can be achieved through the NAT control mechanism. If a source IP address is among those addresses that have been denied access to external networks, the NAT gateway will not translate this address.

The "many-to-many NAT" can be realized through definition of an address pool whereas NAT control can be achieved through ACLs.

- Address pool: a set of consecutive public IP addresses intended for address translation. The address pool should be configured according to the number of legal IP addresses, the number of internal hosts, and the actual network requirements. The NAT gateway will select an address from the address pool and use it as the source public IP address during address translation.
- NAT control through ACLs: NAT is only applied to the packets that match the ACL rules. This makes the use of NAT more flexible.

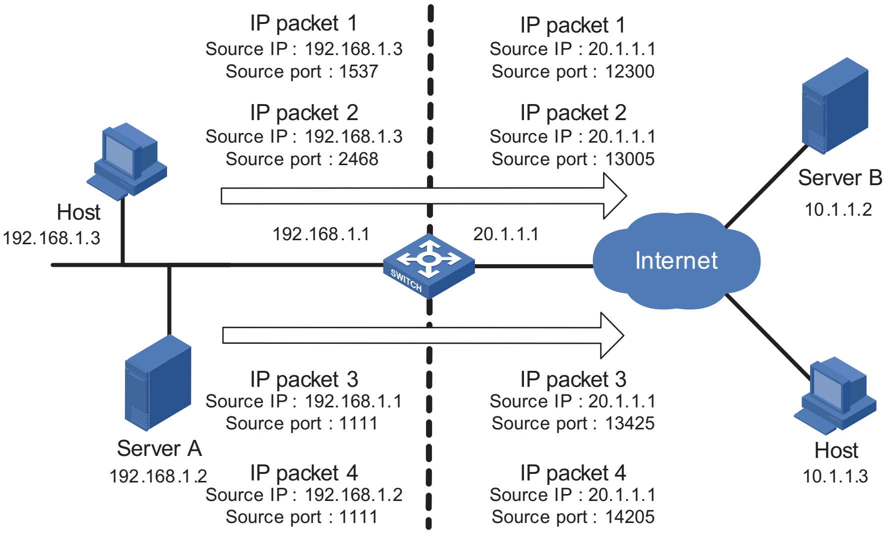
NAPT

Another form of NAT is network address port translation (NAPT for short). NAPT allows multiple internal addresses to be mapped to the same external public IP address, namely "multiple-to-one NAT", or "address multiplexing".

The destination addresses of the packets from different internal hosts are mapped to the same external IP address but with different port numbers. In other words, NAPT maps the combination of a private IP address and a port number to the combination of a public IP address and a port number.

Figure 297 depicts an NAPT process.

Figure 297 An NAPT process



As illustrated in the above figure, four data packets arrive at the NAT gateway. Packets 1 and 2 have the same internal address but different source port numbers. Packets 3 and 4 have different internal addresses but the same source port number. NAPT maps the four data packets to the same external address but with different source port numbers. Therefore, the packets can still be discriminated. When response packets arrive, the NAT gateway can forward them to the corresponding hosts based on the destination address and port numbers.

Internal server

NAT hides the internal network structure, including the identities of internal hosts. However, in practice, external contacts to internal hosts are sometimes also necessary. In this case, you need an internal server, such as a WWW server or an FTP server to provide such services. With NAT, you can deploy an internal server easily and flexibly. For instance, you can use 20.1.1.10 as the WWW server's external address, 20.1.1.11 as the FTP server's external address; or you can even use such address 20.1.1.12:8080 as the WWW server's external address.

Currently, this feature is available on the device. When an external user accesses an internal server, NAT translates the destination address in the request packet to the private IP address of the internal server. When the internal server returns a packet, NAT translates the source address (a private IP address) of the packet into a public IP address.

Easy IP

Easy IP allows the NAT gateway to use the public IP address of an interface as the translated source address for NAT. Besides, the NAT gateway can use ACLs to define the internal IP addresses for NAT.

Support for special protocols

Apart from the basic address translation function, NAT also provides a perfect application layer gateway mechanism that supports various special application protocols without modifying the NAT platform. Because of this, NAT offers high scalability. The special protocols supported by the Switch 8800s include: Internet control message protocol (ICMP), domain name system (DNS), Internet locator service (ILS), and NetBIOS over TCP/IP (NBT).

NAT multiple-instance

This feature allows users from different MPLS VPNs to access external networks through the same outbound interface. It also allows them to have the same internal network address. The process works as follows:

When an MPLS VPN user communicates with an external network, NAT replaces its internal IP address and port number with the NAT gateway's external IP address and port number. It also records the relevant MPLS VPN information, such as the protocol type and router distinguisher (RD for short). When the response packet arrives, the NAT gateway then restores the external IP address and port number to the internal IP address and port number. Additionally, the NAT gateway can identify the users who access the external network. Besides NAT, NAPT also supports multiple-instance.

The multiple-instance feature can also apply to internal servers so that external users can access an internal host of an MPLS VPN. For example, in MPLS VPN1, the host that provides WWW service has an internal address 10.110.1.1. The host can use 202.110.10.20 as an external IP address so that the Internet users can access the WWW service in MPLS VPN1 through this external address.

NAT Configuration Task List

Follow the following steps to configure NAT:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Define an address pool	**nat address-group** *group-number start-address end-address*	Optional Not necessary when the switch has been configured with Easy IP.
Configure address translation	Refer to "Configuring Address Translation" on page 984.	Required
Configure an internal server	Refer to "Configuring Internal Server" on page 985.	Optional
Enable NAT application layer gateway	**nat alg** { **all** | **dns** | **ftp** | **ils** | **nbt** }	Optional Enabled by default Currently, the NAT ALG supports only standard ports for DNS, FTP, ILS, and NBT.
Configure the binding	Refer to "Configuring the Binding" on page 986	Required
Configure NAT log	Refer to "Configuring NAT Log" on page 987	Optional Disabled by default
Configure connection-limit	Refer to "Configuring Connection-limit" on page 990	Optional Disabled by default

- *The addresses in the address pool referenced by NAT must be different from the interface address. Otherwise, the service can be implemented. To use the interface address as the translation address, Easy IP must be used.*

Configuring Address Translation

Introduction to Address Translation

Address translation is implemented by associating an ACL with an address pool (or an interface address in case of Easy IP). This association specifies what packets (defined by ACLs) can use which address (one in the address pool, or the interface address itself) to access the external network. When an internal host needs to send data packets to an external network, the NAT gateway checks the first packet against the ACL to see if it is permitted. If so, NAT chooses an address from the address pool (or the interface address, depending on the association) to perform address translation. This address mapping is recorded in an address translation table so that subsequent packets can be translated directly according to this mapping entry.

For details about ACL, refer to *"ACL Overview" on page 801*.

The configuration for different forms of address translation varies somewhat:

- Easy IP

This feature is implemented using the **nat outbound** *acl-number* command, without the **address-group** keyword specified. When address translation, the NAT gateway directly uses an interface's public IP address as the translated IP address, and uses ACLs to restrict the traffic.

- Many-to-many NAT

You only need to associate an ACL with an address pool, without considering port numbers.

- NAPT

You need to associate an ACL with an address pool, and deal with both IP addresses and port numbers.

- NAT multiple-instance

You need to configure **vpn instance** *vpn-instance-name* in the **rule** of an ACL to specify the MPLS VPN users that need address translation and add a static route to the public network into the routing table of the private network. NAT multiple-instance is supported on Easy IP, Many-to-many NAT, and NAPT.

Configuring Address Translation

Configuring Easy IP

Follow these steps to configure Easy IP:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Enable Easy IP by associating the ACL with the interface IP address	**nat outbound** *acl-number*	Required

Configuring many-to-many NAT

Follow these steps to configure many-to-many NAT:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Enable many-to-many NAT, and associate an ACL with an IP address pool to translate IP address alone	**nat outbound** *acl-number* **address-group** *group-number* **no-pat**	Required

Configuring NAPT

Follow these steps to configure NAPT:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Enable NAPT and associate an ACL with an IP address pool to translate both IP address and port number.	**nat outbound** *acl-number* **address-group** *group-number*	Required

- *For the ACL referenced by NAT, only the source IP address, destination IP address, and VPN instance take effect.*
- *For NO-PAT translation, if multiple NAT rules are configured on a VLAN interface, the device will determine the rule priority based on the ACL numbers bound with the NAT rules and always match the NAT rule with a greater ACL number. The priorities of the rules of an ACL are based on rule number. The smaller the rule number, the higher the priority.*
- *In PAT translation, ACLs are matched according to the "depth-first" order.*

Configuring Internal Server

Introduction to Internal Server

To configure an internal server, you need to map an external IP address and port to the internal server. This is done through the **nat server** command.

Internal server configurations include: external IP address, external port, internal server IP address, internal server port, and internal server protocol type.

If an internal server belongs to an MPLS VPN instance, you should specify the *vpn-instance-name* argument. With this argument not provided, the internal server is considered belonging to a private network.

Configuring an Internal Server

Follow the following steps to configure an internal server:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter VLAN interface view	**interface** *interface-type interface-number*	-
Configure an internal server	**nat server** [**vpn-instance** *vpn-instance-name*] **protocol** *pro-type* **global** *global-address* [*global-port*] **inside** *host-address* [*host-port*]	Use either command
	nat server [**vpn-instance** *vpn-instance-name*] **protocol** *pro-type* **global** *global-address global-port*1 *global-port2* **inside** *host-address1 host-address2 host-port*	

Configuring the Binding

Introduction to Binding

Through the use of the L3+NAT module on a switch, the NAT services can be handled centrally and more efficiently thanks to the quick handling capability of the hardware.

When a VLAN interface is configured with NAT, you can bind the VLAN interface with the NAT virtual interface so that all the packets that pass through the VLAN interface are redirected to the L3+NAT module for handling.

Before configuring the binding, you must configure the VLAN interface first.

Configuration Procedure

Follow these steps to configure the binding:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter NAT service interface view	**interface nat** *number*	-
Configure the binding	**nat binding interface** *interface-type interface-number*	Required Only VLAN interfaces can be bound. A NAT service interface can be bound with multiple NAT-enabled interfaces.

⚠ ***CAUTION:*** *Once bound to the NAT virtual interface, a VLAN interface can no longer serve as the outbound interface of QoS redirection. This is because the* *packets that pass through the VLAN interface have been redirected to the L3+NAT module, causing the QoS redirection function ineffective.*

Configuring NAT Log

Introduction to NAT Log

NAT log is a type of system information generated by the NAT gateway during the IP address translation. NAT log contains such information as the packet's source IP address, source port address, destination IP address, destination port address, translated source IP address, translated source port address and other user operations. The log only traces operations of private network users in accessing an external network, not those in the opposite direction.

As multiple private users share one public IP address when accessing an external network through a NAT gateway, it is hard to identify each of the users. The log function, however, can enhance network security (for supervising purpose) by keeping records of the private network users that access the external network.

Enabling NAT Log Function

Follow these steps to enable NAT log function:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable log function	**nat log enable** [**acl** *acl-number*]	Required Disabled by default
Generate NAT log when establishing a NAT session	**nat log flow-begin**	Required By default, no log is generated when establishing NAT session.
Enable and set the interval for logging active flows	**nat log flow-active** *minutes*	Required Disabled by default

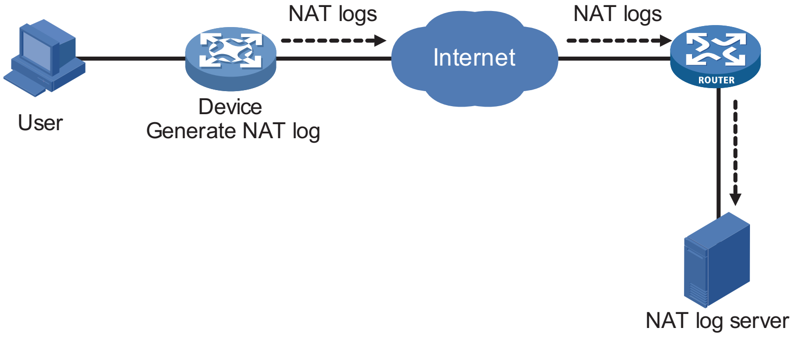
Exporting NAT Logs

NAT logs can be exported in two directions, either to the information center or to the NAT log server.

In the former case, NAT logs are first converted into system logs and exported to the local device's information center. Depending on the configuration of the information system, NAT logs are again exported to their final destination. At most 10 NAT logs can be exported to the information center at one time.

In the latter case, NAT logs are encapsulated into UDP packets and sent to the log server, as shown in Figure 298. The UDP packets may come in several versions, each with different packet formats. Only version 1 is used presently. A UDP packet is composed of a header and several NAT logs.

Figure 298 Export NAT logs



Exporting NAT logs to the information center

Follow these steps to export NAT logs to the information center:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the information source of the log channel	**info-center source USERLOG channel console log level debugging**	-
Export NAT logs to the information center	**userlog nat syslog**	Required NAT logs are exported to the NAT log server by default.

- *Exporting NAT logs to the information center occupies storage space. This approach is recommended when the volume of NAT logs is small.*
- *NAT logs exporting to the information center are prioritized as informational, meaning that they are ordinary information.*
- *For detailed information about data priority, refer to "Information Center Configuration" on page 1111.*

Exporting NAT logs to log server

When exporting NAT logs to the log server in UDP packets, you can configure the following three parameters:

- IP address and UDP port number of the NAT log server. NAT logs cannot be exported successfully without configuring the information center export direction and specifying the log server address.
- Source IP address of NAT logs. This address allows the log server to identify the log source. You are recommended to use the loopback interface address as the source IP address of NAT logs.
- Version number of NAT logs. NAT logs may come in several versions, each with different packet formats. However, the device supports only version 1 currently.

Follow these steps to configure a NAT log server:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Specify the IP address and UDP port number of the NAT log server	**userlog nat export** [**slot** *slot-number*] **host** *ip-address udp-port*	Required
Specify the source IP address of the UDP packet that carries NAT logs	**userlog nat export source-ip** *ip-address*	Optional By default, the source IP address is the interface IP address through which the packet is sent.
Specify the version number of NAT logs	**userlog nat export version** *version-number*	Optional Version 1 is used by default

- *The IP address of the NAT log server must be a valid unicast address.*
- *As for the UDP port number of the log server, you are recommended to use a port number greater than 1024 to avoid conflicts with the system-defined port numbers.*

Configuring User Resource Limit

Introduction to User Resource Limit

User resource limit is a function that defines the maximum number of ordinary users (non-VPN users in an internal network) or VPN users as well as their connections in accessing external network(s). This can help distributing resources more reasonably.

This function only applies to NAPT with its application layer gateway function not enabled.

Configuring User Resource Limit

Follow these steps to configure user resource limit:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set limits for ordinary users or VPN users.	**nat limit** { **public** | **vpn-instance** *vpn-instance-name* } **user-amount** *user-limit* **connection-amount** *connection-limit*	Optional By default, the ordinary users occupy all the system resources.

- *On a newly started system without any configuration, the system resources are completely occupied by ordinary users.*
- *Before a user resource limit is configured for public network users, resources are allocated from those for public network users to a VPN user until the public network user resources are used up.*
- *After the administrator configures a limit on the resources for public network users, resources can be allocated only from the remaining resources to a VPN user until the remaining system resources are used up.*
- *The user resource configuration is performed on a single L3+NAT module, but takes effect to all L3+NAT modules if there are multiple L3+NAT modules.*

Configuring Connection-limit

Introduction to Connection-limit

The connection-limit function allows you to limit user connections in three ways: connection number, connection rate or both. This can avoid the situation where a single user establishes too many connections in a short time as to affect other users in using the network.

Limiting connection number means that when the number of connections initiated by a user reaches a certain upper limit, the user cannot establish new connections. The user must wait (for at least 5 minutes) till the connection number is lower than the upper limit in order to create new connections. This feature applies to VPN users as well.

Limiting connection rate means that a user connection rate cannot exceed a predefined maximum value. This also applies to VPN users.

For the connection-limit function to take effect, you must set a connection-limit policy, bind the policy with the NAT module, and meanwhile activate the connection-limit switch.

CAUTION:

- *For parameters not configured in a connection-limit policy, the global configurations take effect.*
- *For user connections not covered in a connection-limit policy, the global configurations take effect.*

Configuration Procedure

Configuring global connection-limit parameters

Follow these steps to configure global connection-limit parameters

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable connection-limit function	**connection-limit enable**	Required Disabled by default
Configure connection-limit action globally	**connection-limit default action** [**permit** | **deny**]	Optional User connections are not counted and limited by default.
Configure connection number limits globally	**connection-limit default amount upper-limit** *max-amount*	Optional 200 by default
Set the maximum connection rate globally	**connection-limit default rate max-rate** *max-rate*	Optional 100 by default

Configuring connection-limit policy

Follow these steps to configure a connection-limit policy:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Create or edit a connection-limit policy and enter the connection-limit policy view	**connection-limit policy** *policy-number*	Required
Configure the rules of connection-limit	**limit** *limit-index* **source** *ip-address* [**vpn-instance** *vpn-instance-name*] { **amount** *max-amount* | **rate** } *	Required
Set connection-limit mode	**limit mode** { **all | amount | rate** }	Optional By default, both the number and rate of user connections are limited.
Set the maximum connection rate in a policy	**limit rate** *max-rate*	Optional 100 by default

Binding a connection-limit policy to a NAT module

Follow these steps to bind a connection-limit policy to a NAT module

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Bind a connection-limit policy to the NAT module	**nat connection-limit-policy** *policy-number*	Required

CAUTION:

- *A NAT module limits user connections based on the policy bound to it. Each NAT module can be bound with one policy only.*
- *The global connection-limit configuration does not take effect until you bind the connection-limit policy with the NAT module.*
- *If multiple NAT modules exist in the system, the connection limit policy applies to all these NAT modules.*
- *A connection limit policy does not take effect in NO-PAT translation.*

Displaying and Maintaining NAT

To do...	Use the command...	Remarks
Display information about the NAT address pool	**display nat address-group**	Available in any view
Display configurations about all forms of NAT	**display nat all**	Available in any view
Display the connection-limit information	**display nat connection-limit** { **all | ip** *user-ip* [**vpn-instance** *vpn-instance-name*] }	Available in any view
Display the address translation configuration	**display nat outbound**	Available in any view
Display the internal server information	**display nat server**	Available in any view

To do...	Use the command...	Remarks
Display the information about active connections	**display nat session slot** *slot-number* **protocol** { **tcp | udp** } [**vpn-instance** *vpn-instance-name*] **source** { **global** *global-address global-port* | **inside** *inside-address inside-port* } **destination** *dst-address destination-port*	Available in any view
Display NAT statistics	**display nat statistics slot** *slot-number*	Available in any view
Display information about the connection-limit policy	**display connection-limit policy** { *policy-number* | **all** }	Available in any view
Display NAT log information	**display nat log**	Available in any view
Display information about the resource allocation and utilization	**display nat limit** { **all | public | vpn-instance** *vpn-instance--name* }	Available in any view
Display NAT log configuration and statistics	**display userlog export slot** *slot-numbe*r	Available in any view
Clears the records in the NAT log buffer	**reset userlog nat logbuffer slot** *slot-number*	Available in user view
Clears NAT log statistics	**reset userlog export slot** *slot-number*	Available in user view
Clears the address translation mapping table in the memory and release the memory dynamically allocated for storing the mapping table	**reset nat session slot** *slot-number*	Available in user view

Clearing the NAT log buffer implies loss of all NAT logs. In general, you are not recommended to use this command.

NAT Configuration Example

NAT Configuration Example

Network requirements

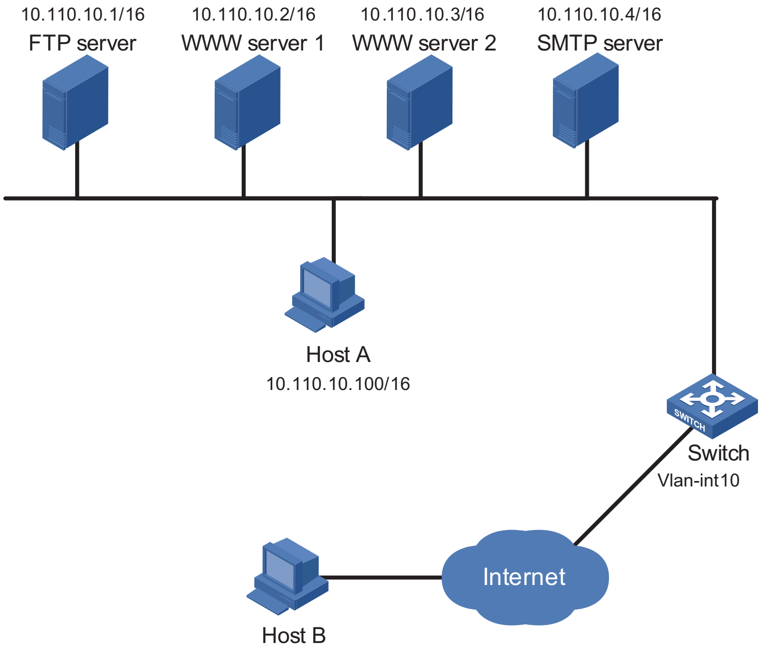
As illustrated in Figure 299, a company accesses the Internet through VLAN 10 of the NAT-enabled device. The company provides two WWW servers, one FTP server, and one SMTP server for external users to access. The internal network address is 10.110.0.0/16. The internal address for the FTP server is 10.110.10.1, for the WWW server 1 is 10.110.10.2, for the WWW server 2 is 10.110.10.3, and for the SMTP server 10.110.10.4. The company wants to provide a unified IP address to external users. Specifically, the company has the following requirements:

- The internal users in subnet 10.110.10.0/24 can access the Internet, while users in other network segments cannot.
- External PCs can access an internal server.
- The company has 6 legal IP addresses ranging from 202.38.160.100/24 to 202.38.160.105/24. Address 202.38.160.100/24 is used as the one for external access and port 8080 is used for WWW server 2.

- Configure a connection-limit policy and bind it to the NAT module. Configure the upper limit of connections as 1000 (based on the source address) respectively, which means the number of connections initiated from internal user cannot exceed 1000.

Network diagram

Figure 299 NAT network diagram



Configuration procedure

# Configure an address pool and an ACL.

```
<Switch> system-view
[Switch] nat address-group 1 202.38.160.101 202.38.160.103
[Switch] acl number 2001
[Switch-acl-basic-2001] rule permit source 10.110.10.0 0.0.0.255
[Switch-acl-basic-2001] quit
```

# Apply NAT to 10.110.10.1 only

```
[Switch] vlan 10
[Switch-vlan10] port Ethernet 1/1/1
[Switch-vlan10] quit
[Switch] interface vlan-interface 10
[Switch-Vlan-interface10] nat outbound 2001 address-group 1
[Switch-Vlan-interface10] quit
[Switch] interface nat 2/0/1
[Switch-NAT2/0/1]nat binding interface vlan-interface 10
```

# Configure the internal FTP server.

```
[Switch-Vlan-interface10] nat server protocol tcp global 202.38.160.
100 8021 inside 10.110.10.1 ftp
```

# Configure the internal WWW server 1.

```
[Switch-Vlan-interface10] nat server protocol tcp global 202.38.160.
100 8081 inside 10.110.10.2 www
```

# Configure the internal WWW server 2.

```
[Switch-Vlan-interface10] nat server protocol tcp global 202.38.160.
100 8080 inside 10.110.10.3 www
```

# Configure the internal SMTP server.

```
[Switch-Vlan-interface10] nat server protocol tcp global 202.38.160.
100 8025 inside 10.110.10.4 smtp
[Switch-Vlan-interface10] quit
```

# Enable the connection-limit function.

```
[[Switch] connection-limit enable
```

# Configure a connection-limit policy and rules.

```
[Switch] connection-limit policy 1
[Switch-connection-limit-policy-1] limit mode amount
[Switch-connection-limit-policy-1] limit 1 source 10.110.10.1 amount 1000
[Switch-connection-limit-policy-1] quit
```

# Bind the connection-limit policy with the NAT module.

```
[Switch] nat connection-limit-policy 1
```

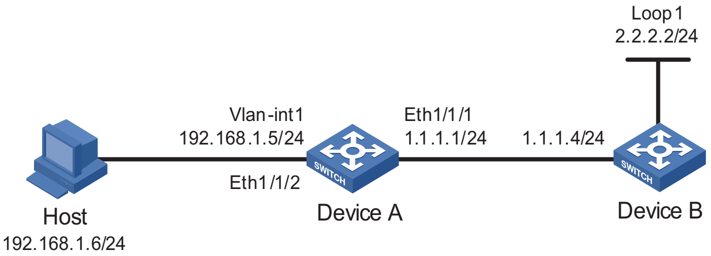
Exporting NAT Logs to the Information Center

Network requirements

- A host in the private network accesses Device B in the public network through Device A, which is enabled with NAT;
- Device A sends NAT logs to the information center in the form of system logs;
- You can view the records on the information center to supervise the private network users.

Network diagram

Figure 300 Export NAT logs to information center



Configuration procedure

The following only lists configurations pertinent to NAT logs. Configurations regarding the IP addresses of the devices and NAT function are omitted here.

# Specify to export the NAT logs of Device A to the information center.

```
<Sysname> system-view
[Sysname] userlog nat syslog
```

# Enable the NAT log function on Device A.

```
[Sysname] nat log enable
```

# View the log buffer to monitor access records.

```
[Sysname] quit
<Sysname> dir
Directory of cf:/
   0   -rw-  16850028  Aug 07 2009 04:02:42   mainpack.bin
   1   drw-         -  Aug 07 2005 05:13:48   logfile
   2   -rw-      1747  Aug 07 2009 04:05:38   config.cfg
   3   -rw-    524288  Aug 13 2009 01:27:40   basicbtm.bin
   4   -rw-    524288  Aug 13 2009 01:27:40   extendbtm.bin
249852 KB total (232072 KB free)
File system type of cf: FAT32
<Sysname> cd logfile
<Sysname> more logfile.log
......omitted......
%@250005%Jul  7 04:20:04:72 2005 Sysname USERLOG/7/NAT:
 ICMP; 192.168.1.6:768--->1.1.1.1:12288; 2.2.2.2:768;
 [2005/07/07 04:20:03-0000/00/00 00:00:00];
 Operator 8: Data flow created
%@250006%Jul  7 04:20:10:72 2005 Sysname USERLOG/7/NAT:
 ICMP; 192.168.1.6:768--->1.1.1.1:12288; 2.2.2.2:768;
 [2005/07/07 04:20:03-2005/07/07 04:20:09];
 Operator 1: Normal over
%@250007%Jul  7 04:20:30:72 2005 Sysname USERLOG/7/NAT:
 ICMP; 192.168.1.6:768--->1.1.1.1:12288; 2.2.2.2:768;
 [2005/07/07 04:20:29-0000/00/00 00:00:00];
 Operator 8: Data flow created
......omitted......
```

Apart from NAT logs, the log file includes other system logs. The following table shows the description of NAT logs:

Field	Description
ICMP	ICMP
192.168.1.6:768	Source IP address and port number before translation
1.1.1.1:12288	Source IP address and port number after translation
2.2.2.2:768	Destination IP address and port number
2005/07/07 04:20:03	Start time of the NAT session (In this example, the time displayed is the device’s system time. When the logs are exported in UDP packet, the UDP packet records the interval in seconds between the current system time and Greenwich time 0 AM, Jan 1st, 1970. The log server, based on its own system time, converts this interval and exports it.
2005/07/07 04:20:29	
2005/07/07 04:20:09	End time of the NAT session
0000/00/00 00:00:00	0000/00/00 00:00:00 means that this time is uncertain.

Field	Description
Operator	Reasons for generating NAT logs come from: ■ Aged for reset or config-change" refers to logs generated due to configuration change or manual session deletion; ■ Aged for no-pat of NAT" refers to logs generated when the no-pat session ages; ■ Active data flow timeout" refers to logs generated when the duration of NAT session exceeds the active data flow time; ■ Data flow created" refers to logs generated when a NAT session is established; ■ Normal over" refers to logs generated when the session is aged out.

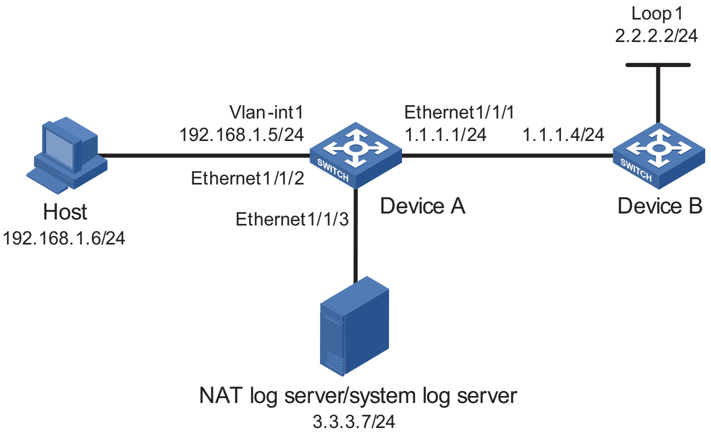
Exporting NAT logs to Log Server

Network requirements

- A PC in the private network accesses Device B on the public network through Device A, which is enabled with NAT.
- Device A sends NAT logs to the information center in UDP packets;

Network diagram

Figure 301 Export NAT log to log server



Configuration procedure

The following only lists configurations pertinent to NAT logs. Configurations regarding the IP addresses of the devices and NAT function are omitted here.

# Specify to export the NAT logs of Device A to the NAT log server.

```
<Sysname> system-view
[Sysname] userlog nat export host 3.3.3.7 9021
```

# Set the source IP address of NAT log packets for Device A to 9.9.9.9

```
[Sysname] userlog nat export source-ip 9.9.9.9
```

# Enable the NAT log function on Device A.

```
[Sysname] nat log enable
```

You must run XLog on the NAT log server or the system log server to view NAT log information.

Troubleshooting NAT

Symptom 1: Abnormal Translation of IP Addresses

Solution: Enable debugging for NAT. Try to locate the problem based on the debugging display. Use other commands, if necessary, to further identify the problem. Pay special attention to the translated source address and ensure that this address is the address that you intend to change to. If not, there may be an address pool bug. Also ensure a route is available between the destination network and the address pool segment. Be aware of the possible effects that the firewall or the ACLs have to NAT, and also note the route configurations.

Symptom 2: Internal Server Functions Abnormally

Solution: Check whether the internal server host is properly configured; whether the router is correctly configured with respect to the internal server parameters, such as the internal server IP address. It is also possible that the firewall that has denied external access to the internal network. You can use the **display acl** command to verify this.

77 DEVICE MANAGEMENT

File names in this document comply with the following rules:

- Path + file name (namely, a full file name): File on a specified path. A full file name consists of 1 to 135 characters.
- File name" (namely, only a file name without a path): File on the current working path. The file name without a path consists of 1 to 91 characters.

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

- "Device Management Overview" on page 999
- "Configuring Device Management" on page 999
- "Displaying and Maintaining Device Management Configuration" on page 1002
- "Device Management Configuration Example" on page 1003

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.

Currently, the following device management functions are available:

- Rebooting a device
- Rebooting a device at a specified time
- Specifying a Boot ROM file for the next device reboot
- Upgrading a Boot ROM file
- Configuring temperature alarm thresholds for a card
- Clearing the 16-bit Interface Indexes Not Used in the Current System

Configuring Device Management

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. You can set a time at which the device can automatically reboot. 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 card	**reboot** [**slot** *slot-number*]	Optional Available in user view.
Enable the scheduled reboot function 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 and specify a reboot waiting time	**schedule reboot delay** { *hh:mm* | *mm* }	

CAUTION:

- *The precision of the rebooting timer is 1 minute. One minute before the rebooting time, the device will prompt a specific reboot time and date and will reboot one minute after this reboot time.*
- *The execution of the **reboot**, **schedule reboot at**, and **schedule reboot delay** commands can reboot a device. As a result, the ongoing services will be interrupted. Be careful to use these commands.*
- *If a primary boot file fails or does not exist, the device cannot be rebooted with this command. In this case, you can re-specify a primary boot file to reboot the device, or you can power off the device then power it on and the system automatically uses the secondary boot file to restart the device.*
- *If you are performing file operations when the device is to be rebooted, the system removes the reboot operation for the sake of security.*

Specifying a Boot ROM File for the Next Device Boot

A Boot ROM file is an application file used to boot the device. When multiple Boot ROM files are available on the storage device, you can specify a file for the next device boot by executing the following command.

Follow these steps to specify a file for the next device boot:

To do...	Use the command...	Remarks
Specify a Boot ROM file on a card	**boot-loader file** *file-url* **slot** *slot-number* { **main** | **backup** }	Required Available in user view.

CAUTION: *The file for the next device boot must be saved under the root directory of the device (for a device supporting storage device partition, the file must be saved on the first partition). You can copy or move a file to change the path of it to the root directory.*

Upgrading Boot ROM

During the operation of the device, you can use Boot ROM in the storage device to upgrade those that are running on the device.

Follow these steps to upgrade Boot ROM:

To do...	Use the command...	Remarks
Upgrade the Boot ROM program on a card(s)	**bootrom update file** *file-url* **slot** *slot-number-list*	Required Available in user view

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 30 seconds by default.

Configuring Temperature Alarm Thresholds for a Card

You can set temperature alarm thresholds for a card by using the following command. When the temperature of a card exceeds the threshold, the device will generate alarm signals.

Follow these steps to configure temperature alarm thresholds for a card:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure temperature alarm thresholds for a card	**temperature-limit** *slot-number lower-value upper-value*	Optional The lower value is 10 and the upper level is 70 by default.

Configuring the Load Mode for the Active and Standby Cards

Two load modes are available between active and standby cards of a device: load sharing (BALANCE) and active/standby mode (SINGLE). You can use the **xbar** command to configure the load mode for a card.

Follow these steps to configure the load mode for active and standby cards:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the load mode for main and backup cards	**xbar** { **load-balance** | **load-single** }	Optional The active and standby cards work in the active/standby mode by default.

CAUTION:

- *Only when both the active module and the standby module are in the slot can the load sharing mode be valid; otherwise, even if the load sharing mode is configured the active module will automatically switch to the active/standby mode.*

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 16bit interface index when a card or logical interface is removed.

If you repeatedly insert and remove different subcards or interface cards to create or delete a large amount of logical interface, the interface indexes will be used up, which will result in interface creation failures. To avoid such a case, you can clear all 16bit 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 the step below to clear the 16bit interface indexes not used in the current system:

To do...	Use the command...	Remarks
Clear the 16-bit interface indexes saved but not used in the current system	**reset unused porttag**	Required

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.*

Displaying and Maintaining Device Management Configuration

To do...	Use the command...	Remarks
Display the Boot ROM file used for the next boot	**display boot-loader** [**slot** *slot-number*]	Available in any view
Display the statistics of the CPU usage	**display cpu-usage** [*number* [[*offset*] [**verbose**] [**slave** | **slot** *slot-number*] [**from-device**]] | **slave** | **slot** *slot-number*]	Available in any view
Display information about a specified device on the switch	**display device** [**cf-card**] [[**shelf** *shelf-number*] [**frame** *frame-number*] [**slot** *slot-number* [**subslot** *subslot-number*]] | **verbose**]	Available in any view
Display manufacture information of the device	**display device manuinfo** [**slot** *slot-number*]	Available in any view
Display the temperature information of devices	**display environment**	Available in any view
Display the operating state of fans in a device	**display fan** [*fan-id*]	Available in any view

To do...	Use the command...	Remarks
Display the usage of the memory of a device	**display memory** [**slave** | **slot** *slot-number*]	Available in any view
Display the power state of a device	**display power** [*power-id*]	Available in any view
Display the reboot time of a device	**display schedule reboot**	Available in any view
Display the load mode of the current active and standby cards	**display xbar**	Available in any view

Device Management Configuration Example

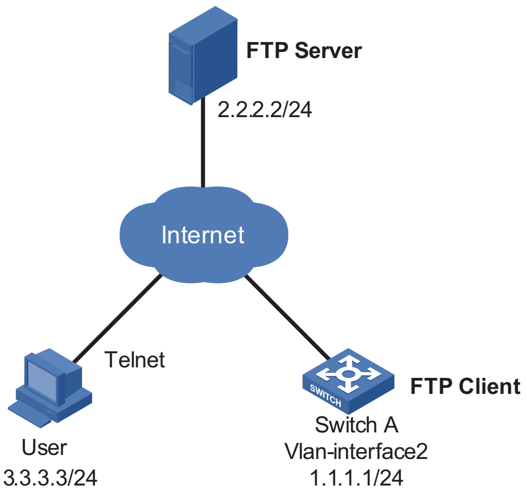
Remote Upgrade Configuration Example

Network requirements

- Switch A serves as the FTP Client. The aaa.app program and the boot.app program are both saved under the aaa directory of the FTP Server.
- The IP address of Vlan-interface 2 on Switch A is 1.1.1.1/24, the IP address of the FTP Server is 2.2.2.2/24, and the IP address of User is 3.3.3.3/24.
- A route exists between the FTP server, Switch A and User.
- User can log in to Switch A via Telnet to perform operations on Switch A (that is, download the application program from FTP Server and remotely upgrade Switch A through command lines).

Network diagram

Figure 302 Network diagram for remote upgrade



Configuration procedure

1 Configure the IP address of each device and a route between FTP server and Switch A, Switch A and User.

Configure the IP address of each device as shown in Figure 302

Configure a route between FTP server and Switch A, Switch A and User. The configuration procedure is omitted here.

2. Configure the username and password on the FTP server.

# Set the FTP username to aaa and password to hello and configure the user to have access to the aaa directory. The configuration procedure is omitted here.

3. Telnet from User to Switch A.

Perform the operation as needed. The procedure is omitted.

4. Configuration on Switch A

CAUTION: *If the size of the Flash on the device is not large enough, delete the original application programs from the Flash before downloading.*

# Enter the following command in user view to log in to FTP Server after telnetting to Switch A.

```
<Sysname> ftp 2.2.2.2
Trying ...
Press CTRL+K to abort
Connected.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(none): aaa
331 Give me your password, please
Password:
230 Logged in successfully
[ftp]
```

# Download the aaa.app and boot.app programs on FTP Server to the Flash of Switch A.

```
[ftp] get aaa.app
[ftp] get boot.app
```

# Terminate the FTP connection and return to user view.

```
[ftp] quit
<Sysname>
```

# Enter system view.

```
<Sysname> system-view
```

# Upgrade the Boot ROM file of the Fabric.

```
<Sysname> bootrom update file boot.app slot 0
```

# Specify the application program for the next boot on Fabric 0.

```
<Sysname> boot-loader file aaa.app slot 0 main
```

# Reboot the device. The application program is upgraded now.

```
<Sysname> reboot
```

78 POE CONFIGURATION

PoE Overview

Introduction to PoE

Power over Ethernet (PoE) means that power sourcing equipment (PSE) supplies power to powered devices (PDs) such as IP telephone, wireless LAN access point, and web camera from Ethernet ports through twisted pair cables.

A PoE device can provide 48 VDC power to its PDs and provide power supply monitoring and PD priority management.

Among the interface cards for Switch 8800 Family, the 3C17528 and 3C17532 interface cards support PoE.

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, 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 is an entity used to manage PoE for ports on a card or subcard. The PSEs on cards/subcards manage the PoE interfaces on their own cards/subcards independently. Through the PoE interface cables, a PSE detects PDs, classifies them, supplies power to them, and stops supplying power to a PD when it detects that the PD is removed.

An Ethernet interface with the PoE capability is called a PoE interface. Currently, a PoE interface can be an FE (Fast Ethernet) or GE (Gigabit Ethernet) interface.

- PD

A PD is a device accepting power from a 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 unit 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 Power" on page 1006	Required
"Configuring a PSE" on page 1007	Required
"Configuring a PoE Interface" on page 1007	Required
"Configuring PoE Power Management" on page 1009	Optional
"Configuring PoE Monitoring" on page 1011	Optional
"Enabling the PSE to Detect Nonstandard PDs" on page 1012	Optional
"Displaying and Maintaining PoE" on page 1012	Optional

- *When the PoE power is shut down or unavailable, all PoE related configuration commands will fail.*
- *Turning off the PoE power during the startup of the device might result in the failure to restore the PoE configuration.*

Configuring the PoE Power

The maximum PoE power refers to the maximum power that the device can provide for all PSEs. To avoid a power failure to the PSE owing to overload, the power consumption of all PSEs should not exceed the maximum power of the device.

Follow these steps to configure the PoE power:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the maximum PoE power	**poe power max-value** *max-power*	Optional The default maximum PoE power is 1125 for 220 VAC input and 562 for 110 VAC input.

- *In the case of 220 VAC input, when the switch uses single or dual power supply modules, you can set at most 2250 W as the maximum PoE power; when the switch uses triple power modules, you can set at most 4500 W as the maximum PoE power.*
- *In the case of 110 VAC input, when the switch uses single or dual power supply modules, you can set at most 1125 W as the maximum PoE power; when the switch uses triple power modules, you can set at most 2250 W as the maximum PoE power.*

Configuring a PSE

Follow these steps to configure a PSE:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable PoE for a PSE	**poe enable pse** *pse-id*	Required By default, PoE is disabled for a PSE.
Configure the maximum PoE power of a PSE	**poe max-power** *max-power* **pse** *pse-id*	Optional By default, the maximum PoE power of a PSE is 806 W.

- *When the remaining power of the PoE system is lower than the maximum power of the PSE, PoE is disabled for the PSE.*
- *The maximum power of the PSE must be greater than or equal to the sum of maximum power of all critical PoE interfaces on the PSE so as to guarantee the power supply to these PoE interfaces.*
- *The relation between the ID and slot number of a PSE is: PSE ID = SlotNo x 3 + 1.*

Configuring a PoE Interface

You can configure a PoE interface in either of the following two ways:

- Adopt the command line.
- Configure a PoE configuration file and apply 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.*

For the Switch 8800s, PoE interfaces can only use signal cables to supply power.

Configuring a PoE Interface Through the Command Line

Follow these steps to configure 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 By default, PoE is disabled on the PoE interface.
Configure the maximum power for the PoE interface	**poe max-power** *max-power*	Optional By default, the maximum power on the PoE interface is 15,400 milliwatts.

To do...	Use the command...	Remarks
Configure the PoE mode for the PoE interface	**poe mode signal**	Optional By default, the PoE mode is **signal** (power over signal cables).
Configure the PoE priority for the PoE interface	**poe priority** { **critical** | **high** | **low** }	Optional By default, the priority is **low**.
Configure a description for the PD connected to the PoE interface	**poe pd-description** *string*	Required By default, no description is configured.

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 common 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
Enable PoE for the PoE interface		**poe enable**	Required By default, PoE is disabled on a PoE interface.
Configure the maximum power for the PoE interface		**poe max-power** *max-power*	Optional By default, the maximum power on the PoE interface is 15,400 milliwatts.
Configure the PoE mode for the PoE interface		**poe mode signal**	Optional By default, the PoE mode is **signal** (power over signal cables).
Configure the PoE priority for the PoE interface		**poe priority** { **critical** | **high** | **low** }	Optional By default, the priority is **low**.
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:

- *Before you can configure another PoE configuration file on a POE interface, you should first remove the original PoE configuration file applied to the PoE interface; otherwise, your configuration will fail.*
- *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.*
- *You must use the same mode (command line or PoE configuration file) to configure the* ***poe max-power*** *and* ***poe priority*** *commands.*

Configuring PoE Power Management

PoE power management involves PSE power management and PD power management.

Configuring PSE Power Management

Where the maximum PoE power may be lower than the sum of the maximum power required by all PSEs, PSE power management is applied to guarantee power supply to important PSE. Where the maximum PoE power of the device is higher than the sum of the maximum power required by all PSEs, it is unnecessary to enable PSE power management.

Power supply to the PSE is subject to PSE power management policies.

When the PoE power supplies power to the PSE,

- By default, no power will be supplied to new PSE when the PoE power is overloaded.
- Under the control of a priority policy, the PSE with a lower priority is first disconnected to guarantee the power supply to the new PSE with a higher priority when the PoE power is overloaded.

The power priority levels of PSE include critical, high and low in descending order.

If the guaranteed remaining PoE power (guaranteed maximum PoE power - power allocated to the critical PSE, regardless of whether PoE is enabled for the PSE) is lower than the maximum power of the PSE, you will fail to set the power priority of the PSE to **critical**. Otherwise, you can succeed in setting the power priority to **critical**, and this PSE will preempt the power of the PSE with a lower priority level. In the latter case, the PSE whose power is preempted will be disconnected, but its configuration will remain unchanged. After you change the priority of the PSE from **critical** to a lower level, other PSEs will have an opportunity of seizing power.

Configuration prerequisites

Enable PoE for the PSE.

Configuration procedure

Follow these steps to configure PSE power management:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Configure the power priority for the PSE	**poe priority** { **critical** | **high** | **low** } **pse** *pse-id*	Optional By default, the power priority level of the PSE is **low**.
Configure a PSE power management priority policy	**poe pse-policy priority**	Optional By default, no PSE power management priority policy is configured.

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.

If the guaranteed remaining PSE power (maximum PSE power - power to be 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 seizing power.

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 approach. 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 PoE Monitoring

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 PoE Power Monitoring

Only an external PoE power supports this feature.

Follow these steps to configure PoE power monitoring:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure an AC input under-voltage threshold for the PoE power	**poe-power input-threshold lower** *value*	Optional The default is: 181 for 220 VAC input 90 for 110 VAC input
Configure an AC input over-voltage threshold for the PoE power	**poe-power input-threshold upper** *value*	Optional The default is: 264 for 220 VAC input 132 for 110 VAC input
Configure a DC output under-voltage threshold for the PoE power	**poe-power output-threshold lower** *value*	Optional The default is 45.
Configure a DC output over-voltage threshold for the PoE power	**poe-power output-threshold upper** *value*	Optional The default is 57.

CAUTION: *The under-voltage threshold should be less than the over-voltage threshold.*

Configuring a Power Alarm Threshold for a PSE

Follow these steps to configure a power alarm threshold for a PSE:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do...	Use the command...	Remarks
Configure a power alarm threshold for a PSE	**poe utilization-threshold** *utilization-threshold-value* **pse** *pse-id*	Optional By default, the power alarm threshold for a PSE is 80%.

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 By default, the PSE is disabled from detecting nonstandard PDs.

Displaying and Maintaining PoE

To do...	Use the command...
Display the mapping between ID, module, and slot of all PSEs.	**display poe device**
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 power information of the PoE power and all PSEs	**display poe power-usage**
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 information of the PoE power	**display poe-power**
Display the state information of the AC input power	**display poe-power ac-input state**
Display the alarm information of the PoE power	**display poe-power alarm**
Display the state information of the DC output power	**display poe-power dc-output state**
Display the parameter values of the DC output power	**display poe-power dc-output value**
Display the status information of the PoE power	**display poe-power status**
Display the information of the monitoring module of the PoE power	**display poe-power supervision-module**
Display the switch information of the PoE power	**display poe-power switch state**

To do...	Use the command...
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*

The **display** commands are available in any view.

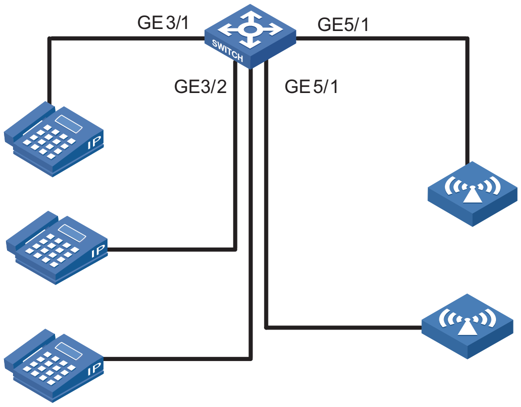
PoE Configuration Example

Network requirements

- The device is equipped with two PoE-supporting cards, which are inserted in slot 3 and slot 5 respectively.
- Allocate 400 watts to the card in slot 3 and full power to the card in slot 5 to guarantee normal power supplying to all PSEs.
- GigabitEthernet3/1/1 and GigabitEthernet3/1/2 are connected to IP telephones.
- GigabitEthernet5/1/1 and GigabitEthernet5/1/2 are connected to access point (AP) devices.
- The power priority of GigabitEthernet3/1/2 is critical.
- The power of the AP device connected to GigabitEthernet5/1/1 cannot exceed 9,000 milliwatts.

Network diagram

Figure 303 Network diagram for PoE



Configuration procedure

# Enable PoE for the PSE.

```
<Sysname> system-view
[Sysname] poe enable pse 10
[Sysname] poe enable pse 16
```

# Set the maximum power for the card in slot 3 to 400 watts. You do not need to configure the maximum power for the card in slot 5 because it is full by default.

```
[Sysname] poe max-power 400 pse 10
```

# Enable PoE on GigabitEthernet3/1/1, GigabitEthernet3/1/2, GigabitEthernet5/1/1, and GigabitEthernet5/1/2.

```
<Sysname> system-view
[Sysname] interface gigabitEthernet 3/1/1
[Sysname-GigabitEthernet3/1/1] poe enable
[Sysname-GigabitEthernet3/1/1] quit
[Sysname] interface gigabitEthernet 3/1/2
[Sysname-GigabitEthernet3/1/2] poe enable
[Sysname-GigabitEthernet3/1/2] quit
[Sysname] interface gigabitEthernet 5/1/1
[Sysname-GigabitEthernet5/1/1] poe enable
[Sysname-GigabitEthernet5/1/1] quit
[Sysname] interface gigabitEthernet 5/1/2
[Sysname-GigabitEthernet5/1/2] poe enable
[Sysname-GigabitEthernet5/1/2] quit
```

# Set the power priority level of GigabitEthernet3/1/2 to critical.

```
<Sysname> system-view
[Sysname] interface gigabitEthernet 3/1/2
[Sysname-GigabitEthernet3/1/2] poe priority critical
[Sysname-GigabitEthernet3/1/1] quit
```

# Set the maximum power of GigabitEthernet5/1/1 to 9,000 milliwatts.

```
[Sysname] interface gigabitEthernet 5/1/1
[Sysname-GigabitEthernet5/1/1] poe max-power 9000
```

After the configuration takes effect, the IP phone and AR device are powered and can work normally.

Troubleshooting PoE

Symptom: Setting 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 former 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 latter case, you should first remove the priority already configured.

Symptom: 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.

- A PoE configuration file is already applied to the PoE interface.

Solution:

- In case 1, you can solve the problem by removing the original configurations of those configurations.
- In case 2, you need to need to modify some configurations in the PoE configuration file.
- In case 3, you need to remove the application of the undesired PoE configuration file to the PoE interface.

Symptom: 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.

79 SYSTEM MAINTENANCE AND DEBUGGING

When maintaining and debugging the system, go to these sections for information you are interested in:

- "System Maintaining and Debugging Overview" on page 1017
- "System Maintaining and Debugging" on page 1019
- "System Maintaining Example" on page 1020

System Maintaining and Debugging Overview

Introduction to System Maintaining and Debugging

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. Display related statistics.

Output of the **ping** command includes:

- Information on the destination's responses towards each ICMP echo request, if the source device has received the 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), and the response time.
- If within the period set by the timeout timer, the destination device has not received the ICMP response, it displays the prompt information.
- The **ping** command can apply to the destination's name or IP address. If the destination's name is unknown, the prompt information is displayed.
- The statistics during the ping operation, which include number of packets sent, number of echo reply messages received, percentage of messages not received, the minimum, average, and maximum response time.

The tracert command

By using the **tracert** command, you can trace the switches 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 switch that first receives the packet; the value of TTL decreases by 1 by each hop) 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 switch.
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 switch.
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 switches that a packet traverses from the source device to the destination device.

Introduction to System Debugging

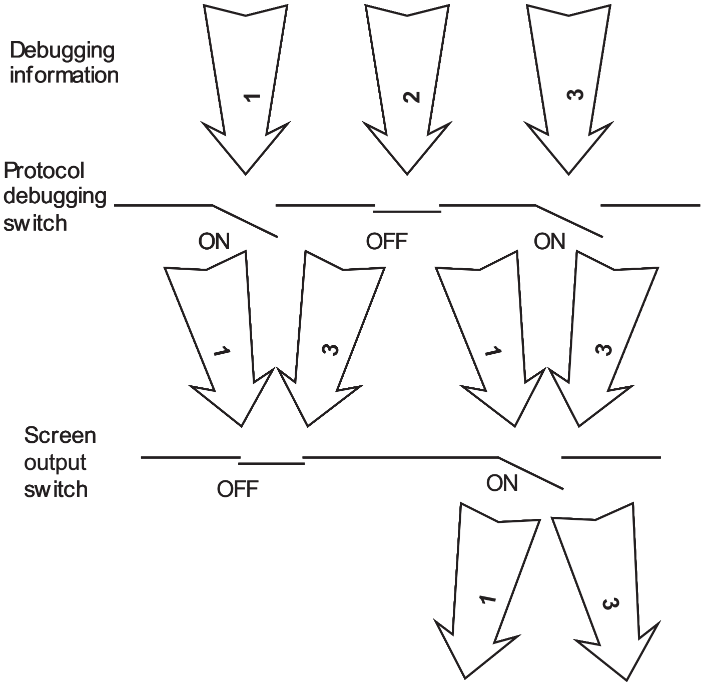
For the majority of protocols and new features provided, the system provides corresponding debugging functions to help users diagnose errors.

The following two switches control the display of debugging information:

- Protocol debugging switch, which controls the output of protocol-specific debugging information
- Screen output switch, which controls whether to display the debugging information on a certain screen.

Figure 304 illustrates the relationship between the protocol debugging switch and the screen output switch. Only when both are turned on can debugging information be output on a terminal.

Figure 304 The relationship between the protocol and screen debugging switch



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-ip* | **-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 routes 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

- *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 debugging for a specified module	**debugging** { **all** [**timeout** *time*] | *module-name* [*option*] }	Required Disabled by default Available in user view
Enable the terminal monitoring	**terminal monitor**	Optional The terminal monitoring on the console is enabled by default and that on the monitoring terminal is disabled by default.
Enable the terminal debugging	**terminal debugging**	Required Disabled by default Available in user view
Display the enabled debugging functions	**display debugging** [**interface** *interface-type interface-number*] [*module-name*]	Optional Available in any view

- *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 the Switch 8800 Command Reference Guide.*

System Maintaining Example

Network requirements

- The IP address of the destination device is 10.1.1.4.
- Display the switches a packet traverses from the current device to the destination device.

Network diagram (omitted here)

Configuration procedure

```
<Sysname> tracert nis.nsf.net
traceroute to nis.nsf.net (10.1.1.4) 30 hops max, 40 bytes packet
1  128.3.112.1  19 ms  19 ms  10 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 a packet traverses nine switches from the source to the destination device.

80 FILE SYSTEM MANAGEMENT CONFIGURATION

Throughout this document, a filename can be entered as either of the following:

- *A fully qualified filename with the path included to indicate a file under a specific path. The filename can be 1 to 135 characters in length.*
- *A short filename with the path excluded to indicate a file in the current path. The filename can be 1 to 91 characters in length.*

When configuring the file system management, go to these sections for information you are interested in:

- "File System Management" on page 1023
- "Configuration File Management" on page 1027
- "Displaying and Maintaining Device Configuration" on page 1030

File System Management

This section covers these topics:

- "File System Overview" on page 1023
- "Directory Operations" on page 1023
- "File Operations" on page 1024
- "Storage Device Operations" on page 1025
- "File System Prompt Mode Setting" on page 1026
- "File System Operations Examples" on page 1026

File System Overview

A major function of the file system is to manage storage devices. 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, may cause problems such as data loss or corruption, the file system will ask you to confirm the operation by default.

Depending on the managed object, file system operations fall into "Directory Operations" on page 1023, "File Operations" on page 1024, "Storage Device Operations" on page 1025, and "File System Prompt Mode Setting" on page 1026.

Directory Operations

Directory operations include create, delete, display the current path, display specified directory or file information as shown in the following table:

To do...	Use the command...	Remarks
Create a directory	**mkdir** *directory*	Optional
Remove a directory	**rmdir** *directory*	Optional

To do...	Use the command...	Remarks
Display the current path	**pwd**	Optional
Display files or directories	**dir** [**/all**] [*file-url*]	Optional
Change the current path	**cd** *directory*	Optional

- *The directory to be removed must be empty, meaning 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 in the Switch 8800 Command Reference Guide.*
- *After the execution of the* ***rmdir*** *command, the files in this directory will be automatically deleted forever.*

File Operations

File operations include delete (removing files into the recycle bin), restore the deleted, permanently delete (deleting files from the recycle bin), display, rename, copy, and move files, and display specified directory or file information as shown in the following table:

To do...	Use the command...	Remarks
Remove a file to the recycle bin or delete it permanently	**delete** [**/unreserved**] *file-url*	Optional
Restore a file from the recycle bin	**undelete** *file-url*	Optional
Empty the recycle bin	**reset recycle-bin** [*file-url*] [**/force**]	Optional
Display the contents of a file	**more** *file-url*	Optional Currently only a .txt file can be displayed.
Rename a file	**rename** *fileurl-source fileurl-dest*	Optional
Copy a file	**copy** *fileurl-source fileurl-dest*	Optional
Move a file	**move** *fileurl-source fileurl-dest*	Optional
Display files or directories	**dir** [**/all**] [*file-url*]	Optional
Enter system view	**system-view**	-
Execute the batch file	**execute** *filename*	Optional

You can create a file by copying or downloading or using the ***save*** *command.*

CAUTION:

- *Timely empty the recycle bin with the* ***reset recycle-bin*** *command to save memory space.*
- *As the* ***delete /unreserved*** *file-url command deletes a file permanently and the action cannot be undone, use it with caution.*
- *The original and target directory of the file to be moved must be on the same device. The* ***move*** *command does not support cross-device file moving.*
- *The* ***execute*** *command cannot ensure the execution of each command. For example, if a certain command is not correctly configured, the system will omit*

this command and go to the next one. Therefore, each configuration command in a batch file must be a standard configuration command, meaning the valid configuration information which can be displayed with the ***display current-configuration*** *command after this command is configured successfully; otherwise, this command may not be executed correctly.*

Storage Device Operations

Memory space management

You can use the **fixdisk** command to restore the space of a storage device or the **format** command to format a specified storage device as shown in the following table:

To do...	Use the command...	Remarks
Restore the space of a storage device	**fixdisk** *device*	Optional
Format a storage device	**format** *device*	Optional

You may use the two commands when some space of a storage device becomes inaccessible due to abnormal operations for example.

CAUTION: *When you format a storage device, all the files stored on it are erased and cannot be restored. In particular, if there is a startup configuration file on the storage device, formatting the storage device results in loss of the startup configuration file. Format a file under the directions of technical support switch fabricers.*

Mounting/unmounting a storage device

Switch 8800s support hot swappable storage devices, such as CF card, USB device, etc (excluding Flash), you can use the **mount** and **umount** command to mount or unmount the storage device.

When a device is unmounted, it is in a logically disconnected state, you can then remove the storage device from the system safely. To mount a device, you are reconnecting the logically disconnected device to the system.

Follow the steps below to mount/unmount a storage device:

To do...	Use the command...	Remarks
Mount a storage device	**mount** *device*	Optional A storage device is in mounted state when it is connected to the system by default.
Unmount a storage device	**umount** *device*	Optional A storage device is in mounted state by default. Before unplugging a storage device, unmount it.

CAUTION:

- *Do not remove the storage device or swap the module when mounting or unmounting the device, or when you are processing files on the storage device. Otherwise, the file system could be damaged.*
- *When a storage device is connected to a low version system, the system may not be able to recognize the device automatically; you need to use the **mount** command for the storage device to function normally.*
- *Before removing a mounted storage device from the system, you should first unmount it to avoid damaging the device.*

File System Prompt Mode Setting

The file system provides the following two prompt modes:

- **alert**: where the system warns you about operations that may bring undesirable consequence such as file corruption or data loss.
- **quiet**: where the system does not do that in any cases.

Follow these steps to set the operation prompt mode of the file system:

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 Examples

# Display the files under the root directory.

```
<Sysname> dir
Directory of flash:/
   0   drw-          -  May 08 2006 21:27:24   hafile
   1   -rw-        248  May 08 2006 21:40:44   manuinfo.txt
   2   -rw-        118  Jun 16 2006 10:16:05   ls.pwd
   3   -rw-       3530  Oct 16 2006 16:39:53   config.cfg
   4   -rw-     326944  Jul 24 2006 14:03:04   lsbSRP1N43202.app
   5   -rw-     207624  Jul 07 2006 14:27:30   lsblmcua0110y.app
   6   -rw-     326944  Jul 07 2006 11:05:39   srpbt.app
   7   -rw-     326944  Jul 10 2006 10:40:42   switch.app
15621 KB total (14363 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 files under the **test** directory.

```
<Sysname> dir
Directory of flash:/test/

   0   drw-          -  Feb 16 2006 15:28:14   mytest

2540 KB total (2519 KB free)
```

# Return to the upper directory.

```
<Sysname> cd ..
```

Configuration File Management

This section covers these topics:

- "Configuration File Management Overview" on page 1027
- "Saving the Current Configuration" on page 1027
- "Synchronizing Configuration Files Saved on the main fabric to standby fabric" on page 1028
- "Erasing the Startup Configuration File" on page 1028
- "Specifying a Configuration File for Next Startup" on page 1029
- "Backing up/Restoring the Configuration File for Next Startup" on page 1029

Configuration File Management Overview

Types of configuration

The configuration of a device falls into two types:

- Saved configuration, a configuration file used for initialization. If this file does not exist, the default parameters are used.
- Current configuration, which refers to the user's configuration during the operation of a device. This configuration is stored in dynamic random-access memory (DRAM). It is removed when the device is rebooting.

Format of configuration file

Configuration files are saved as text files for ease of reading. They:

- Save configuration in the form of commands.
- Save only non-default configuration settings.
- List commands in sections by view in this view order: system, interface, routing protocol, and so on. Sections are separated with one or multiple blank lines or comment lines that start with a pound sign (#).
- End with a return.

The operating interface provided by the configuration file management function is user-friendly. With it, you can easily manage your configuration files.

Saving the Current Configuration

You can modify the configuration on your device at the command line interface (CLI). To use the modified configuration for your subsequent startups, you must save it (using the **save** command) as a configuration file.

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 quicker but is likely to lose the 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 slower but can retain the configuration file in the flash even if the device reboots or the power fails during the process.

Follow the step below to save the current configuration:

To do...	Use the command...	Remarks
Save the current configuration	**save** [*file-name* | **safely**]	Available in any view

- *When you use the **save** file-name command, if you specify the saving directory in the file-name, the configuration will be saved in the specified directory; if you do not specify a saving directory in the file-name, the configuration will be saved in the current directory.*
- *In interactive mode, if you specify a saving directory in the file name, the directory to be specified must be the directory of the saving device on the main fabric.*
- *To save the configuration file, you can specify either the filename argument or the **safely** keyword.*
- *Fast saving mode is suitable for environments where power supply is stable. The safe mode, however, is preferred where stable power supply is unavailable or remote maintenance is involved.*
- *The extension name of the configuration file must be .cfg.*

Synchronizing Configuration Files Saved on the main fabric to standby fabric

For a Switch 8800 switch, you can only execute commands on the main fabric instead of a standby fabric. After the configuration file saving synchronization function is enabled, when you use the **save** command on the main fabric to save the current configuration, the standby fabric will automatically save the current configuration to its configuration files to keep the consistency of the configuration files on the main fabric and standby fabric.

Follow these steps to configure configuration file saving synchronization on the main fabric and standby fabric:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable configuration file saving synchronization for the main fabric and standby fabric	**slave auto-update config**	Optional Enabled by default.

Erasing the Startup Configuration File

With the configuration file erased, your device will boot up with the default configuration next time it is powered on.

You may need to erase the configuration file for one of these reasons:

- After you upgrade software, the original configuration file does not match the new software.
- The startup configuration file is corrupted or not the one you need.

Follow the step below to erase the configuration file:

To do...	Use the command...	Remarks
Erase the startup configuration file from the storage device	**reset saved-configuration**	Required Available in user view

***CAUTION:** This command will permanently delete the configuration file from the device. Use it with caution.*

Specifying a Configuration File for Next Startup

Follow the step below to specify a configuration file for next startup:

To do...	Use the command...	Remarks
Specify a configuration file for next startup	**startup saved-configuration** *cfgfile*	Required Available in user view

***CAUTION:** The configuration file must use ".cfg" as its extension name and the startup configuration file must be saved under the root directory of the device.*

Backing up/Restoring the Configuration File for Next Startup

Backup/restore function overview

The backup/restore function allows you to backup or restore a configuration file for next startup through operations at the CLI. TFTP is used for intercommunication between the device and the server. The backup function enables you to backup a configuration file to the TFTP server, while the restore function enables you to download the configuration file from the TFTP server for next startup.

For a Switch 8800 switch, when you execute the **restore** command on your main fabric, you are restoring the startup configuration file for both the main fabric and the standby fabric. However, when you execute the **backup** command on your main fabric, your operation has no effect on the standby fabric.

The backup/restore operation applies to the next startup configuration file.

Backing up the configuration file for next startup

To do...	Use the command...	Remarks
Back up the configuration file for next startup	**backup startup-configuration to** *dest-addr* [*filename*]	Required Available in user view

Before backup, 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 verify if you have set the startup configuration file, and use the **dir** command to verify if this file exists. If the file is set as NULL or does not exist, the backup will be unsuccessful.*

Restoring the startup configuration file

To do...	Use the command...	Remarks
Restore the startup configuration file	**restore startup-configuration from** *src-addr filename*	Required Available in user view

- *Before restoring a configuration file, you should ensure that the server is reachable, the server is enabled with TFTP service, and the client has permission to read and write.*
- *After the command is successfully executed, you can use the* ***display startup*** *command (in user view) to verify if the filename of the startup configuration file is the same with the filename argument, and use the* ***dir*** *command to verify if the restored file exists.*

Displaying and Maintaining Device Configuration

To do...	Use the command...	Remarks
Display the configuration file saved in the storage device	**display saved-configuration** [**by-linenum**]	Available in any view
Display the configuration file used for this and next startup	**display startup**	Available in any view
Display the validated configuration in current view	**display this**	Available in any view
Display current configuration	**display current-configuration** [**interface** [*interface-type* [*interface-number*]] | **configuration** [*configuration*] | [**by-linenum**] | [| { **begin** | **exclude** | **include** } *regular-expression*]] *	Available in any view

- *Configuration files are displayed in the same format in which they are saved.*
- *The support for the optional arguments in both the* ***display this*** *and* ***display current-configuration*** *command varies with devices. For detailed description of this command, refer to the Switch 8800 Command Reference Guide.*

81 FTP CONFIGURATION

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

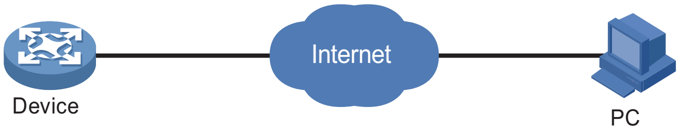
FTP Overview

The file transfer protocol (FTP) is an application layer protocol for sharing files between server and client over a TCP/IP network.

FTP adopts the server/client model. Your device can function either as client or as server (as shown in Figure 305). They work in the following way:

- When the device serves as the FTP client, a PC user first telnets or connects to the device through an emulation program, then executes the **ftp** command to establish the connection to the remote FTP server, and gain access to the files on the server. The device must obtain FTP username and password first to log onto the remote FTP server.
- When the device serves as the FTP server, it must be configured with an IP address so that a user running FTP client program can access it. For the sake of security, the device does not support anonymous FTP. Therefore, you must use an authenticated username and password. By default, authenticated users can access the root directory of the device.

Figure 305 Network diagram for FTP



CAUTION:

- *The FTP function is available when a route exists between the FTP server and the FTP client.*
- *When a device serving as the FTP server logs onto the device using IE, some IE functions are not supported because multiple user connections are established, and the device supports only one connection currently.*

Configuring the FTP Client

Establishing an FTP Connection

To access an FTP server, the FTP client must connect with it. Two ways are available for the connection: using the **ftp** command to establish the connection directly; using the **open** command in FTP client view.

Multiple routes may exist for the FTP client to successfully access the FTP server. You can specify one by configuring the source address of the packets of the FTP client to meet the requirement of the security policy of the FTP client. 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 FTP client is specified, a device uses the IP address of the interface determined by the routing protocol 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 the source address is specified with the **ftp client source** command and then with the **ftp** command, the address specified with the latter one 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** { **ip** *source-ip-address* | **interface** *interface-type interface-number* }	Optional A device uses the IP address of the interface determined by the routing protocol as the source IP address to communicate with the FTP server by default.
Exit to system view	**quit**	-
Log onto the remote FTP server directly in user view	**ftp** [*server-address* [*service-port*] [**source** { **ip** *source-ip-address* | **interface** *interface-type interface-number* }]]	Use either approach. Available in user view
Log onto the remote FTP server indirectly in FTP client view	**ftp**	
	open *server-address* [*service-port*]	

- *If no primary IP address is configured on the source interface, the FTP connection fails.*

- *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 new source IP address will overwrite the current one, and vice versa.*

Follow these steps to establish an FTP connection (In IPv6 networking):

To do...	Use the command...	Remarks
Log onto 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.
Log onto 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 establishing FTP connection, refer to "Establishing an FTP Connection" on page 1032.), the device can perform the following operations for the authorized directory:

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
Use other username to relog after logging onto 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** *pathname*	Optional
Exit the current directory and enter the upper level directory	**cdup**	Optional
Display files/directories information on the FTP server	**dir** [*remotefile* [*localfile*]]	Optional
Check 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 working directory of the remote FTP server	**pwd**	Optional
Find the working path of the FTP client	**lcd**	Optional

To do...	Use the command...	Remarks
Create a directory on the FTP server	**mkdir** *directory*	Optional
Set the data transfer mode to passive	**passive**	Optional Passive by default
Delete specified file on the FTP server	**delete** *remotefile*	Optional
Delete specified directory on the FTP server	**rmdir** *directory*	Optional
Disconnect with the FTP server without exiting the FTP client view	**disconnect**	Optional Equal to the **close** command
Disconnect with the FTP server without exiting the FTP client view	**close**	Optional Equal to the **disconnect** command
Disconnect with 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

- *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 size and date of creation of files or directories.*

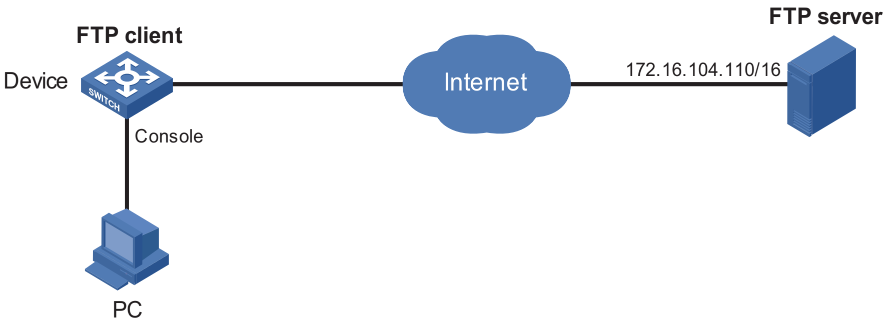
FTP Client Configuration Examples

Network requirements

- Use your device as an FTP client to download an image file from the FTP server.
- The IP address of the FTP server is 172.16.104.110/16.
- On the FTP server, an FTP user account has been created for the FTP client, with the username being **abc** and the password being **pwd**.
- The PC performs operations on the device through Console port.

Network diagram

Figure 306 Network diagram for FTPing an image file from an FTP server



Configuration procedure

# Check files on your device. Remove those redundant to ensure adequate space for the image file to be downloaded.

```
<Sysname> dir
Directory of flash:/

   0   drw-         -  Dec 07 2005 10:00:57   filename
   1   drw-         -  Jan 02 2006 14:27:51   logfile
   2   -rw-      1216  Jan 02 2006 14:28:59   config.cfg
   3   -rw-      1216  Jan 02 2006 16:27:26   backup.cfg

2540 KB total (2511 KB free)
<Sysname> delete flash:/backup.cfg
```

# Download the image file from the server.

```
<Sysname> ftp 172.16.104.110
Trying 172.16.104.110 ...
Connected to 172.16.104.110.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(172.16.104.110:(none)):abc
331 Give me your password, please
Password:
230 Logged in successfully
[ftp] binary
200 Type set to I
[ftp] get aaa.app bbb.app

227 Entering Passive Mode (10.1.1.1,4,1).
125 BINARY mode data connection already open, transfer starting for aaa.app.
.....226 Transfer complete.
FTP: 5805100 byte(s) received in 19.898 second(s) 291.74Kbyte(s)/sec.
[ftp] bye
```

# Specify the image file for next startup with the **boot-loader** command

```
<Sysname> boot-loader file bbb.app
<Sysname> reboot
```

CAUTION: *The image file specified by the* ***boot-loader*** *command for next startup must be saved under the root directory. You can change the directory of a file to the root directory through copy or move operation. For the details of the* ***boot-loader*** *command, refer to the Switch 8800 Command Reference Guide.*

Configuring the FTP Server

Configuring FTP Server Operating Parameters

The FTP server uses two modes to update files when you upload files (use the **put** command) to the FTP server:

- In fast mode, the FTP server starts writing data to the Flash after file transfer completes. This protects the files intended to be overwritten on the device from being corrupted in the event that anomalies, power failure for example, occur during a file transfer.
- In normal mode, the FTP server writes data to the Flash during file transfer. This means that any anomaly, power failure for example, during file transfer might

result in file corruption on the router. 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.
Configure the idle-timeout timer	**ftp timeout** *minutes*	Optional 30 minutes by default. In 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 in FTP	**ftp update** { **fast** | **normal** }	Optional Normal update is used by default.

Configuring Authentication and Authorization for Accessing 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.

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

If FTP server performs authentication, authorization and accounting (AAA) policy on FTP client, AAA related parameters should be configured on the FTP server. For more information about the ***local-user****,* ***password*** *and* ***service-type*** *commands refer to the Switch 8800 Command Reference Guide and to the "AAA, RADIUS and HWTACACS Configuration" on page 873 for more information on the AAA related configuration.*

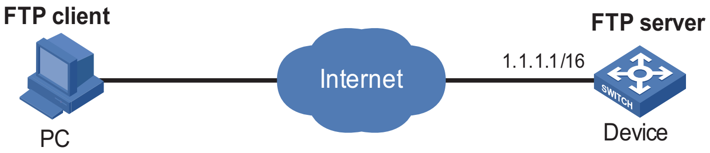
FTP Server Configuration Examples

Network requirements

- Use your device as an FTP server. Create a user account for an FTP user on it, setting the username to **abc** and the password to **pwd**.
- The IP address of the VLAN interface is 1.1.1.1/16.
- The PC serves as the FTP client, which is to upload an image file.

Network diagram

Figure 307 Smooth upgrading using the FTP server



Configuration procedure

1 Configure Device (FTP Server)

# Create an FTP user account abc, setting its password to pwd.

```
<Sysname> system-view
[Sysname] local-user abc
[Sysname-luser-abc] service-type ftp
[Sysname-luser-abc] password simple pwd
```

# Specify abc to use FTP, and authorize its access to certain directory.

```
[Sysname-luser-abc] work-directory flash:
[Sysname-luser-abc] quit
```

# Enable FTP server.

```
[Sysname] ftp server enable
[Sysname] quit
```

# Check files on your device. Remove those redundant to ensure adequate space for the image file to be uploaded.

```
<Sysname> dir
Directory of flash:/

   0    drw-          -   Dec 07 2005 10:00:57   filename
   1    drw-          -   Jan 02 2006 14:27:51   logfile
   2    -rw-       1216   Jan 02 2006 14:28:59   config.cfg
   3    -rw-       1216   Jan 02 2006 16:27:26   back.cfg
   4    drw-          -   Jan 02 2006 15:20:21   ftp

2540 KB total (2511 KB free)
<Sysname> delete /unreserved flash:/back.cfg
```

2 Configure the PC (FTP Client)

# Upload the image file to the FTP server and save it under the root directory of the FTP server.

```
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.
ftp> put aaa.app bbb.app
```

- *When upgrading the configuration file with FTP, put the new file under the root directory*
- *After you finish upgrading the Boot ROM program through FTP, you must execute the **bootrom upgrade** command to refresh the system configuration.*

# Specify the image file for next startup with the **boot-loader** command.

```
<Sysname> boot-loader file bbb.app
<Sysname> reboot
```

***CAUTION:** The image file specified by the **boot-loader** command for next startup must be saved under the root directory. You can change the directory of the file to the root directory through copy or move operation. For details of the **boot-loader** command, refer to the Switch 8800 Command Reference Guide.*

Displaying and Maintaining FTP

Use the following **display** commands to display and maintain the FTP server:

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

82 TFTP CONFIGURATION

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

TFTP Overview

The trivial file transfer protocol (TFTP) provides functions similar to those provided by FTP, but it is not as complex as FTP in interactive access interface and authentication. Therefore, it is more suitable where complex interaction is not needed between client and server.

TFTP uses the UDP service for data delivery. 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 for programming files and ASCII for text files.

Only the TFTP client service is available with your device at present.

Configuring the TFTP Client

When a device acts as a TFTP client, you can upload files on the device to a TFTP server and download files from the TFTP server to the local device. You can use either of the following ways to download files:

- Normal download: The device writes the obtained files to the storage device directly. In this way, 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 files to its memory and does not write them to the storage device until all user files are 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 securer but consumes more memory.

You are recommended to use the latter mode or use a filename not existing in the current directory as the target filename when downloading startup file or configuration file.

Multiple routes may exist for a TFTP client to successfully access the TFTP server. You can specify one by configuring the source address of the packets from the TFTP client to meet the requirement of the security policy of the TFTP client. 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 routing protocol 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 the source address is specified with the **tftp client source** command and then with the **tftp** command, the source address configured with the latter one 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**	-
Reference an access control list (ACL) to the TFTP server	**tftp-server** [**ipv6**] **acl** *acl-number*	Optional
Configure the source address of the TFTP client	**tftp client source** { **ip** *source-ip-address* | **interface** *interface-type interface-number* }	Optional A device uses the source address determined by the routing protocol to communicate with the TFTP server by default.
Return to user view	**quit**	-
Download or upload a file in IPv4 network	**tftp** *server-address* { **get** | **put** | **sget** } *source-filename* [*destination-filename*] [**source** { **ip** *source-ip-address* | **interface** *interface-type interface-number* }]	Optional
Download or upload a file in IPv6 network	**tftp ipv6** *tftp-ipv6-server* [**-i** *interface-type interface-number*] { **get** | **put** } *source-file* [*destination-file*]	Optional

- *If no primary IP address is configured on the source interface, TFTP connection fails.*

- *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

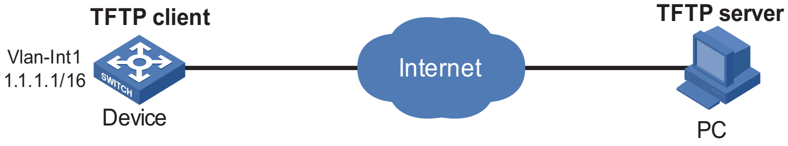
TFTP Client Configuration Examples

Network requirements

- Use a PC as the TFTP server and your device as the TFTP client.
- PC uses IP address 1.1.1.2/24 and a TFTP working directory has been defined for the client.
- On your device, VLAN-interface 1 is assigned an IP address 1.1.1.1/24, making sure that the port connected to PC belongs to the same VLAN.
- TFTP an image file from PC for upgrading and a configuration file config.cfg to PC for backup.

Network diagram

Figure 308 Smooth upgrading using the TFTP client function



Configuration procedure

1 Configure PC (TFTP Server), the configuration procedure omitted.

- On the PC, enable TFTP server
- Configure a TFTP working directory

2 Configure the device (TFTP Client)

CAUTION: *If the free memory space of the device is not big enough, you should delete the existing programs before downloading new ones.*

# Enter system view.

```
<Sysname> system-view
```

# Assign VLAN-interface 1 an IP address 1.1.1.1, making sure that the port connected to PC belongs to the same VLAN.

```
[Sysname] interface Vlan-interface 1
[Sysname-Vlan-interface1] ip address 1.1.1.1 255.255.255.0
[Sysname-Vlan-interface1] return
```

# Download an application file aaa.app from the TFTP server.

```
<Sysname> tftp 1.1.1.2 get aaa.app bbb.app
```

# Upload a configuration file config.cfg to the TFTP server.

```
<Sysname> tftp 1.1.1.2 put config.cfg config.cfg
```

# Specify the image file for next startup with the **boot-loader** command

```
<Sysname> boot-loader file bbb.app
<Sysname> reboot
```

CAUTION: *The image file specified by the **boot-loader** command for next startup must be saved under the root directory. You can change the directory of the file to the root directory through copy or move operation. For details of the **boot-loader** command, refer to the Switch 8800 Command Reference Guide.*

83 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 information, modify information, find and diagnose network problems, plan for network growth, and generate reports on any 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 manufactures, especially so in small, fast and low cost network environments.

SNMP Mechanism

An SNMP enabled network is comprised of network management station (NMS) and Agent.

- NMS is a station that runs the SNMP client software. It offers a user friendly human computer interface, making it easier for network administrators to perform most network management tasks. Currently, the most commonly used NMSs include Sun NetManager and IBM NetView.
- 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.
- NMS manages an SNMP enabled network, whereas Agent is the managed network device. They exchange management information through the SNMP protocol.

SNMP provides the following four basic operations:

- Get operation: NMS gets the behavior information of the Agent through this operation.
- Set operation: NMS can reconfigure certain values in the Agent MIB (management information base) to make the Agent perform certain tasks by means of this operation.
- Trap operation: Agent sends Trap information to the NMS through this operation.
- Inform operation: NMS sends Trap information to other NMSs through this operation.

SNMP Protocol Version

Currently, SNMP agents support SNMPv3 and are compatible with SNMPv1 and SNMPv2c.

SNMPv1 and SNMPv2c authenticate by means of community name, 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.

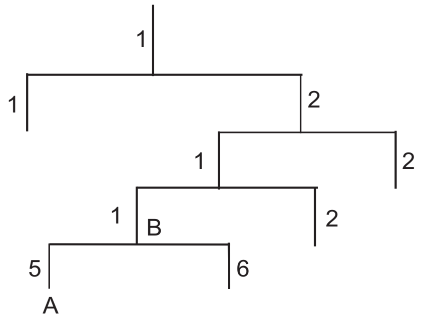
SNMPv3 offers an authentication that is implemented with a User-Based Security Model (USM for short), which could be authentication with privacy, authentication without privacy, or no authentication no privacy. USM regulates the access from NMS to Agent in a more efficient way.

MIB Overview

Management information base (MIB) is a collection of all the objects managed by NMS. It defines the set of characteristics associated with the managed objects, such as the object identifier (OID), access right and data type of the objects.

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 309 MIB tree



SNMP Configuration

As configurations for SNMPv3 differ substantially from those of SNMPv1 and SNMPv2c, their SNMP functionalities will be 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 Technologies Co.,Ltd. for contact, Hangzhou China for location, and SNMPv3 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
Add a new user to an SNMP agent group	**snmp-agent usm-user v3** *user-name group-name* [**authentication-mode** { **md5** | **sha** } *auth-password* [**privacy-mode** { **des56** | **aes128** } *priv-password*]] [**acl** *acl-number*]	Required
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 switch fabric ID for a local SNMP agent	**snmp-agent local-switch fabricid** *switch fabricid*	Optional Company ID and device ID by default
Create or update the MIB view content for an SNMP agent	**snmp-agent mib-view** { **included** | **excluded** } *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**".

To do...			Use the command...	Remarks
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 Technologies Co.,Ltd. for contact, Hangzhou China for location and SNMPv3 for the version.
Configure SNMP NMS access right	Configure directly	Configure a community name	**snmp-agent community** { **read** | **write** } *community-name* [**acl** *acl-number* | **mib-view** *view-name*]*	Use either approach. The community name of SNMPv1 or SNMPv2c is used in direct configuration. The second approach was introduced to be compatible with SNMPv3. Adding a user to a specified group equals to the configuration of the community name of SNMPv1 and SNMPv2c. The community name configured on NMS should be consistent with the corresponding 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 15,00 bytes by default
Configure the switch fabric ID for a local SNMP agent			**snmp-agent local-switch fabricid** *switch fabricid*	Optional Company ID and device ID by default
Create or update MIB view content for an SNMP agent			**snmp-agent mib-view** { **included** | **excluded** } *view-name oid-tree* [**mask** *mask-value*]	Optional ViewDefault by default

CAUTION: *The validity of a USM user depends on the switch fabric ID of the SNMP agent. If the switch fabric ID used for USM user creation is not identical to the current switch fabric ID, the USM user is invalid.*

Trap Configuration

SNMP Agent sends Trap messages to NMS to alert the latter of critical and important events (such as restart of the managed device).

Configuration Prerequisites

Basic SNMP configurations have been completed.

Configuration Procedure

Enabling Trap message transmission

Follow these steps to enable Trap packet transmission:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-

To do…	Use the command…	Remarks
Set to enable the device to send Trap packets globally	**snmp-agent trap enable** [**bgp** | **configuration** | **flash** | **mpls** | **ospf** [*process-id*] [*ospf-trap-list*] | **standard** [**authentication** | **coldstart** | **linkdown** | **linkup** | **warmstart**]* | **system** | **vrrp** [**authfailure** | **newmaster**]]	Optional All types of Trap packets are allowed by default.
Enter Ethernet interface view	**interface** *interface-type interface-number*	-
Set to enable the device to send Trap packets of interface state change	**enable snmp trap updown**	Optional Transmission of Trap packets of interface state change is allowed by default.

CAUTION: *To enable an interface to send SNMP Trap packets when its state changes, you need to enable the Link up/down Trap packet transmission function on an interface and globally. Use the* ***enable snmp trap updown*** *command to enable this function on an interface, and use the* ***snmp-agent trap enable*** *[* ***standard*** *[* ***linkdown*** **|** ***linkup*** *] *] command to enable this function globally.*

Configuring Trap message transmission parameters

Follow these steps to configure Trap:

To do…	Use the command…	Remarks
Enter system view	**system-view**	-
Configure target host attribute for Trap messages	**snmp-agent target-host trap address udp-domain** { *ip-address* | **ipv6** *ipv6-address* } [**udp-port** *port-number*] **params securityname** *security-string* [**v1** | **v2c** | **v3** [**authentication** | **privacy**]]	Required
Configure the source address for Trap messages	**snmp-agent trap source** { *interface-type interface-number* }	Optional
Configure the queue size for sending Trap messages	**snmp-agent trap queue-size** *size*	Optional 100 by default
Configure the life for Trap messages	**snmp-agent trap life** *seconds*	Optional 120 seconds by default

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 switch fabric ID	**display snmp-agent local-switch fabricid**	
Display SNMP agent group information	**display snmp-agent group** [*group-name*]	
Display SNMP v3 agent user information	**display snmp-agent usm-user** [**switch fabricid** *switch fabricid* | **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*]	

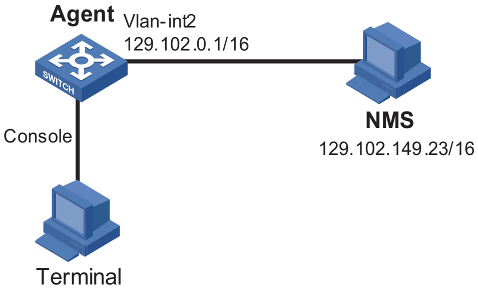
SNMP Configuration Examples

Network requirements

- The NMS connects to the agent, a switch, through an Ethernet.
- The IP address of the NMS is 129.102.149.23/16.
- The IP address of VLAN interface on the switch is 129.102.0.1/16.
- On the switch, configure the following: community name, access right, administrator ID, contact, location, enabling sending of Trap messages.

Network diagram

Figure 310 Network diagram for SNMP



Configuration procedure

1 Configuring SNMP Agent

# Configure the community name, the SNMP agent group, and SNMP agent user.

```
<Sysname> system-view
[Sysname] snmp-agent sys-info version all
```

```
[Sysname] snmp-agent community read public
[Sysname] snmp-agent community write private
[Sysname] snmp-agent mib-view include internet 1.3.6.1
[Sysname] snmp-agent group v3 managev3group write-view internet
[Sysname] snmp-agent usm-user v3 managev3user managev3group
```

# Configure VLAN-interface 2 (with the IP address of 129.102.0.1/16) for network management. Add port Ethernet 2/1/3 used for network management to VLAN 2.

```
[Sysname] vlan 2
[Sysname-vlan2] port ethernet 2/1/3
[Sysname-vlan2] interface Vlan-interface 2
[Sysname-Vlan-interface2] ip address 129.102.0.1 255.255.0.0
[Sysname-Vlan-interface2] quit
```

# Configure the system information of the switch.

```
[Sysname] snmp-agent sys-info version all
[Sysname] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Sysname] snmp-agent sys-info location telephone-closet,3rd-floor
```

# Enable the sending of Trap messages to the NMS with an IP address of 129.102.149.23/16, using **public** as the community name.

```
[Sysname] snmp-agent trap enable
[Sysname] snmp-agent target-host trap address udp-domain 129.102.149
.23 udp-port 5000 params securityname public
```

2 Configuring SNMP NMS

SNMPv3 uses authentication and privacy security model. In NMS, the user needs to specify username and security level, and based on that level, configure the authentication mode, authentication password, privacy mode, privacy password. In addition, the time-out time and number of retries should also be configured. The user can inquire and configure the switch through NMS. For detailed information, refer to the NMS manuals.

The configurations on the agent and the NMS must match in order to perform the related operations.

84 RMON CONFIGURATION

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

- "RMON Overview" on page 1051
- "Configuring RMON" on page 1053
- "Displaying and Maintaining RMON" on page 1054
- "RMON Configuration Examples" on page 1055

RMON Overview

This section covers these topics:

- "Introduction" on page 1051
- "RMON Groups" on page 1052

Introduction

Remote Monitoring (RMON) is a type of IETF-defined MIB. It is the most important enhancement to the MIB II standard. It allows you to monitor traffic on network segments and even the entire network.

RMON is implemented based on the simple network management protocol (SNMP) and is fully compatible with the existing SNMP framework.

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 probe for an interface. It monitors and gathers information about traffic over the network segment connected to the interface to provide statistics about packets over a specified period and good packets sent to a host for example.

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 with 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.

Switch 8800 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

RMON categorizes objects into ten groups. This section describes only the major implemented five groups.

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 events in the event log table
- Sending traps to NMSs
- Both logging and sending traps
- 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 rising threshold, a rising alarm event is triggered; if the sampled value of the monitored variable is lower than or equal to the falling threshold, a falling alarm 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.

If a monitored 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.

History group

The history group controls the periodic statistical sampling of data, such as bandwidth utilization, number of errors, and total number of packets.

Note that each value provided by the group is a cumulative sum 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, and so on.

After the creation of a valid event entry on a specified interface, the Ethernet statistics group counts the number of packets received on the current interface. The result of the statistics is a cumulative sum.

Currently, Switch 8800s do not support statistics about oversize frames and bytes received.

Configuring RMON

Configuration Prerequisites

Before configuring RMON, configure the SNMP agent as described in the *SNMP Configuration* in *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** | **trap** *trap-community* | **log-trap** *log-trapcommunity* | **none** } [**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**	Required
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

To do...	Use the command...	Remarks
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

- *Two entries with the same configuration cannot be created. If the parameters of a newly created entry are identical to the corresponding parameters of an existing entry, the system considers their configurations the same and the creation fails. Refer to Table 41 for the parameters to be compared for different entries.*
- *The system limits the total number of all types of entries (Refer to Table 41 for the detailed numbers). When the total number of an entry reaches the maximum number of entries that can be created, the creation fails.*

Table 41 Limitations 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
Pri-alarm	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

To do...	Use the command...	Remarks
Display RMON history information and the latest 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 RMON event log information	**display rmon eventlog** [*event-number*]	Available in any view

RMON Configuration Examples

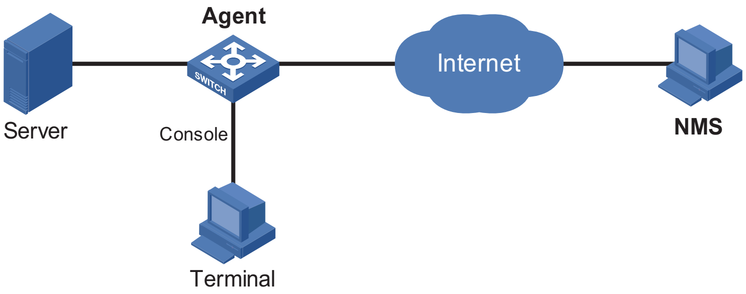
Network requirements

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 4/2/2, and logging is enabled after received bytes exceed the specified threshold.

Network diagram

Figure 311 Network diagram for RMON



Configuration procedure

# Configure RMON to gather statistics for interface GigabitEthernet 4/2/2.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 4/2/2
[Sysname-GigabitEthernet4/2/2] rmon statistics 1 owner user1-rmon
[Sysname-GigabitEthernet4/2/2] quit
```

# Display RMON statistics for interface GigabitEthernet 4/2/2.

```
<Sysname> display rmon statistics GigabitEthernet 4/2/2
Statistics entry 1 owned by user1-rmon is VALID.
  Interface : GigabitEthernet4/2/2<ifIndex.157>
  etherStatsOctets         : 0          , etherStatsPkts          : 0
  etherStatsBroadcastPkts  : 0          , etherStatsMulticastPkts : 0
  etherStatsUndersizePkts  : 0          , etherStatsOversizePkts  : 0
  etherStatsFragments      : 0          , etherStatsJabbers       : 0
  etherStatsCRCAlignErrors : 0          , etherStatsCollisions    : 0
  etherStatsDropEvents (insufficient resources): 0
  Packets received according to length:
  64     : 0         ,  65-127  : 0         ,  128-255  : 0
  256-511: 0         ,  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
[Sysname] display rmon event 1
Event table 1 owned by 1-rmon is VALID.
  Description: null.
  Will cause log when triggered, last triggered at 2day(s) 03h:56m:06s.
```

# Configure an alarm group.

```
[Sysname] rmon alarm 1 1.3.6.1.2.1.16.1.1.1.4.1 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
```

85 NTP CONFIGURATION

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

i *The term router and the router icons used in this chapter refer to the routers in a generic sense and the switches running routing protocols.*

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

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 increment backup between a backup server and clients, timekeeping must be synchronized between the backup server and all the clients.

An administrator can by no means keep synchronized time 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.

Advantages of NTP:

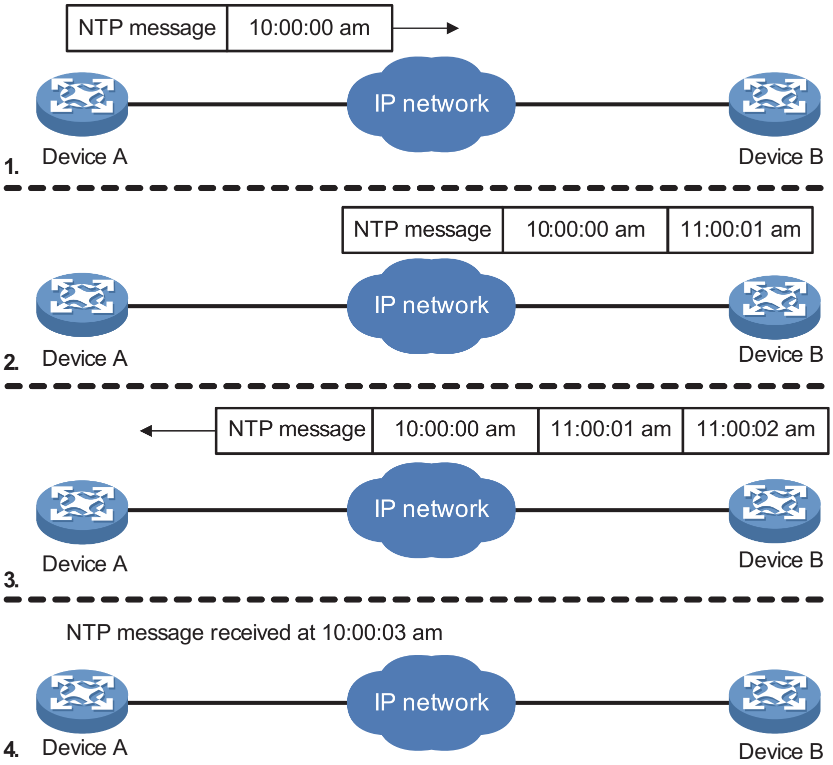
- NTP uses a stratum to describe the clock precision, and is able to synchronize time among all devices within the network.
- NTP supports access control and MD5 authentication.
- NTP can unicast, multicast or broadcast protocol messages.

How NTP Works

Figure 312 shows the basic work flow 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 10:00:00am while that of Device B is set to 11:00:00am.
- 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 312 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 10:00:00am (T_1).
- When this NTP message arrives at Device B, it is timestamped by Device B. The timestamp is 11:00:01am (T_2).
- When the NTP message leaves Device B, Device B timestamps it. The timestamp is 11:00:02am (T_3).
- When Device A receives the NTP message, the local time of Device A is 10:00:03am (T_4).

Up to now, Device A has sufficient information to calculate the following two important parameters:

- The roundtrip delay of NTP message: Delay = $(T_4-T_1) - (T_3-T_2)$ = 2 seconds.
- Time difference between Device A and Device B: Offset = $((T_2-T_1) + (T_3-T_4))/2$ = 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.

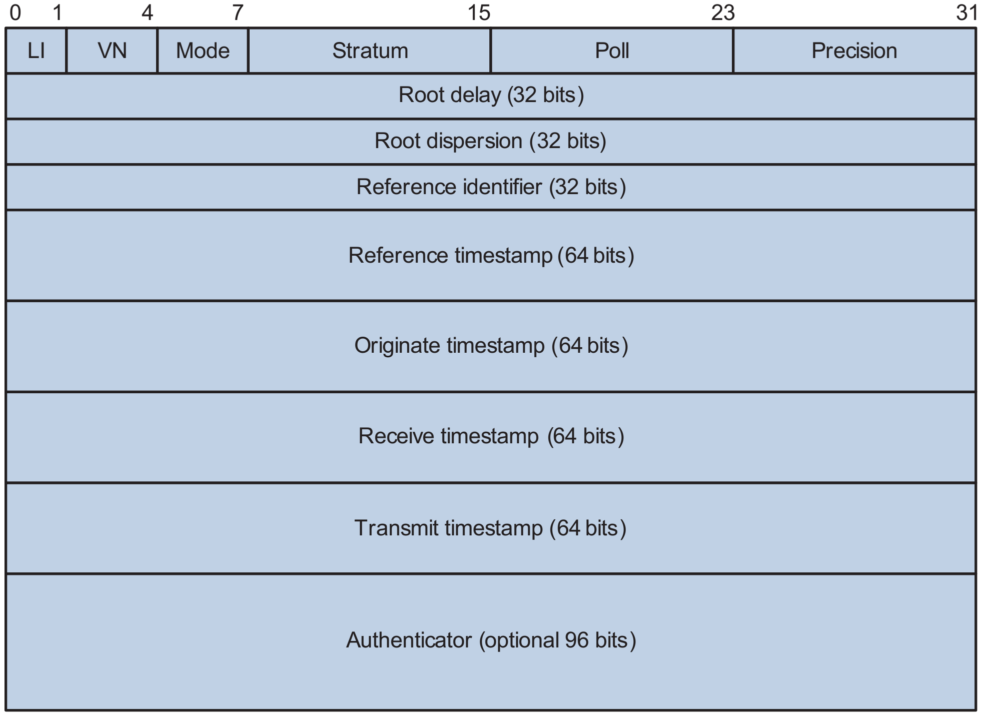
NTP Message Format

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.

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 313.

Figure 313 Clock synchronization message format


0
1
4
7
15
23
31
LI
VN
Mode
Stratum
Poll
Precision
Root delay (32 bits)
Root dispersion (32 bits)
Reference identifier (32 bits)
Reference timestamp (64 bits)
Originate timestamp (64 bits)
Receive timestamp (64 bits)
Transmit timestamp (64 bits)
Authenticator (optional 96 bits)

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 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 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:

Server/client mode

In server/client mode, a client can be synchronized to a server, but not vice versa. When working in the server/client 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.

Symmetric peers mode

After the symmetric peers mode is configured, the symmetric active peer sends clock synchronization messages with the Mode field set to 3 (client mode) to the symmetric passive peer. The device that receives the message automatically enters the symmetric passive mode and sends a reply, with the Mode field in the message set to 4 (server mode). 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. In this case, the Mode field is set to 1 (symmetric active peer) in the clock synchronization messages sent by the symmetric active peer, and that is set to 2 (symmetric passive peer) in the response messages sent by the symmetric passive peer. If both devices have reference clocks, the device whose local clock has a lower stratum level will synchronize the clock of the other device.

Broadcast mode

In the broadcast mode, a server periodically sends clock synchronization messages to the broadcast address 255.255.255.255. Clients listen to the broadcast messages from servers. After a client receives the first broadcast message, the client initiates the server/client mode request to acquire the network delay between the 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

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. Clients listen to the multicast messages from servers. After a client receives the first multicast message, the client initiates the server/client mode request to acquire the network delay between the 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.

Multiple Instances of NTP

The server/client mode and symmetric mode support multiple instances of NTP and thus support clock synchronization within an MPLS VPN network. Namely, network devices (CEs and PEs) at different physical location can get their clocks synchronized through MPLS VPN connection, as long as they are in the same VPN. The specific functions are as follows:

- The NTP client on a customer edge device (CE) can be synchronized to the NTP server on another CE.
- The NTP client on a CE can be synchronized to the NTP server on a provider edge device (PE).
- The NTP client on a PE can be synchronized to the NTP server on a CE through a designated VPN instance.
- The NTP server on a PE can synchronize multiple NTP clients on different CEs.

- *A CE is a device that has an interface directly connecting to the service provider (SP). A CE is not "aware of" the presence of the VPN.*
- *A PE is a device that directly connecting to CEs. In an MPLS network, all events related to VPN processing occur on the PE.*

NTP Configuration Task List

Complete these tasks to configure NTP:

Configuration tasks	Remarks
"Configuring the Operation Modes of NTP" on page 1062	Required
"Configuring the Local Clock as a Reference Source" on page 1066	Optional
"Configuring Optional Parameters of NTP" on page 1066	Optional
"Configuring Access-Control Rights" on page 1067	Optional
"Configuring NTP Authentication" on page 1068	Optional

Configuring the Operation Modes of NTP

According to the devices' position in the network and the network structure, devices can implement clock synchronization in one of the following modes:

- Server/client mode
- Symmetric peers mode
- Broadcast mode
- Multicast mode

For the server/client mode or symmetric peers 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.

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 server/client 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 Server/Client Mode

For devices working in the server/client mode, you only need to make configurations on the clients, and 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

- *In the* ***ntp-service unicast-server*** *command, ip-address must be a host address, rather than a broadcast address, a multicast address or the IP address of the local clock.*
- *When the interface sending the NTP packet is specified by the* ***source-interface*** *argument, the source IP address of the NTP packet 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 Mode

For devices working in the symmetric mode, you need to specify a symmetric-passive 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**	-

To do...	Use the command...	Remarks
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

- *In the symmetric mode, you should use the* ***ntp-service refclock-master*** *command or any NTP configuration command in section "Configuring the Operation Modes of NTP" on page 1062 to enable NTP; otherwise, a symmetric-passive peer will not process NTP packets from a symmetric-active peer.*
- *In the* ***ntp-service unicast-peer*** *command, ip-address must be a host address, rather than a broadcast address, a multicast address or the IP address of the local clock.*
- *When the interface used to send NTP messages is specified by the* ***source-interface*** *argument, the source IP address of the NTP message 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

For devices working in the broadcast mode, you need to configure both the server and clients. The broadcast server periodically sends NTP broadcast messages to the broadcast address 255.255.255.255. Because an interface need 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 commands for NTP broadcast mode can be configured only in the specific interface view.

Configuring a broadcast client

Follow these steps to configure an NTP 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

Follow these steps to configure the NTP broadcast server:

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 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

A broadcast server can synchronize broadcast clients only after its clock has been synchronized.

Configuring NTP Multicast Mode

If using the multicast mode, you need to configure both the server and clients. The multicast server periodically sends NTP multicast messages to multicast clients. The commands for the NTP multicast mode must be configured in the specific interface view. You can configure a maximum of 1,024 multicast clients, among which 128 can take effect at the same time.

Configuring a multicast client

Follow these steps to configure an NTP multicast 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 multicast messages
Configure the device to work in the NTP multicast client mode	**ntp-service multicast-client** [*ip-address*]	Required The multicast IP address must be 224.0.1.1.

Configuring the multicast server

Follow these steps to configure the NTP multicast server:

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 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

A multicast server can synchronize broadcast clients only after its clock has been synchronized.

Configuring the Local Clock as a Reference Source

A network device can get its clock synchronized in one of the following two ways:

- Synchronized to the local clock, which as the reference source.
- Synchronized to another device on the network in any of the four NTP operation modes previously described.

If you configure two synchronization modes, the device will choose the optimal clock as the reference source.

Follow these steps to configure the local clock as a reference source:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the local clock as a reference source	**ntp-service refclock-master** [*ip-address*] [*stratum*]	Required

In this command, ip-address must be 127.127.1.u, where u ranges from 0 to 3, representing the NTP process ID.

Configuring Optional Parameters of NTP

Configuring the Interface to Send NTP Messages

After you specify the interface used to send NTP messages, the source IP address of the NTP message will be configured as the primary IP address of the specified interface.

Following these steps to configure the local interface used to send NTP messages:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the interface used to send NTP messages	**ntp-service source-interface** *interface-type interface-number*	Required

CAUTION: *If you have specified an interface in the* ***ntp-service unicast-server*** *or* ***ntp-service unicast-peer*** *command, this interface will be used for sending NTP messages.*

Disabling an Interface from Receiving NTP Messages

Follow these steps to disable an interface from receiving NTP messages:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enter interface view	**interface** *interface-type interface-number*	Required
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

Follow these steps to configure the maximum number of dynamic sessions allowed to be established locally:

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 device to perform control query to the NTP service on the local device but does not permit the peer device to synchronize its clock to 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 the peer device to synchronize its clock to the local device but does not permit the peer device to perform control query.
- **server**: server access and query permitted. This level of right permits the peer device to perform synchronization and control query to the local device but does not permit the local device to synchronize its clock to the peer device.
- **peer**: full access. This level of right permits the peer device to perform synchronization and control query to the local device and also permits the local device to synchronize its clock to the 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 Overview" on page 801*.

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 to the local device	**ntp-service access** { **peer** | **query** | **server** | **synchronization** } *acl-number*	Required **peer** by default

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 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 server/client 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 server/client mode, if the NTP authentication feature has not been enabled for the client, the client can synchronize with the server regardless the NTP authentication feature has been enabled for the server or not.
- For all synchronization modes, the server side and the client side must be consistently configured.
- 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.

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

To do...	Use the command...	Remarks
Associate the specified key with an NTP server	Server/client mode: **ntp-service unicast-server** { *ip-address* | *server-name* } **authentication-keyid** *keyid* Symmetric peers mode: **ntp-service unicast-peer** { *ip-address* | *peer-name* } **authentication-keyid** *keyid*	Required

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 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* Multicast server mode: **ntp-service multicast-server authentication-keyid** *keyid*	Required

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**]	
View the brief information of the NTP servers from the local device back to the primary reference source	**display ntp-service trace**	

NTP Configuration Examples

Unless otherwise specified, the examples given in this section apply to all switches and routers that support NTP.

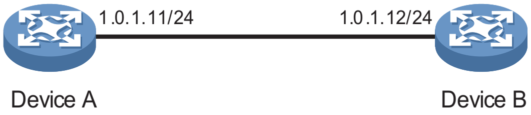
Configuring NTP Server/Client Mode

Network requirements

The local clock of Device A is to be used as a reference source, with the stratum level of 2. Device A is to be used as the NTP server of Device B, with Device B as the client mode, and Device A then as the server automatically.

Network diagram

Figure 314 Network diagram for NTP server/client mode configuration



Configuration procedure

1 Configuration on Device A:

# Specify the local clock as the reference source, with the stratum level of 2.

```
<DeviceA> system-view
[DeviceA] ntp-service refclock-master 2
```

2 Configuration on Device B:

# View the NTP status of Device B before clock synchronization.

```
<DeviceB> 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 Device A as the NTP server.

```
<DeviceB> system-view
[DeviceB] ntp-service unicast-server 1.0.1.11
```

# (After the above configurations, Device B is synchronized to Device A.) View the NTP status of Device B after clock synchronization.

```
[DeviceB] 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, Device B has been synchronized to Device A, and the clock stratum level of Device B is 3, while that of Device A is 2.

# View the NTP session information of Device B, which shows that an association has been set up between Device B and Device A.

```
[DeviceB] 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
```

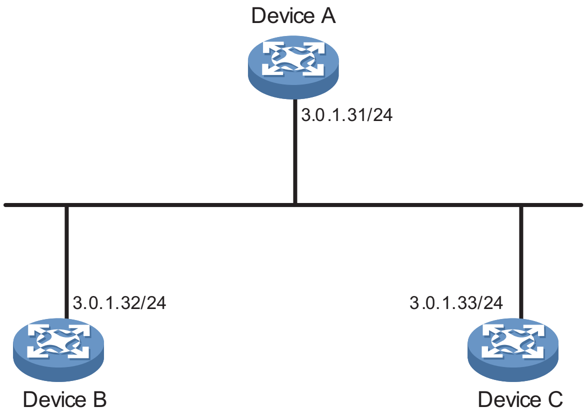
Configuring the NTP Symmetric Mode

Network requirements

The local clock of Device A is to be configured as a reference source, with the stratum level of 2. Device A is to be used as the NTP server of Device B, with Device B as the client, and Device A then as the server automatically. At the same time, Device B acts as the peer of Device C, with Device C as the symmetric-active peer and Device B as the symmetric-passive peer.

Network diagram

Figure 315 Network diagram for NTP symmetric peers mode configuration



Configuration procedure

1 Configuration on Device A:

# Specify the local clock as the reference source, with the stratum level of 2.

```
<DeviceA> system-view
[DeviceA] ntp-service refclock-master 2
```

2 Configuration on Device B:

# Specify Device A as the NTP server.

```
<DeviceB> system-view
[DeviceB] ntp-service unicast-server 3.0.1.31
```

3 Configuration on Device C (after Device B is synchronized to Device A):

# Specify the local clock as the reference source, with the stratum level of 1.

```
<DeviceC> system-view
[DeviceC] ntp-service refclock-master 1
```

# Configure Device B as a symmetric peer after local synchronization.

```
[DeviceC] ntp-service unicast-peer 3.0.1.32
```

In the step above, Device B and Device C are configured as symmetric peers, with Device C in the symmetric-active mode and Device B in the symmetric-passive mode. Because the stratus level of the local clock of Device C is 1 while that of Device B is 3, Device B is synchronized to Device C.

# View the NTP status of Device B after clock synchronization.

```
[DeviceB] 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, Device B has been synchronized to Device C, and the clock stratum level of Device B is 2, while that of Device C is 1.

# View the NTP session information of Device B, which shows that an association has been set up between Device B and Device C.

```
[DeviceB] 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
```

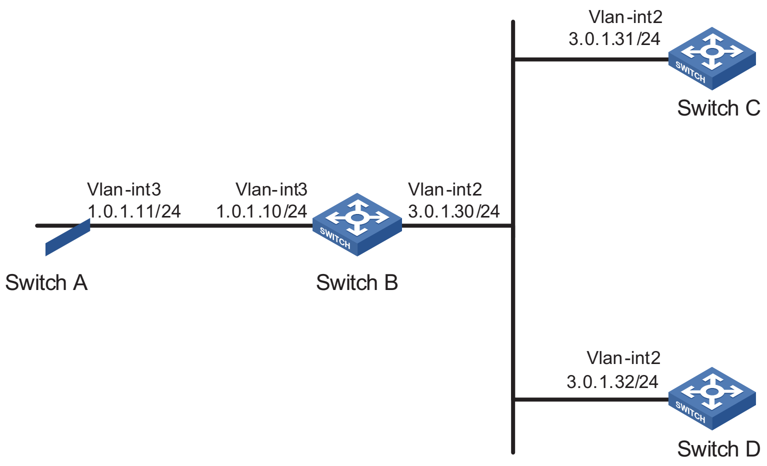
Configuring NTP Broadcast Mode

Network requirements

Switch C's local clock is to be used as a reference source, with the stratum level of 2, and Switch C sends out broadcast messages from VLAN-interface 2. Switch D and Switch A listen to broadcast messages through their own VLAN-interface 2 and VLAN-interface 3 respectively.

Network diagram

Figure 316 Network diagram for NTP broadcast mode configuration



Configuration procedure

1 Configuration on Switch C:

# Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 2
```

# Configure to send broadcast messages through VLAN-interface 2.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server
```

2 Configuration on Switch D:

# Enter system view.

```
<SwitchD> system-view
```

# Enter VLAN-interface 2 view.

```
[SwitchD] interface vlan-interface 2
```

# Specify Switch D as the broadcast client.

```
[SwitchD-Vlan-interface2] ntp-service broadcast-client
```

3 Configuration on Switch A:

# Enter system view.

```
<SwitchA> system-view
```

# Enter VLAN-interface 3 view

```
[SwitchA] interface vlan-interface 3
```

# Specify Switch A as the broadcast client.

```
[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] 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 A, 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] 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
```

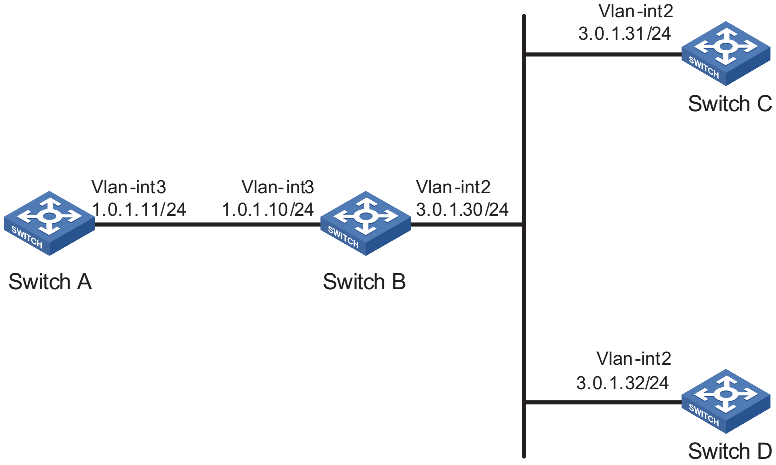
Configuring NTP Multicast Mode

Network requirements

Switch C's local clock is to be used as a reference source, with the stratum level of 2, and Switch C sends out multicast messages from VLAN-interface 2. Switch D and Switch A listen to multicast messages through VLAN-interface 2 and VLAN-interface 3 respectively.

Network diagram

Figure 317 Network diagram for NTP multicast mode configuration



Configuration procedure

1 Configuration on Switch C:

# Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 2
```

# Configure Switch C to work in the multicast server mode and send multicast messages through VLAN-interface 2.

```
[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 IGMP-enabled 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 IGMP 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 ethernet 4/1/3
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] igmp enable
[SwitchB-Vlan-interface3] quit
[SwitchB] interface ethernet 4/1/3
[SwitchB-Ethernet4/1/3] 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
```

Refer to “IGMP Configuration” on page 523 for how to configure IGMP.

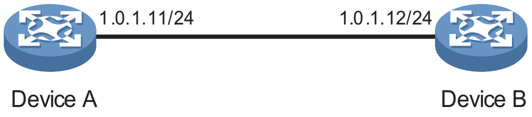
Configuring NTP Server/Client Mode with Authentication

Network requirements

The local clock of Device A is to be configured as a reference source, with the stratum level of 2. Device A is to be used as the NTP server of Device B, with Device B as the client and Device A then as the server automatically. At the same time, NTP authentication is to be enabled for Device A and Device B.

Network diagram

Figure 318 Network diagram for configuration of NTP server/client mode with authentication



Configuration procedure

1 Configuration on Device A:

# Specify the local clock as the reference source, with the stratum level of 2.

```
<DeviceA> system-view
[DeviceA] ntp-service refclock-master 2
```

2 Configuration on Device B:

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

# Enable NTP authentication on Device B.

```
[DeviceB] ntp-service authentication enable
[DeviceB] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
[DeviceB] ntp-service reliable authentication-keyid 42
[DeviceB] ntp-service unicast-server 1.0.1.11 authentication-keyid 42
```

Before Device B can synchronize its clock to that of Device A, you need to enable NTP authentication for Device A.

Perform the following configuration on Device A:

# Enable NTP authentication.

```
[DeviceA] ntp-service authentication enable
```

# Set an authentication key.

```
[DeviceA] ntp-service authentication-keyid 42 authentication-mode md
5 aNiceKey
```

# Specify the key as key as a trusted key.

```
[DeviceA] ntp-service reliable authentication-keyid 42
```

After the above configurations, Device B can be synchronized to Device A.

# View the NTP status of Device B after clock synchronization.

```
[DeviceB] 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, Device B has been synchronized to Device A, and the clock stratum level of Device B is 3, while that of Device A is 2.

# View the NTP session information of Device B, which shows that an association has been set up Device B and Device A.

```
[DeviceB] 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
```

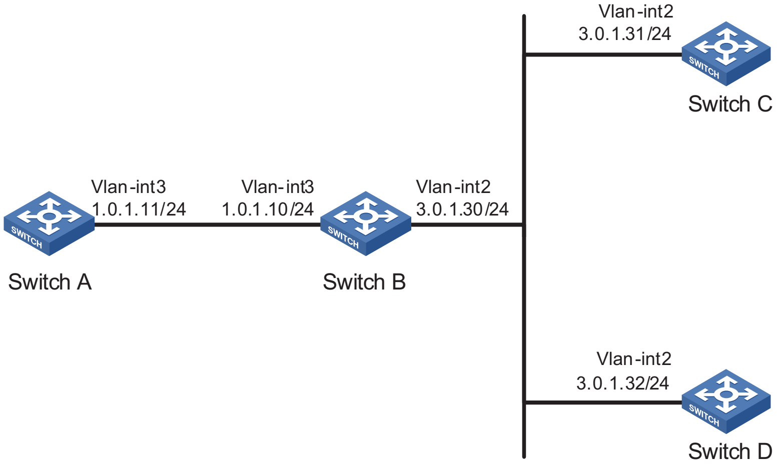
Configuring NTP Broadcast Mode with Authentication

Network requirements

Switch C's local clock is to be used as a reference source, with the stratum level of 2, and Switch C sends out broadcast messages from VLAN-interface 2. Switch D is to receive broadcast client through VLAN-interface 2, with NTP authentication enabled on both the server and client.

Network diagram

Figure 319 Network diagram for configuration of NTP broadcast mode with authentication



Configuration procedure

1 Configuration on Switch C:

# Specify the local clock as the reference source, with the stratum level of 3.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 3
```

# Configure NTP authentication

```
[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 ID.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server authentication-id 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 with that of Switch C.

# View the NTP status of Switch D after clock synchronization.

```
[SwitchD] 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 Device 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] 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
```

Configuring MPLS VPN Time Synchronization in Server/Client Mode

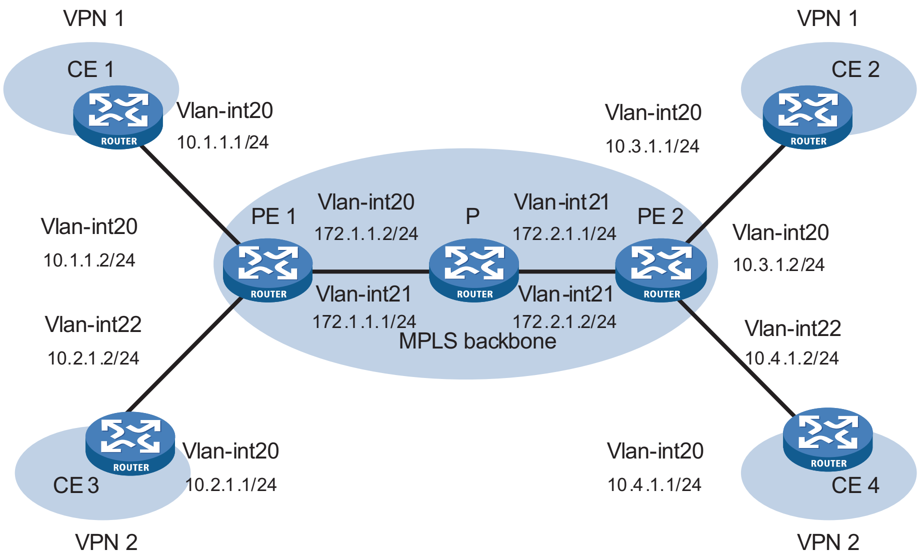
Network requirements

As shown in Figure 320, two VPNs are present on PE 1 and PE 2: VPN 1 and VPN 2. CE 1 and CE 2 are devices in VPN 1, while CE 3 and CE 4 are devices in VPN 2. It is required that CE 2 can be synchronized to CE 1 in the server/client mode. CE 1 is synchronized to the local reference source, with the clock stratum level being 1.

At present, MPLS VPN time synchronization can be implemented only in the unicast mode (server/client mode or symmetric peers mode), but not in the multicast or broadcast mode.

Network diagram

Figure 320 Network diagram for MPLS VPN time synchronization configuration



Configuration procedure

1 Configuration on CE 1:

# Specify the local clock as the reference source, with the stratum level of 1.

```
<CE1> system-view
[CE1] ntp-service refclcok-master 1
```

2 Configuration on CE 2:

# Specify CE 1 as the NTP server of CE 2 in VPN 1.

```
<CE2> system-view
[CE2] ntp-service unicast-server 10.1.1.1
```

# View the NTP session information and status information on CE 2 a certain period of time later. You can see that CE 2 has been synchronized to CE 1, with the clock stratum level being 2.

```
[CE2] display ntp-service status
 Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 10.1.1.1
 Nominal frequency: 63.9100 Hz
 Actual frequency: 63.9100 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 47.00 ms
 Root dispersion: 0.18 ms
 Peer dispersion: 34.29 ms
 Reference time: 02:36:23.119 UTC Jan 1 2001(BDFA6BA7.1E76C8B4)
[CE2] display ntp-service sessions
source          reference       stra reach poll  now offset  delay disper
**************************************************************************
[12345]10.1.1.1       LOCL          1     7    64    15     0.0    47.0     7.8
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
```

```
[CE2]  display ntp-service trace
 server 127.0.0.1,stratum 2, offset -0.013500, synch distance 0.03154
 server 10.1.1.1,stratum 1, offset -0.506500, synch distance 0.03429
 refid 127.127.1.0
```

Configuring MPLS VPN Time Synchronization in Symmetric Peers Mode

Network requirements

It is required that PE 2 can get synchronized to PE 1 in the symmetric peers mode, with PE 1 synchronized to the local reference source, having a clock stratum level of 1.

Network diagram

See Figure 320.

Configuration procedure

1 Configuration on PE 1:

# Specify the local clock as the reference source, with the stratum level of 1.

```
<PE1> system-view
[PE1] ntp-service refclcok-master 1
```

2 Configuration on PE 2:

# Specify PE 1 in VPN 1 as the symmetric-passive peer of PE 2.

```
<PE2> system-view
[PE2] ntp-service unicast-peer vpn-instance vpn1 10.1.1.2
```

# View the NTP session information and status information on PE 2 a certain period of time later. The information should show that PE 2 has been synchronized to PE 1, with the clock stratum level of 2.

```
[PE2] display ntp-service status
Clock status: synchronized
 Clock stratum: 2
 Reference clock ID: 10.1.1.2
 Nominal frequency: 63.9100 Hz
 Actual frequency: 63.9100 Hz
 Clock precision: 2^7
 Clock offset: 0.0000 ms
 Root delay: 32.00 ms
 Root dispersion: 0.60 ms
 Peer dispersion: 7.81 ms
 Reference time: 02:44:01.200 UTC Jan 1 2001(BDFA6D71.33333333)
[PE2] display ntp-service sessions
source          reference       stra reach poll  now offset  delay disper
**************************************************************************
[12345]10.1.1.2       LOCL          1    1   64   29    -12.0   32.0    15.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations :  1
[PE2] display ntp-service trace
 server 127.0.0.1,stratum 2, offset -0.012000, synch distance 0.02448
 server 10.1.1.2,stratum 1, offset 0.003500, synch distance 0.00781
 refid 127.127.1.0
```

86 NQA CONFIGURATION

The term router and the icon router in this document refer to a router in a generic sense or an Ethernet switch running routing protocols.

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

- "NQA Overview" on page 1083
- "Configuring NQA Tests" on page 1084
- "Configuring Optional Parameters for NQA Tests" on page 1103
- "Displaying and Maintaining NQA" on page 1106

NQA Overview

This section covers these topics:

- "Introduction to NQA" on page 1083
- "NQA Server and NQA Client" on page 1083
- "NQA Test Operation" on page 1084

Introduction to NQA

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. NQA (network quality analyzer) is an enhanced Ping tool used for testing the performance of protocols running on networks. Besides the Ping functions, NQA can provide the following functions:

- Detecting the availability and the response time of DHCP, FTP, HTTP, and SNMP services.
- Testing the delay jitter of the network.
- Verifying the availability of TCP, UDP, and DLSw packets.

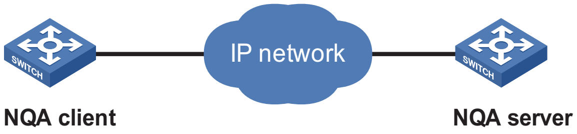
Different from Ping, NQA does not display the roundtrip time or time-out time of each packet on the console terminal in a realtime way. In this case, you have to execute the **display nqa results** command to view NQA test results. In addition, NQA can help you to set parameters for various tests and start these tests through the network management system (NMS).

For the detailed description on TCP, UDP, Jitter, ICMP, HTTP, FTP, DHCP, DLSw and SNMP, refer to the corresponding manuals.

NQA Server and NQA Client

In most NQA test systems, you only need to configure an NQA client. However, when you perform a TCP, UDP, or jitter test, you need to configure an NQA server. Figure 321 shows the relationship between an NQA client and an NQA server.

Figure 321 Relationship between NQA client and NQA server



The NQA server listens to test requests originated by the NQA client and makes a response to these requests. The NQA server can respond to requests originated by the NQA client only when the NQA server is enabled and the corresponding destination address and port number are configured on the server. The IP address and port number specified for a listening service on the server must be consistent with those on the client.

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.

NQA Test Operation

NQA can test multiple protocols. A test group must be created for each type of NQA test. Each test group can be related to only one type of NQA test. Each test group has an administrator name and an operation tag. The administrator name and the operation tag uniquely identify a test group.

After you create a test group and enter test group view, you can configure related test parameters. Test parameters vary with the test type. For details, see the configuration procedure below.

For optional parameters common to different types of tests, refer to "Configuring Optional Parameters for NQA Tests" on page 1103.

To perform an HW test successfully, proceed as follows:

1. Enable the NQA client.
2. Create a test group and configure test parameters according to the test type.
3. Perform the NQA test through the related **enable** command.
4. View the test results through the related **display** or **debugging** command.

i *After you enable the NQA client, you can create multiple test groups to perform tests. In this way, you do not need to enable the NQA client repeatedly.*

Configuring NQA Tests

i

- *You need to configure the NQA client and NQA server for TCP, UDP, and jitter tests, while you need to configure only the NQA client for other tests.*
- *You are not recommended to perform TCP, UDP, or jitter test on the ports from 1 to 1023 (known ports). Otherwise, the NQA probes will fail or the corresponding services of this known port will be unavailable.*

This section covers these topics:

- "Configuring the ICMP Test" on page 1085

- "Configuring the DHCP Test" on page 1087
- "Configuring the FTP Test" on page 1088
- "Configuring the HTTP Test" on page 1090
- "Configuring the Jitter Test" on page 1092
- "Configuring the SNMP Query Test" on page 1095
- "Configuring the TCP Test" on page 1097
- "Configuring the UDP Test" on page 1099
- "Configuring the DLSw Test" on page 1101

Configuring the ICMP Test

The ICMP test is mainly used to test whether an NQA client can send packets to a specified destination and test the roundtrip time of packets.

Configuration procedure

Follow these steps to configure the ICMP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to ICMP	**test-type icmp**	Optional ICMP by default.
Configure a destination address for a test	**destination-ip** *ip-address*	Required
Configure the size of test packets sent	**datasize** *size*	Optional 56 bytes by default.
Configure a string of fill characters of a test packet	**datafill** *text*	Optional The string of fill characters of an ICMP packet is the string corresponding with the ASCII code 00 to 09 by default.
Specify a VPN instance	**vpninstance** *name*	Optional No VPN instance is specified by default. When there are multiple VPNs, you need to use this command to specify a VPN instance for test.
Specify the IP address of an interface as the source IP address of an ICMP test request packet	**source-interface** *interface-type interface-number*	Optional The interface specified by this command can only be a layer 3 Ethernet interface or VLAN interface. In addition, the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103.	Optional
Enable the NQA test	**test-enable**	Required

To do...	Use the command...	Remarks
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

Configuration example

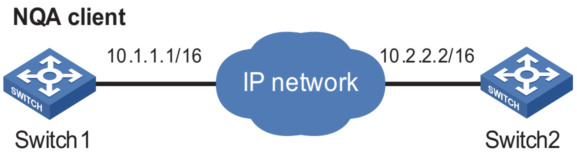
1 Network requirements

Use the NQA ICMP function to test whether the NQA client (Switch 1) can send packets to the specified destination (Switch 2) and test the roundtrip time of packets.

- Switch 1 serves as the NQA client, with the IP address being 10.1.1.1/16.
- Switch 2 serves as the device to be tested, with the IP address being 10.2.2.2/16.

2 Network diagram

Figure 322 Network diagram for the ICMP test



3 Configuration procedure

Perform the following configurations on Switch 1:

# Enable the NQA client, create an ICMP test group, and configure related test parameters.

```
<Switch1> system-view
[Switch1] nqa-agent enable
[Switch1] nqa admin icmp
[Switch1-nqa-admin-icmp] test-type icmp
[Switch1-nqa-admin-icmp] destination-ip 10.2.2.2
```

# Configure optional parameters.

```
[Switch1-nqa-admin-icmp] count 10
[Switch1-nqa-admin-icmp] timeout 5
```

# Enable the ICMP test.

```
[Switch1-nqa-admin-icmp] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch1-nqa-admin-icmp] display nqa results admin icmp
  NQA entry(admin admin, tag icmp) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 10             Receive response times: 10
      Min/Max/Average Round Trip Time: 1/3/1
      Square-Sum of Round Trip Time: 29
      Last succeeded test time: 2009-08-15 15:02:03.0
```

```
Extend result:
  Packet lost in test: 0%
  Failures due to Timeout: 0
  Failures due to System Busy: 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
```

Configuring the DHCP Test

The 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

The interface specified by the **source-interface** command must be up.

Before the DHCP test, you need to perform some configurations on the DHCP server. For example, you need to enable the DHCP service and configure an address pool. If the NQA (DHCP) client and the DHCP server are not in the same network segment, you need to configure a DHCP relay. For detailed configurations, refer to *“DHCP Server Configuration” on page 721*.

Configuration procedure

Follow these steps to configure the DHCP test

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to DHCP	**test-type dhcp**	Required
Specify an interface for a DHCP test	**source-interface** *interface-type interface-number*	Required The interface in the command must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to “Configuring Optional Parameters for NQA Tests” on page 1103	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

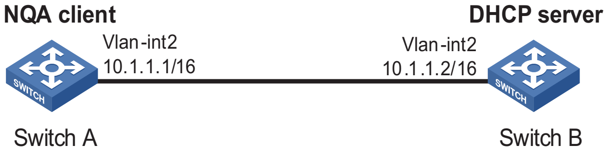
Configuration example

1 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.

2 Network diagram

Figure 323 Network diagram for the DHCP test



3 Configuration procedure

For the configuration of DHCP Server, refer to "DHCP Server Configuration" on page 721.

Perform the following configurations on Switch A:

# Enable the NQA client, create a DHCP test group, and configure related test parameters.

```
<SwitchA> system-view
[SwitchA] nqa-agent enable
[SwitchA] nqa admin dhcp
[SwitchA-nqa-admin-dhcp] test-type dhcp
[SwitchA-nqa-admin-dhcp] source-interface Vlan-interface 2
```

# Enable the DHCP test.

```
[SwitchA-nqa-admin-dhcp] test-enable
```

# View the test results with the **display nqa results** command.

```
[SwitchA-nqa-admin-dhcp] display nqa results admin dhcp
  NQA entry(admin admin, tag dhcp) test result:
      Send operation times: 1             Receive response times: 1
      Min/Max/Average Round Trip Time: 527/527/527
      Square-Sum of Round Trip Time: 277729
      Last succeeded test time: 2006-06-07 13:15:07.3
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the FTP Test

The FTP test is mainly used to test the connection with a specified FTP server and the time necessary for the FTP client to transfer a file to the FTP server.

Configuration prerequisites

Before the 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 configurations, refer to *"FTP Configuration" on page 1031*.

Configuration procedure

Follow these steps to configure the FTP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to FTP	**test-type ftp**	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required Here it is the IP address of the FTP server.
Configure the source IP address of a test request packet	**source-ip** *ip-address*	Required 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	**ftp-operation** { **get** | **put** }	Optional **get** by default, meaning to get files from the FTP server.
Configure a login username	**username** *name*	Required
Configure a login password	**password** *password*	Optional
Specify a file to be transferred between the FTP server and the FTP client.	**filename** *file-name*	Required
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

- *Transfer a small file for a get operation. If the file is too large, the test may fail because of time-out.*
- *When you perform a get operation, the file obtained from the FTP server will not be saved on the device, either. If there is no such file-name file on the FTP server, the FTP test will fail.*
- *When you perform a put operation, a file-name file with a fixed size and contents will be created on the FTP server, but the uploaded file will not be saved.*

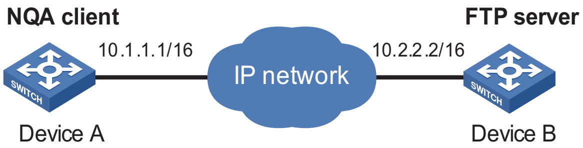
Configuration example

1 Network requirements

Use the NQA FTP function to test the connection with a specified FTP server and the time necessary for the FTP client to upload a file to the FTP server. The login username is admin, the login password is nqa, and the file to be transferred to the FTP server is config.txt.

2 Network diagram

Figure 324 Network diagram for the FTP test



3 Configuration procedure

For the configuration of FTP Server, refer to "FTP Configuration" on page 1031.

Perform the following configurations on Device A:

# Enable the NQA client, create an FTP test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin ftp
[Switch-nqa-admin-ftp] test-type ftp
[Switch-nqa-admin-ftp] destination-ip 10.2.2.2
[Switch-nqa-admin-ftp] source-ip 10.1.1.1
[Switch-nqa-admin-ftp] ftp-operation put
[Switch-nqa-admin-ftp] username admin
[Switch-nqa-admin-ftp] password nqa
[Switch-nqa-admin-ftp] filename config.txt
```

# Enable the FTP test.

```
[Switch-nqa-admin-ftp] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-ftp] display nqa results admin ftp
  NQA entry(admin admin, tag ftp) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average Round Trip Time: 191/191/191
      Square-Sum of Round Trip Time: 36481
      Last succeeded test time: 2000-06-07 13:21:23.9
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the HTTP Test

The HTTP test is mainly used to test the connection with a specified HTTP server and the time required to obtain data from the HTTP server.

Configuration procedure

Follow these steps to configure the HTTP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to HTTP	**test-type http**	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required Here it is the IP address of the HTTP server.
Configure the HTTP operation type	**http-operation** { **get** | **post** }	Optional **get** by default, meaning to get data from the HTTP server.
Configure an HTTP operation string	**http-string** *string version*	Required
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

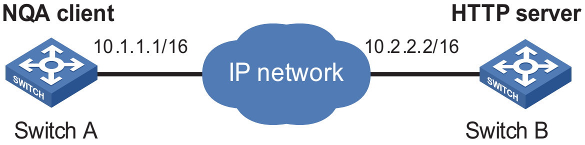
Configuration example

1 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.

2 Network diagram

Figure 325 Network diagram for the HTTP test



3 Configuration procedure

Perform the following configurations on Switch A:

# Enable the NQA client, create an HTTP test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin http
[Switch-nqa-admin-http] test-type http
[Switch-nqa-admin-http] destination-ip 10.2.2.2
```

```
[Switch-nqa-admin-http] http-operation get
[Switch-nqa-admin-http] http-string /index.htm HTTP/1.0
```

# Enable the HTTP test.

```
[Switch-nqa-admin-http] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-http] display nqa results admin http
  NQA entry(admin admin, tag http) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average Round Trip Time: 15/15/15
      Square-Sum of Round Trip Time: 225
      Last succeeded test time: 2006-12-28 11:01:07.6
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the Jitter Test

CAUTION: *You are not recommended to perform an NQA jitter test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.*

The jitter test is used to take statistics of delay jitter of UDP packet transmission. Delay jitter refers to the difference between the interval of receiving two packets consecutively and the interval of sending these two packets. During the test, the source port sends data packets to the destination port at regular intervals. The destination port affixes a time stamp to each packet that it receives and then sends it back to the source port. After the source port receives the data packet, the delay jitter can be calculated.

To improve the accuracy of the statistics results, you must send multiple test packets when you perform a test. The more test packets are sent, the more accurate the statistics results are. However, it takes a longer time to complete the test. You can quicken a jitter test by reducing the interval of sending test packets. However, doing so will cause an impact on the network.

The error in the statistics results of a jitter test is big since there is a delay in both sending and receiving data packets.

A jitter test requires cooperation between the NQA server and the NQA client. You must configure the UDP listening function on the NQA server, and a destination address and a destination port on the NQA client, and ensure that the destination address and destination port on the NQA client are respectively the listening IP address and port on the NQA server.

Configuration procedure

1 Configure the NQA server

Follow these steps to configure the NQA server for a jitter test:

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 listening function on the NQA server	**nqa-server udpecho** *ip-address port-number*	Required The listening IP address and port number must be the destination IP address and port on the NQA client.

2 Configure the NQA client

Follow these steps to configure the NQA client for a jitter test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to jitter	**test-type jitter**	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required The destination address is the listening IP address on the NQA server.
Configure a destination port	**destination-port** *port-number*	Required The destination port is the listening port on the NQA server.
Configure the number of jitter test packets sent in a probe	**jitter-packetnum** *number*	Optional 10 by default.
Configure the interval for sending jitter test packets	**jitter-interval** *interval*	Optional 20 ms by default.
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103.	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.
View the recorded delay jitter of UDP packet transmission in the last NQA jitter test	**display nqa jitter** [*admin-name operation-tag*]	Optional You can execute the command in any view.

The number of probes made in a jitter test depends on the ***count*** *command, while the number of test packets sent in each probe depends on the* ***jitter-packetnum*** *command.*

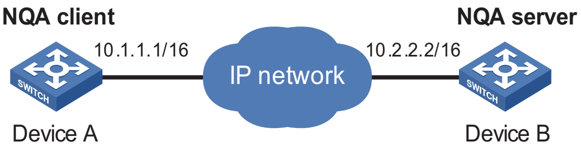
Configuration example

1 Network requirements

Use the NQA jitter function to test the delay jitter of packet transmission between the local port (Device A) and the specified destination port (Device B).

2 Network diagram

Figure 326 Network diagram for the jitter test



3 Configuration procedure

- Configuration on Device B

# Enable the NQA server and configure the listening IP address and port number.

```
<DeviceB> system-view
[DeviceB] nqa-server enable
[DeviceB] nqa-server udpecho 10.2.2.2 9000
```

- Configuration on Device A

# Enable the NQA client, create a jitter test group, and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa-agent enable
[DeviceA] nqa admin jitter
[DeviceA-nqa-admin-jitter] test-type jitter
[DeviceA-nqa-admin-jitter] destination-ip 10.2.2.2
[DeviceA-nqa-admin-jitter] destination-port 9000
```

# Enable the jitter test.

```
[DeviceA-nqa-admin-jitter] test-enable
```

# View the test results with the **display nqa results** and **display nqa jitter** commands.

```
[DeviceA-nqa-admin-jitter] display nqa results admin jitter
  NQA entry(admin admin, tag jitter) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 10             Receive response times: 10
      Min/Max/Average Round Trip Time: 1/9/2
      Square-Sum of Round Trip Time: 114
      Last succeeded test time: 2009-08-15 15:19:10.9
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
    Jitter result:
      RTT Number: 10
      SD Maximal delay: 4                    DS Maximal delay: 4
      Min Positive SD: 1                     Min Positive DS: 0
      Max Positive SD: 1                     Max Positive DS: 0
      Positive SD Number: 1                  Positive DS Number: 0
      Positive SD Sum: 1                     Positive DS Sum: 0
      Positive SD average: 0                 Positive DS average: 0
      Positive SD Square Sum: 1              Positive DS Square Sum: 0
      Min Negative SD: 1                     Min Negative DS: 1
      Max Negative SD: 6                     Max Negative DS: 1
      Negative SD Number: 2                  Negative DS Number: 1
      Negative SD Sum: 7                     Negative DS Sum: 1
      Negative SD average: 4                 Negative DS average: 1
      Negative SD Square Sum: 37             Negative DS Square Sum: 1
      SD lost packets number: 0              DS lost packet number: 0
      Unknown result lost packet number: 0
```

Configuring the SNMP Query Test

The 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.

Configuration procedure

Follow these steps to configure the SNMP query test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to SNMP query	**test-type snmpquery**	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required
Configure common optional parameters	Refer to “Configuring Optional Parameters for NQA Tests” on page 1103.	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

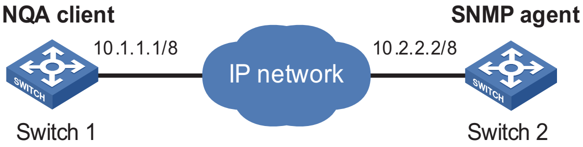
Configuration example

1 Network requirements

Use the NQA SNMP query function to test the time it takes Switch 1 to send an SNMP query packet to the SNMP agent and receive a response packet.

2 Network diagram

Figure 327 Network diagram for the SNMP query test



3 Configuration procedure

- Configure Switch 2.

# Enable the SNMP agent service and set the SNMP version to v2c, the read community to public, and the write community to private.

```
<Switch> system-view
[Switch] snmp-agent sys-info version v2c
[Switch] snmp-agent community read public
[Switch] snmp-agent community write private
```

- *SNMP must be enabled on the SNMP agent. Otherwise, no response packet will be received.*
- *In this example, the configuration is based on the SNMP v2c. If the SNMP of other versions is enabled, the configuration may be different. For details, "SNMP Configuration" on page 1043.*

- Configure Switch 1:

# Enable the NQA client, create an SNMP query test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin snmp
[Switch-nqa-admin-snmp] test-type snmpquery
[Switch-nqa-admin-snmp] destination-ip 10.2.2.2
```

# Enable the SNMP query test.

```
[Switch-nqa-admin-snmp] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-snmp] display nqa results admin snmp
  NQA entry(admin admin, tag snmp) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average Round Trip Time: 5/5/5
      Square-Sum of Round Trip Time: 25
      Last succeeded test time: 2006-06-09 11:19:28.2
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the TCP Test

***CAUTION:** You are not recommended to perform an NQA TCP test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.*

The TCP test is used to test the TCP connection between the client and the specified server and the setup time for the connection.

The TCP test includes TCP-Public test and TCP-Private test.

- For the TCP-Public test, a connection setup request is permanently initiated to TCP port 7 of the destination address. No destination port needs to be configured on the client, but TCP port 7 used for listening needs to be configured on the server. Even if a port is configured on the client, the port does not take effect.
- For the TCP-Private test, a connection setup request is initiated to the specified port of the destination address.

Configuration procedure

1 Configure the NQA server

Follow these steps to configure the NQA server for the TCP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA server	**nqa-server enable**	Required Disabled by default
Configure the TCP listening function on the NQA server	**nqa-server tcpconnect** *ip-address port-number*	Required The listening IP address and port number must be the destination IP address and port on the NQA client. If the test type is TCP-Public, the port number must be set to 7.

2 Configure the NQA client

Follow these steps to configure NQA client for the TCP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to TCP	**test-type** { **tcpprivate | tcppublic** }	Required

To do...	Use the command...	Remarks
Configure a destination address for a test	**destination-ip** *ip-address*	Required The destination address must be the listening IP address on the NQA server.
Configure a destination port	**destination-port** *port-number*	If the test type is TCP-Public, no port needs to be configured. If the test type is TCP-Private, a port must be configured and it must be the listening port configured on the NQA server.
Configure common optional parameters	"Configuring Optional Parameters for NQA Tests" on page 1103	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

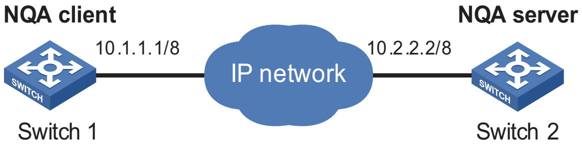
Configuration example

1 Network requirements

Use the NQA TCP-Private function to test the time for setting up a TCP connection between the local port (Switch 1) and the specified destination port (Switch 2). The port number used is 9000.

2 Network diagram

Figure 328 Network diagram for the TCP-Private test



3 Configuration procedure

- Configuration on Switch 2

# Enable the NQA server and configure the listening IP address and port number.

```
<Switch> system-view
[Switch] nqa-server enable
[Switch] nqa-server tcpconnect 10.2.2.2 9000
```

- Configuration on Switch 1

# Enable the NQA client, create a TCP test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin tcpprivate
[Switch-nqa-admin-tcpprivate] test-type tcpprivate
[Switch-nqa-admin-tcpprivate] destination-ip 10.2.2.2
[Switch-nqa-admin-tcpprivate] destination-port 9000
```

# Enable the TCP test.

```
[Switch-nqa-admin-tcpprivate] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-tcpprivate] display nqa results admin tcpprivate
  NQA entry(admin admin, tag tcpprivate) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average Round Trip Time: 1/1/1
      Square-Sum of Round Trip Time: 1
      Last succeeded test time: 2009-08-15 15:24:34.8
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the UDP Test

CAUTION: *You are not recommended to perform an NQA UDP test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.*

The UDP test is used to test the roundtrip time of a UDP packet from the client to the specified server.

The UDP test includes UDP-Public test and UDP-Private test.

- For the UDP-Public test, a connection setup request is permanently initiated to UDP port 7 of a destination address. No port needs to be configured on the client, but port 7 for listening needs to be configured on the server. Even if a port is configured on the client, the port does not take effect.
- For the UDP-Private test, a connection setup request is initiated to the specified port of the destination address.

Configuration procedure

1 Configure the NQA server

Follow these steps to configure the NQA server for the UDP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA server	**nqa-server enable**	Required Disabled by default.

To do...	Use the command...	Remarks
Configure the UDP listening function on the NQA server	**nqa-server udpecho** *ip-address port-number*	Required The listening IP address and port number must be the destination IP address and port on the NQA client. If the test type is UDP-Public, the port number must be set to 7.

2 Configure the NQA client

Follow these steps to configure the NQA client for the UDP test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to UDP	**test-type** { **udpprivate** | **udppublic** }	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required The destination address must be the listening IP address configured on the NQA server.
Configure a destination port	**destination-port** *port-number*	If the test type is UDP-Public, no port needs to be configured. If the test type is UDP-Private, a port must be configured and it must be the listening port configured on the NQA server.
Configure the size of test packets sent	**datasize** *size*	Optional 100 bytes by default.
Configure a string of fill characters of a test packet	**datafill** *text*	Optional The string of fill characters of a UDP packet is the string corresponding with the ASCII code 00 to FF by default.
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

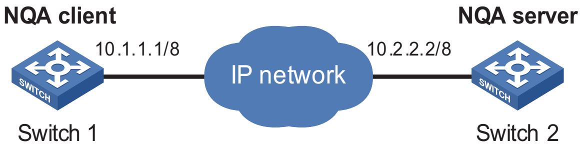
Configuration example

1 Network requirements

Use the NQA UDP-Private function to test the setup time for the UDP connection between the local port (Switch 1) and the specified destination port (Switch 2). The port number used is 8000.

2 Network diagram

Figure 329 Network diagram for the UDP-Private test



3 Configuration procedure

- Configuration on Switch 2

# Enable the NQA server and configure the listening IP address and port number.

```
<Switch> system-view
[Switch] nqa-server enable
[Switch] nqa-server udpecho 10.2.2.2 8000
```

- Configuration on Switch 1

# Enable the NQA client, create a UDP-Private test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin udpprivate
[Switch-nqa-admin-udpprivate] test-type udpprivate
[Switch-nqa-admin-udpprivate] destination-ip 10.2.2.2
[Switch-nqa-admin-udpprivate] destination-port 8000
```

# Enable the TCP test.

```
[Switch-nqa-admin-udpprivate] test-enable
```

# View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-udpprivate] display nqa results admin udpprivate
  NQA entry(admin admin, tag udpprivate) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1             Receive response times: 1
      Min/Max/Average Round Trip Time: 11/11/11
      Square-Sum of Round Trip Time: 121
      Last succeeded test time: 2009-08-15 15:26:01.8
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring the DLSw Test

Data link switching (DLSw) was jointly developed by Advanced Peer-to-Peer Networking (APPN) Implementers Workshop (AIW) and the Data-Link Switching Related Interest Group (DLSw RIG) for transmitting Systems Network Architecture

(SNA) traffic over a TCP/IP network. The DLSw test is used to test the response time of the DLSw device.

Configuration prerequisites

Before the DLSw test, a TCP connection can be set up between the NQA client and the specified device.

Configuration procedure

Follow these steps to configure the DLSw test:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the NQA client	**nqa-agent enable**	Required
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	-
Set the test type to DLSw	**test-type dlsw**	Required
Configure a destination address for a test	**destination-ip** *ip-address*	Required
Configure common optional parameters	Refer to "Configuring Optional Parameters for NQA Tests" on page 1103.	Optional
Enable the NQA test	**test-enable**	Required
View the test results	**display nqa results** [*admin-name operation-tag*]	Required You can execute the command in any view.

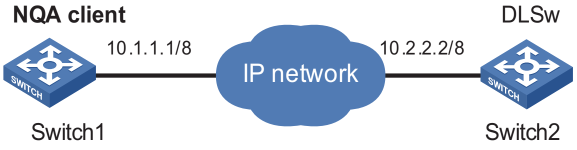
Configuration example

1 Network requirements

Use the NQA DLSw function to test the response time of the DLSw device.

2 Network diagram

Figure 330 Network diagram for the DLSw test



3 Configuration procedure

Enable the NQA client, create a DLSw test group, and configure related test parameters.

```
<Switch> system-view
[Switch] nqa-agent enable
[Switch] nqa admin dlsw
[Switch-nqa-admin-dlsw] test-type dlsw
[Switch-nqa-admin-dlsw] destination-ip 10.2.2.2
```

# Enable the DLSw test.

```
[Switch-nqa-admin-dlsw] test-enable
```

View the test results with the **display nqa results** command.

```
[Switch-nqa-admin-dlsw] display nqa results admin dlsw
  NQA entry(admin admin, tag dlsw) test result:
    Destination ip address: 10.2.2.2
      Send operation times: 1              Receive response times: 1
      Min/Max/Average Round Trip Time: 5/5/5
      Square-Sum of Round Trip Time: 25
      Last succeeded test time: 2006-06-07 13:25:45.1
    Extend result:
      Packet lost in test: 0%
      Failures due to Timeout: 0
      Failures due to System Busy: 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
```

Configuring Optional Parameters for NQA Tests

Unless otherwise specified, the following parameters are applicable to all test types and they can be configured according to the actual conditions. Optional parameters common to NQA are valid for all NQA tests, while those common to an NQA test group are valid only for tests in this test group. This section covers these topics:

- "Configuring Optional Parameters Common to NQA" on page 1103
- "Configuring Optional Parameters Common to an NQA Test Group" on page 1103
- "Configuring Trap Delivery" on page 1105

Configuring Optional Parameters Common to NQA

Follow these steps to configure optional parameters common to NQA:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Configure the maximum number of tests that the NQA client can simultaneously perform	**nqa-agent max-requests** *max-number*	Optional 5 by default

Configuring Optional Parameters Common to an NQA Test Group

Follow these steps to configure the 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** *admin-name operation-tag*	-
Configure a descriptive string for a test group	**description** *text*	Optional No descriptive string by default.

To do...	Use the command...	Remarks
Configure the interval of performing a cyclic test	**frequency** *interval*	Optional No cyclic test is performed by default. This command is invalid for the DHCP test.
Configure the number of probes in a test	**count** *times*	Optional 1 by default. For the TCP test, a probe means a connection. For the jitter test, the number of test packets sent in a probe is determined by the **jitter-packetnum** command. For the SNMP test, three test packets are sent in a probe. For the other tests, one test packet is sent in a probe.
Configure the NQA probe time-out time	**timeout** *time*	Optional Three seconds by default. If no response packet is received within the time-out time of a request packet, the probe fails.
Configure the maximum number of history records that can be saved in a test group	**history-records** *number*	Optional 50 by default If the number of history records exceeds this value, the earliest test results are discarded.
Configure the maximum number of hops a test request packet traverses in the network	**ttl** *number*	Optional 20 by default. This command is invalid for the DHCP test.
Configure the ToS field in an IP packet header	**tos** *value*	Optional 0 by default. This command is invalid for the DHCP test.
Configure the source IP address of a test request packet	**source-ip** *ipaddress*	This command is required for the FTP test but optional for other tests. You can specify an IP address as the source IP address of a test request packet. Otherwise, the IP address of the interface sending packets serves as the source IP address of the test request packet. The source IP address in the command must be the IP address of an interface on the device and the interface must be up. Otherwise, the test will fail. This command is invalid for the DHCP test.

To do...	Use the command...	Remarks
Configure the source port of a test request packet	**source-port** *port-number*	Optional You can specify a port as the source port of a test request packet. Otherwise, the system automatically assigns a port to serve as the source port of the test request packet. This command is only valid for jitter, UDP, and SNMP tests.
Enable the routing table bypass function	**sendpacket passroute**	Optional Disabled by default. If you want to test the connectivity between the local address and the destination address, you can enable this function. After this function is enabled, the routing table will not be searched, and the packet is directly sent to the destination in the directly connected network. If the destination is not in the directly connected network, an error will be prompted. This command is invalid for the DHCP test.

Configuring Trap Delivery

Configuration prerequisites

Before configuring Trap delivery, you should configure the address of the network management server which receives the Trap message. For detailed configuration procedure, refer to *"Trap Configuration" on page 1046*.

Configuring Trap delivery

A trap message is generated no matter whether an NQA test succeeds or fails. You can set a switch to control the delivery of the trap message to the network management server.

Follow these steps to configure Trap:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Create an NQA test group and enter its view	**nqa** *admin-name operation-tag*	Required
Enable trap debugging to send a trap message to the network management server	**send-trap** { **all** | { **probefailure** | **testcomplete** | **testfailure** }* }	Optional No trap message is sent to the network management server by default.
Configure the minimum number of probe failures in an NQA test before a test failure trap message is sent	**test-failtimes** *times*	Optional 1 by default.

To do...	Use the command...	Remarks
Configure the number of consecutive probe failures in an NQA test before a trap message is sent to indicate a probe failure	**probe-failtimes** *times*	Optional 1 by default.

Displaying and Maintaining NQA

To do...	Use the command...	Remarks
Display history information of tests	**display nqa history** [*admin-name operation-tag*]	Available in any view
Display the results of the last NQA jitter test	**display nqa jitter** [*admin-name operation-tag*]	Available in any view
Display the results of the last test	**display nqa results** [*admin-name operation-tag*]	Available in any view

87 HIGH AVAILABILITY CONFIGURATION

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

- "Introduction to HA" on page 1107
- "Configuring HA" on page 1107
- "Displaying and Maintaining HA" on page 1108
- "HA Configuration Example" on page 1108

Introduction to HA

High Availability (HA for short) feature can be used to achieve a higher degree of system availability. Devices supporting HA are normally equipped with two Switching and Routing Process Units (Fabric for short), with one being the active card that works under master mode and the other being the standby card that works under slave mode. With the synchronization function standby card can work under the same system configuration. Thus, in the event of an active card failure, the standby card will immediately switch over to function as an active card (referred to as switchover hereafter), ensuring that the device works properly and its configuration is synchronized with that of the active card.

Switchover occurs in the following cases:

- Active card failure
- The active card is plugged out
- SNMP
- Manual switchover

A switchover takes place in the following order in the event of an active card failure:

1. The standby card automatically connects and controls system bus while the original active card disconnects from it.
2. The original standby card becomes the new active card whereas the original active card automatically reboots and functions as the new standby card.

We discuss switchover through user command line in this chapter.

Configuring HA

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the manual switchover between the active card and standby card	**slave switchover** { **enable** | **disable** }	Optional Enabled by default

To do...	Use the command...	Remarks
Manually configure switchover between the active card and standby card	**slave switchover**	Optional
Manually configure the standby card restart	**slave restart**	Optional
Enable Full Mesh forwarding enhance mode	**fullmesh-enhance** { **enable** | **disable** }	Optional Disabled by default.

CAUTION:

- *The standby card does not support any system configuration commands. Therefore, users cannot execute any commands in the standby card unless it has replaced the original active card and has become the new active card.*
- *After the standby card has restarted, the active card will perform initial synchronization on the standby card. During this process if the user presses <Enter>, the system will prompt the user no command can be input. After the initial synchronization is completed, the user can execute all the configuration commands in the active card and the active card and standby card will keep a real-time synchronization process, meaning the configuration of the user on the active card will be copied to the standby card to ensure the consistency of the current configuration of the active card and standby card.*
- *A switch over capable device cannot be updated with a software specific with centralized device as this will result in an unavailable system.*
- *If only one Fabric is in the slot, the Full Mesh enhance mode does not take effect.*
- *The Switch 8807 switch does not support the **fullmesh-enhance** command.*
- *When the system works in load balancing mode, it activates the Full Mesh enhance mode if this mode is configured.*
- *When the system works in active and standby mode, it does not activate the Full Mesh enhance mode if this mode is configured. In this case, you need to switch the system to the load balancing mode to activate the Full Mesh enhance mode.*

Displaying and Maintaining HA

To do...	Use the command...	Remarks
Display the switchover state	**display switchover state** [*slot-id*]	Available in any view
Display Full Mesh forwarding information	**display fullmesh-enhance**	Available in any view

HA Configuration Example

Dynamic Upgrade of the System Through HA

Configuration prerequisites

Dynamic upgrade of image files of the active card and standby card.

Configuration procedure

1 Download the update software

Through remote online update commands, download new application program to the active card. Use the FTP, TFTP, or XModem to download the application program to the active card and save it in the flash.

2 Copy the software to the standby card

Assume the update application is platform.app, slot0 is the active card, and slot1 is the standby card.

```
<Sysname> copy platform.app slot1#flash:/platform.app
```

3 Designate the newly downloaded update application to be the executive software of the active card and standby card.

```
<Sysname> boot-loader file flash:/platform.app slot 0 main
<Sysname> boot-loader file slot1#flash:/platform.app slot 1 main
```

4 Restart the standby card.

```
<Sysname> system-view
[Sysname] slave restart
The slave will reset! Continue?[Y/N]:y
```

5 Copy the configuration file to the standby card

If the standby card is working properly, system will prompt that it is in an active state. As long as the auto-update function is enabled, the configuration file will be copied to the standby card while the active card is saving the same file.

```
[Sysname] slave auto-update config
[Sysname] quit
<Sysname> save
```

6 Manually configure the switch over between the active card and standby card.

```
<Sysname> system-view
[Sysname] slave switchover
Caution!!! Confirm switch slave to master[Y/N]?y
Starting.....
RAM Line....OK
```

After the switchover, the original active card will reset, restart, and update its application file. Thus, the whole system has its application upgraded dynamically.

Full Mesh Forwarding Mode Configuration Example

Configuration prerequisites

- Two Fabrics work in active and standby mode.
- Set the Full Mesh forwarding to enhance mode to optimize the Full Mesh forwarding performance in the system.

Configuration procedure

# Enter system view.

```
<Sysname> system-view
```

# Switch to load balancing mode.

```
[Sysname] xbar load-balance
# Enable Full Mesh mode.
[Sysname] fullmesh-enhance enable
```

# Check whether Full Mesh mode is activated.

```
[Sysname] display fullmesh-enhance
Configuration status: Enabled
Operation status: Active
```

88 INFORMATION CENTER CONFIGURATION

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

- "Information Center Overview" on page 1111
- "Configuring Information Center" on page 1117
- "Displaying and Maintaining Information Center" on page 1123
- "Information Center Configuration Examples" on page 1123

Information Center Overview

Introduction to Information Center

Acting as the system information hub, information center classifies and manages system information. Together with the debugging functionality, information center offers a powerful support for network administrators and developers in monitoring network performance and diagnosing network problems.

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.

The information center of the system has the following features:

Classification of system information

The system is available with three types of information:

- Log information
- Trap information
- Debug information

Eight levels of system information

The information is classified into eight levels by severity and can be filtered by level. More emergent information has a smaller severity level.

Table 42 Severity description

Severity	Severity value	Description
emergencies	0	The system is unavailable.
alerts	1	Information that demands prompt reaction
critical	2	Critical information

Table 42 Severity description

Severity	Severity value	Description
errors	3	Error information
warnings	4	Warnings
notifications	5	Normal errors with important information
informational	6	Informational information to be recorded
debugging	7	Information generated during debugging

Information filtering by severity works this way: information with the severity value greater than the configured threshold is not output during the filtering.

- If the threshold is set to 0, only information with the severity being emergencies will be output;
- If the threshold is set to 7, information of all severities will be output.

Ten channels and seven output directions of system information

The system supports seven information output directions, including the Console, console terminal (monitor), logbuffer, loghost, trapbuffer, SNMP and logfile.

The system can support ten channels. The channels 0 through 5, and channel 9 have their default channel names and are associated with seven output directions by default. Both the names of the routers and the associations between the channels and output directions can be changed through commands.

Table 43 Information channels and output directions

Information channel number	Default channel name	Default output direction
0	console	Console (Receives log, trap and debug information)
1	monitor	Monitor terminal (Receives log, trap and debug information, facilitating remote maintenance)
2	loghost	Log host (Receives log, trap and debug 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.)
4	logbuffer	Log buffer (Receives log information, a buffer inside the router for recording information.)
5	snmpagent	SNMP NMS (Receives trap information)
6	Not specified	Not specified (Receives log, trap, and debug information)

Table 43 Information channels and output directions

Information channel number	Default channel name	Default output direction
7	Not specified	Not specified (Receives log, trap, and debug information)
8	Not specified	Not specified (Receives log, trap, and debug information)
9	channel9	Log file (Receives log, trap, and debug information)

Configurations for the seven output directions 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, module drivers, and configuration modules. The information can be classified and filtered by source module. Some module names and description are shown in Table 44.

Table 44 Module name list

Module name	Description
8021X	802.1X module
ACL	Access Control List module
ADBM	MAC address management module
ARP	Address Resolution Protocol module
BGP	Border Gateway Protocol module
CFM	Configuration File Management module
CLST	Cluster Configuration module
CMD	Command line module
COMMONSY	Common System MIB module
default	Default setting of all modules
DEV	Device management module
DHCP	Dynamic Host Configuration Protocol module
DIAGCLI	Diagnosis module
DNS	Domain Name System module
DRVMPLS	Multiprotocol label switching driver module
DRVL2	Layer 2 driver module
DRVL3	Layer 3 driver module
DRVL3MC	Layer 3 multicast module
MPLS	Multiprotocol Label Switching module
DRVPOS	POS driver module
DRVQACL	QACL driver module
DRVVPLS	Virtual Private LAN Service driver module
ETH	Ethernet module
FTPS	FTP Server module
GARP	Generic Attribute Registration Protocol module
HA	High Availability module

Table 44 Module name list

Module name	Description
HABP	3Com Authentication Bypass Protocol module
Switch ClusteringS	3Com Group Management Protocol Service module
HWCM	3Com Configuration Management MIB module
IFNET	Interface management module
IGSP	IGMP Snooping module
IP	Internet Protocol module
ISIS	Intermediate System-to-Intermediate System intra-domain routing information exchange protocol module
L2INF	Interface management module
L2V	L2 VPN module
LACL	LAN switch ACLmodule
LAGG	Link Aggregation module
LDP	Label Distribution Protocol module
LINE	Line module
LINKAGG	LINK AGG module
LQOS	LAN switch QoS module
LS	Local Server module
LSPAGENT	Label Switched Path Agent module
LSPM	Label Switch Path Management module
MIX	Dual main control network management module
MMC	MMC module
MODEM	MODEM module
MPLSFW	Multi-protocol Label Switch Forward module
MPM	Multicast Port Management module
MSDP	Multicast Source Discovery Protocol module
MSTP	Multiple Spanning Tree Protocol module
NAT	Network Address Translation module
NTP	Network Time Protocol module
PKI	Public Key Infrastructure module
OSPF	Open Shortest Path First module
PHY	Physical Sublayer & Physical Layer module
POE	Power over Ethernet module
POS_SNMP	POS Simple Network Management Protocol module
PPP	Point to Point Protocol module
PSSINIT	PSSINIT module
QoS	Quality of Service module
RDS	Radius module
RM	Routing Management module
RMON	Remote monitor module

Table 44 Module name list

Module name	Description
RPR	Resilient Packet Ring module
RSA	Revest, Shamir and Adleman module
RTPRO	Routing protocol module
SHELL	User interface module
SNMP	Simple Network Management Protocol module
SOCKET	Socket module
SSH	Secure Shell module
SYSM	System Manage veneer module
SYSMIB	System MIB module
TAC	Terminal Access Controller module
TELNET	Telnet module
UDPH	UDP Helper module
USERLOG	USER Calling Logging module
VFS	Virtual File System module
VLAN	Virtual Local Area Network module
VOS	Virtual Operation System module
VRRP	Virtual Router Redundancy Protocol module
VTY	Virtual Type Terminal module

To sum up, the major task of the information center is to output the three types of information of the modules onto the ten channels in terms of the eight severity levels and according to the user's settings, and then redirect the ten information channels to the seven output directions.

System Information Format

System information has the following format:

```
<priority>timestamp sysname module/level/digest:content
```

- *The closing set of angel brackets < >, the space, the forward slash /, and the colon are all required, and the percent sign % is optional in the above format.*
- *Before the <priority> may have %, "#, or * followed with a space, indicating log, alarm, or debug information respectively.*

Below is an example of the format of log information to be output to a log host:

```
% <188>Sep 28 15:33:46:235 2005 3Com SHELL/5/LOGIN: Console login from con0
```

What follows is a detailed explanation of the fields involved:

Priority

The priority is calculated using the following formula: facility*8+severity, in which facility is local7 by default and the range of severity is 0 to 7. Table 42 details the value and meaning associated with each severity.

Note that there is no space between the priority and timestamp fields and that the priority 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.

Note that there is a space between the timestamp and sysname (host name) fields.

The timestamp is in the format of Mmm dd hh:mm:ss yyyy, where

- Mmm" represents the month, and the available values are: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, and Dec.
- dd" is the date, which shall follow a space if less than 10, for example, " 7".
- hh:mm:ss" is the local time, where "hh" is in the 24-hour format, ranging from 00 to 23, and both "mm" and "ss" range from 00 to 59.
- yyyy" is the year.

Sysname

Sysname is the system name of the current host. You can use the **sysname** command to modify the sysname. (Refer to the *Switch 8800 Command Reference Guide* for details)

Note that there is a space between the sysname and module fields.

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.

Refer to Table 44 for module name and description.

Between "module" and "level" is a "/".

Level (Severity)

System information can be divided into eight levels based on its severity, from 0 to 7. Refer to Table 42 for definition and description of these severity levels. Note that there is a forward slash between the levels (severity) and digest fields.

Digest

The digest field is a string of up to 32 characters, outlining the system information.

Note that there is a colon between the digest and content fields.

Content

This field provides the content of the system information.

Configuring Information Center

Setting to Output System Information to the Console

Setting to output 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
Configure the channel through which system information can be output to the console	**info-center console channel** { *channel-number* | *channel-name* }	Optional System information is output to the console by default, with channel 0 as the default channel.
Configure the information source for an information channel.	**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* } *] *	Required Refer to Table 45 for the default output rules of system information.
Configure the format of the time stamp	**info-center timestamp** { **log** | **trap** | **debugging** } { **boot** | **date** | **none** }	Optional The time stamp for log, trap and debug information is **date** by default.

Table 45 Default output rules for different output directions

		LOG		TRAP		DEBUG	
Output direction	**Modules allowed**	**Enabled/ disabled**	**Severity**	**Enabled/ disabled**	**Severity**	**Enabled/ disabled**	**Severity**
Console	default (all modules)	Enabled	warnings	Enabled	debugging	Enabled	debugging
Monitoring terminal	default (all modules)	Enabled	warnings	Enabled	debugging	Enabled	debugging
Log host	default (all modules)	Enabled	informational	Enabled	debugging	Disabled	debugging
Trap buffer	default (all modules)	Disabled	informational	Enabled	warnings	Disabled	debugging
Log buffer	default (all modules)	Enabled	warnings	Disabled	debugging	Disabled	debugging
SNMP NMS	default (all modules)	Disabled	debugging	Enabled	warnings	Disabled	debugging

Table 45 Default output rules for different output directions

		LOG		TRAP		DEBUG	
Output direction	**Modules allowed**	**Enabled/ disabled**	**Severity**	**Enabled/ disabled**	**Severity**	**Enabled/ disabled**	**Severity**
Log file	default (all modules)	Enabled	debuggin g	Enabled	debuggin g	Disabled	debuggin g

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 monitoring terminal by default.
Enable the display of debug 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

Setting to Output 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, or TTY user interface.

Setting to output 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 43 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 System information is output to the monitor terminal by default with channel 1 as the default channel.

To do...	Use the command...	Remarks
Configure the information source for an information channel	**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* } *] *	Required
Configure the format of the time stamp	**info-center timestamp** { **log** | **trap** | **debugging** } { **boot** | **date** | **none** }	Optional By default, the time stamp for log, trap and debug 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 disabled on the monitoring terminal by default.
Enable the display of debug 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

Setting to Output 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 43 for default channel names.
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* | **language** { **chinese** | **english** }] *	Required Disabled by default with channel 2 as the default channel when enabled.

To do...	Use the command...	Remarks
Configure the source interface through which log information can be output to a log host	**info-center loghost source** *interface-type interface-number*	Optional No source interface is configured by default, and the system selects an interface as the source interface.
Configure the information source for an information channel	**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* } *] *	Required
Configure the format of the time stamp for log information	**info-center timestamp loghost** { **date** | **no-year-date** | **none** }	Optional **date** by default.

Setting to Output System Information 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 43 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 System information is output to the trap buffer by default with channel 3 (known as trapbuffer) as the default channel and a default buffer size of 256.
Configure the information source for an information channel	**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* } *] *	Required
Configure the format of the time stamp	**info-center timestamp** { **log** | **trap** | **debugging** } { **boot** | **date** | **none** }	Optional The time stamp for log, trap and debug information is **date** by default.

Setting to Output System Information to the Log Buffer

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable information center	**info-center enable**	Optional Enabled by default.

To do...	Use the command...	Remarks
Name the channel with a specified channel number	**info-center channel** *channel-number* **name** *channel-name*	Optional Refer to Table 43 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 System information is output to the log buffer by default with channel 4 (known as logbuffer) as the default channel and a default buffer size of 512.
Configure the information source for an information channel	**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* } *] *	Required
Configure the format of the timestamp	**info-center timestamp** { **log** | **trap** | **debugging** } { **boot** | **date** | **none** }	Optional The time stamp for log, trap and debug information is **date** by default.

Setting to Output System Information to the SNMP NMS

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 43 for default channel names.
Configure the channel through which system information can be output to the SNMP NMS	**info-center snmp channel** { *channel-number* | *channel-name* }	Optional System information is output to the SNMP NMS by default with channel 5 (known as snmpagent) as the default channel.
Configure the information source for an information channel	**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* } *] *	Required
Configure the format of the timestamp	**info-center timestamp** { **log** | **trap** | **debugging** } { **boot** | **date** | **none** }	Optional The time stamp for log, trap and debug information is **date** by default.

To ensure that system information can be output to the SNMP NMS, you need to make the necessary configurations on the SNMP agent and the NMS. For detailed information on SNMP, refer to "SNMP Configuration" on page 1043.

Setting to Save System Information to a Log File

With the log file feature enabled, the log information generated by system can be saved to a specified directory with a predefined frequency. This allows you to check the operation history at any time to ensure that the device functions properly.

Follow these steps to set to save system information to a log file:

To do...	Use the command...	Remarks
Enter system view	**system-view**	-
Enable the log file feature	**info-center logfile enable**	Optional Enabled by default
Configure the frequency with which the log file is saved	**info-center logfile frequency** *freq-sec*	Optional The default value varies with devices.
Configure the language mode of the log file	**info-center logfile language** { **chinese** | **english** }	English by default
Configure the maximum storage space reserved for a log file	**info-center logfile size-quota** *size*	Optional
Configure the directory to save the log file	**info-center logfile switch-directory** *dir-name*	Optional By default, the logfile directory under the root directory of the memory device, which varies with devices.
Manually save the log buffer content to the log file	**logfile save**	Optional Available in any view By default, the system saves the log file with the frequency defined by the **info-center logfile frequency** command.

- *To ensure the device to work normally, use the* ***info-center logfile size-quota*** *command to set a logfile to be no smaller than 1 MB and no larger than 10 MB.*
- *Use the* ***info-center logfile switch-directory*** *command to manually configure the directory to which a log file can be saved. The configuration will be invalid after system reboot or the primary/backup switchover.*

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 debug information, then after the completion of system output the system will display a command line prompt (in command editing mode a prompt, 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

- *If you do not input any information following the current command line prompt, the system does not display any command line prompt after system information output.*
- *In the interaction mode, you are prompted for some information input. If the input is interrupted by system output, no system prompt will be made, rather only your input will be displayed in a new line.*

Displaying and Maintaining Information Center

To do...	Use the command...	Remarks
Display channel information for a specified channel	**display channel** [*channel-number* | *channel-name*]	Available in any view
Display the configurations for all information channels except channel 6 to 8	**display info-center**	Available in any view
Display the state of the log buffer and the log information recorded	**display logbuffer** [**level** *severity* | **size** *buffersize* | **slot** *slot number*] * [**|** { **begin** | **exclude** | **include** } *text*]	Available in any view
Display a summary of the log buffer	**display logbuffer summary** [**level** *severity* | **slot** *slotnum*] *	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** [**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

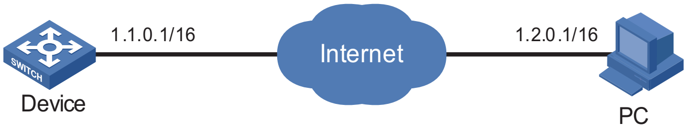
Outputting Log Information to a Unix Log Host

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.

Network diagram

Figure 331 Network diagram for outputting log information to a Unix log host



Configuration procedure

1 Configuring the device

# Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

# Specify the channel to output log information to the log host (loghost by default, optional).

```
[Sysname] info-center loghost 1.2.0.1 channel loghost
```

# Disable the output of log, trap, and debug information of all modules to the log host.

```
[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 vary, ensure that the output of log, trap, and debug information for the specified channel (loghost in this example) of all modules is disabled before the system information can be output to meet the current network requirements.*
- *Use the **display channel** command to display the state of a channel.*

# Set the host with an IP address of 1.2.0.1/16 to be the log host, set the severity to informational, language to English, and the source modules to ARP and IP. (Note that the source modules vary with devices.).

```
[Sysname] info-center loghost 1.2.0.1 facility local4 language english
[Sysname] info-center source arp channel loghost log level informational
[Sysname] info-center source ip channel loghost log level informational
```

2 Configuring 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: issue the following commands as a root user.

```
# mkdir /var/log/3Com
# touch /var/log/3Com/information
```

Step 2: Edit the file /etc/syslog.conf as a root user and add the following selector/action pair.

```
# 3Com configuration messages
local4.info     /var/log/3Com/information
```

Be aware of the following issues while editing the /etc/syslog.conf file:

- Comments must be on a separate line and must begin with the # sign.
- The selector/action pair must be separated with a tab key, rather than a space.
- No redundant spaces are allowed in the file name.
- The device name and the accepted severity of log information specified by the /etc/syslog.conf file must be identical to those configured on the device using the **info-center loghost** or **info-center source** command; otherwise the log information may not be output properly to the log host.

Step three: after the log file information has been created and the configuration file /etc/syslog.conf has been modified, ensure that the configuration file /etc/syslog.conf is reread:

```
# ps -ae | grep syslogd
147
# kill -HUP 147
```

After the above configurations, the system will be able to keep log information in the related file.

ole and displays it on the console.

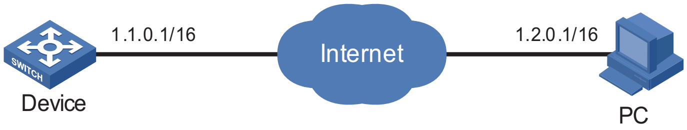
Outputting Log Information to a Linux Log Host

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.

Network diagram

Figure 332 Network diagram for outputting log information to a Linux log host



Configuration procedure

1 Configuring the device

# Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

# Specify the channel to output log information to the log host (optional, loghost by default).

```
[Sysname] info-center loghost 1.2.0.1 channel loghost
```

# Disable the output of log, trap, and debug information of all modules to the log host.

```
[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 vary, ensure that the output of log, trap, and debug information for the specified channel (loghost in this example) of all modules is disabled before the system information can be output to meet the current network requirements.*
- *Use the **display channel** command to display the state of a channel.*

# Set the host with an IP address of 1.2.0.1/16 to be the log host, set the severity to informational, language to English, and the source modules to be all modules.

```
[Sysname] info-center loghost 1.2.0.1 facility local7 language english
[Sysname] info-center source default channel loghost log level informational
```

2 Configuring the log host

Step 1: issue the following commands as a root user.

```
# mkdir /var/log/3Com
# touch /var/log/3Com/information
```

Step 2: Edit the file /etc/syslog.conf as a root user and add the following selector/action pair.

```
# 3Com configuration messages
local7.info    /var/log/3Com/information
```

Be aware of the following issues while editing the /etc/syslog.conf file:

- *Comments must be on a separate line and must begin with the # sign.*
- *The selector/action pair must be separated with a tab key, rather than a space.*
- *No redundant spaces are allowed in the file name.*
- *The device name and the accepted severity of the log information specified by the /etc/syslog.conf file must be identical to those configured on the device using the **info-center loghost** or **info-center source** command; otherwise the log information may not be output properly to the log host.*

Step three: after the log file information has been created and the /etc/syslog.conf file has been modified, issue the following commands to display the process ID of syslogd, terminate a syslogd process, and to restart syslogd using the -r option.

```
# ps -ae | grep syslogd
147
# kill -9 147
# syslogd -r &
```

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 keep log information in the related file.

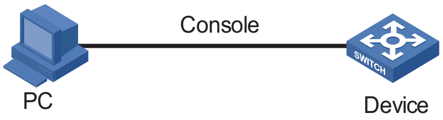
Outputting Log Information to the Console

Network requirements

- Log information with a severity higher than informational will be output to the console;
- The source modules are ARP and IP.

Network diagram

Figure 333 Network diagram for sending log information to the console



Configuration procedure

# Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

# Specify the channel to output log information to the console (optional, Console by default).

```
[Sysname] info-center console channel console
```

# Disable the output of log, trap, and debug information of all modules to the log host.

```
[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 vary, ensure that the output of log, trap, and debug information for the specified channel (console in this example) of all modules is disabled before the system information can be output to meet the current network requirements.*
- *Use the **display channel** command to display the state of a channel.*

# Enable system information output for the ARP and IP modules, with information severity ranging from emergencies to informational.

```
[Sysname] info-center source ARP channel console log level informational
[Sysname] info-center source ip channel console log level informational
[Sysname] quit
```

# Enable the display of log information on a monitor terminal.

```
<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 and displays it on the console.