HP Virtual Connect and HP 5820 Switch Series HP IRF integration guide

Technical white paper

Table of contents

Introduction

HP Intelligent Resilient Framework (IRF) is an innovative technology for switch platform virtualization, which allows dramatic simplification of the design and operations of data center and campus Ethernet networks. IRF overcomes the limitations of traditional STP-based and legacy competitive designs by delivering new levels of network performance and resiliency.

HP Virtual Connect (VC) is an industry standard-based implementation of server-edge virtualization. It cleanly separates server enclosure administration from LAN and SAN administration and allows you to add, move, or replace servers without impacting production LAN and SAN availability.

This white paper provides detailed configuration and testing information of common usage scenarios using HP 5820 and 5800 Switch Series utilizing IRF to form a logical switch, which connects to the server HP BladeSystem-integrated VC modules.

The intended audience for this white paper is HP solution architects and HP technical consultants.

HP IRF and VC setup configurations

- HP 5820/5800 Switch Series and IRF link setup from two standalone switches (on page 10)
- 5820/5800 Switch Series BFD MAD link setup (on page 13)
- LLDP neighbor discovering (on page 14)
- LACP port bundling (long timeout and short timeout) (on page 16)

Failover tests

- 5820 port-channel (bridge-aggregation interface connecting to VC) failure (on page 22)
- 5820 switch failure (on page 24)
- 5820 IRF link failure to test MAD detection (on page 25)
- VC primary module failure (on page 28)

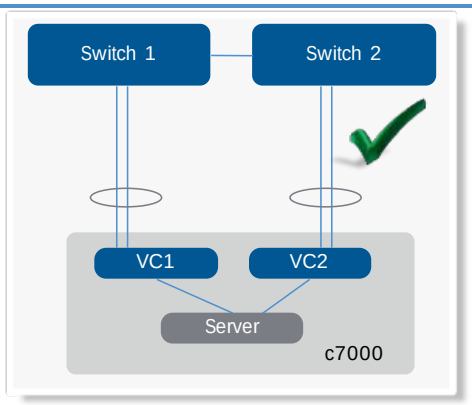
Images of HP Intelligent Management Center and HP Insight Control for VMware vCenter network monitoring

- HP Insight Control for VMware vCenter plug-in screen capture of network monitoring of VC, vSwitch, and access switch (5820) (on page 29)
- HP Networking HP Intelligent Management Center (IMC) screen capture of 5820 and VC monitoring (on page 32)

Design scenarios

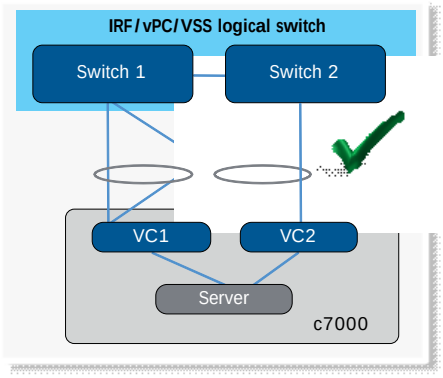
Two typical design scenarios are available to connect VC with network switches.

Figure 1. Design scenario 1



Scenario 1—This is a typical connection scenario, in which VC modules connect with non-IRF/Microsoft® Windows® Virtual PC/Microsoft Visual Source Safe capable switches. VC needs to configure one SUS per VC module (two total). Switch 1 and switch 2, each has one port channel configured to peer with VC SUS.

Figure 2. Design scenario 2

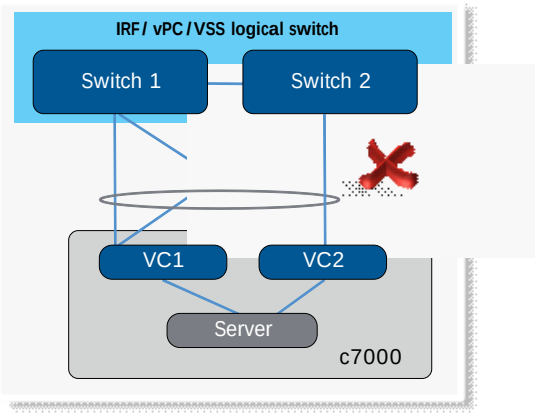


Scenario 2—This is the recommended connection scenario, in which VC modules connect with IRF/Microsoft Windows Virtual PC/Microsoft Visual Source Safe logical switch. VC needs to configure one SUS per VC module (two total). The logical switch also has two port channels configured to peer with VC SUS, which is known as active/active VC design. Active/standby VC design is also available, but because it does not use all available uplink bandwidth, it is not discussed. For more information on active/standby, see Scenario 1—4 in the "HP Virtual Connect Fibre Channel Network Scenarios" cookbook

This design provides two main benefits over the previous design:

- If either switch fails, traffic remains on the same port channel and rehashes to the remaining physical link in less than one second. The server does not require failover tests.
- For the incoming traffic from upstream core switch to server direction, all traffic can be sent to VC. Previously, if the destination MAC was on the other switch, the traffic would have to traverse the inter-switch trunk, so the flow was not optimized.

Figure 3. Design scenario 3

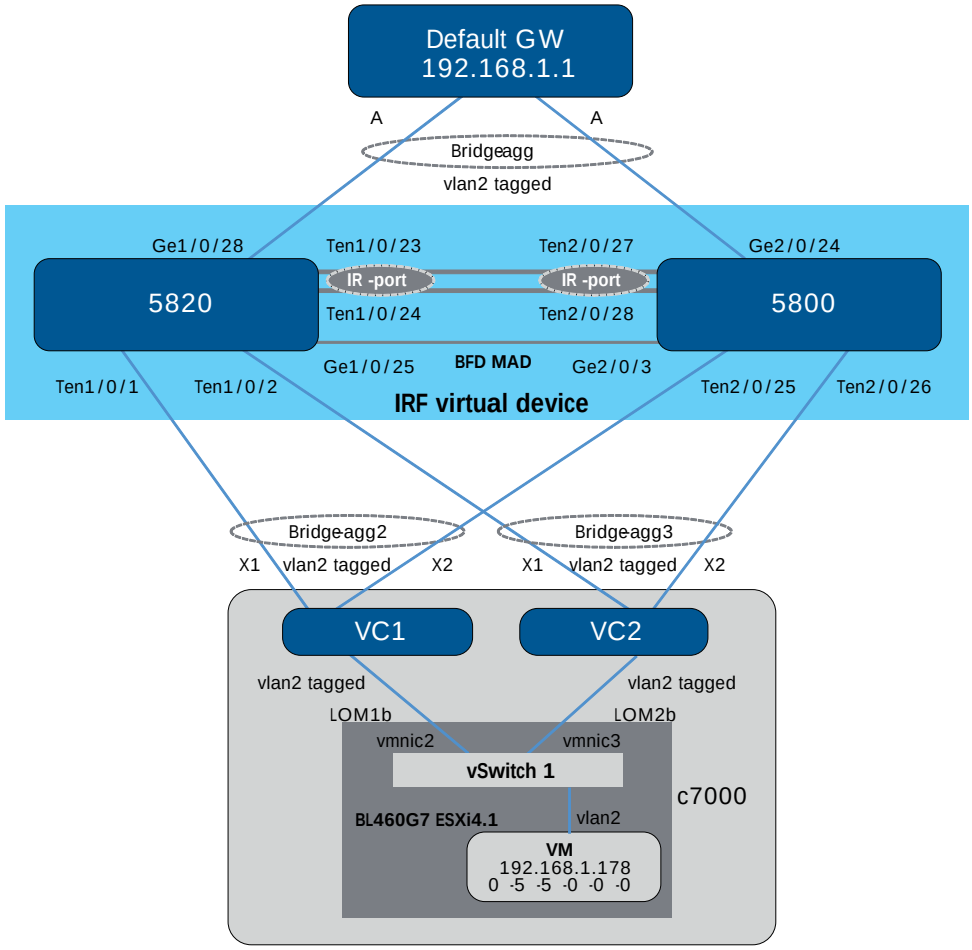


Scenario 3—This configuration does not work. Configuring one port channel on a logical switch side and one SUS on a VC side does not move traffic forward on all four links. VC does not support port channels across different modules. Some links will go into standby and not form port channels.

See Appendix 2 (on page 40) for the results of this scenario.

Network topology

Figure 4. Physical diagram



The IRF cluster consists of one 5820 switch and one 5800-32C switch. HP Comware-based switches support IRF clustering on different switch models if they are compatible with each other for IRF.

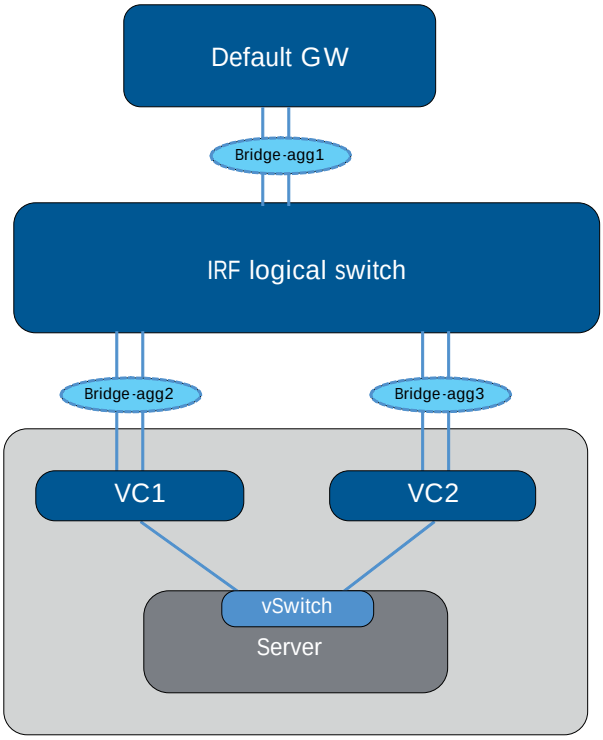
The 5820 and 5800 Switch Series forms an IRF bundle link between them with two 10 GB links. The 5820 switch is switch 1, the master of the domain, and has logical port IRF-port 2. The 5800 switch is switch 2, the slave of the domain, and has logical port IRF-port 1, defined originally before merging with the 5820 switch.

The 5820 and 5800 Switch Series uses one Gigabit Ethernet link as a BFD MAD link for MAD.

VC1 and VC2 are HP Flex-10 modules in interconnect bays 1 and 2 of the HP BladeSystem c7000 Enclosure. Each Flex-10 module has a SUS connecting to an IRF virtual device. A SUS consists of two 10 GB links terminated on 5820 and 5800 Switch Series. With IRF, these two 10 GB links form one bridge-aggregation bundle (the same as port channel on Cisco NX-OS and EtherChannel on Cisco IOS). VC1 connects the IRF cluster with the bridge-aggregation 2 interface, and VC2 connects the IRF cluster with the bridge-aggregation 3 interface. Bridge-aggregation 1 forms a virtual port channel between the IRF cluster and the virtual machine (VM) default gateway (simulated by an HP Networking switch).

Traffic flow testing uses ping packets from VM1 (192.168.1.178) to its default gateway (192.168.1.1). The default gateway is simulated by an HP Networking switch. The VM traffic has two paths to reach its default gateway, depending on how the vSwitch hashes VM traffic to a specific VM NIC. VM NIC2 is mapped to Flex-10 in the following illustration.

Figure 5. Logical diagram



Two bundle interfaces (bridge-aggregation 2 and bridge-aggregation 3) exist between the VC and the IRF logical switch because VC currently does not support link bundling across two different physical modules.

IRF

IRF

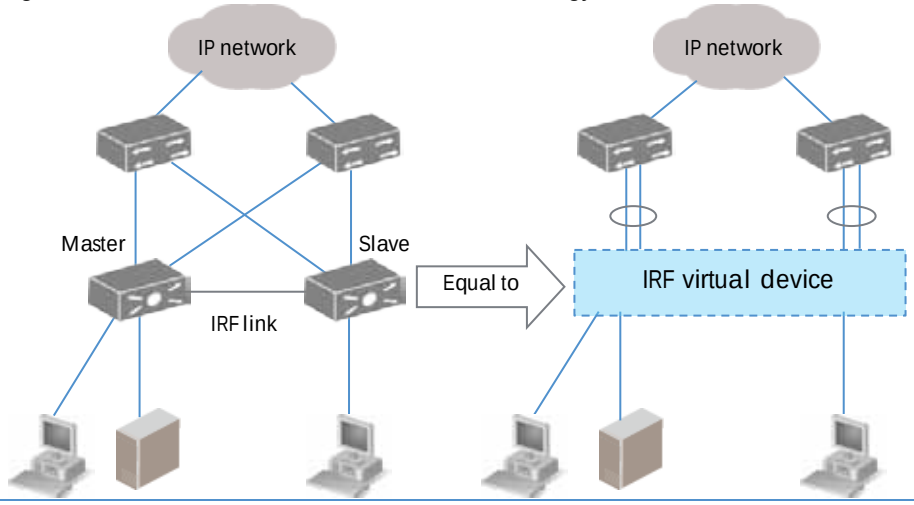
IRF creates one logical switch from two or more physical switches. The HP 5820 switch can support up to nine switches in one IRF domain.

The logical switch uses standard LACP to connect to any vendor, core, distribution, or edge switches with a failure convergence time of less than 40 milliseconds. The switch acts as the following:

- Single IP address for management
- Single Layer 2 switch
- Single Layer 3 router (all protocols)

Implementation is available across multiple products from core to access platforms HP 12500, 10500, 9500, 7500, 5900AF, 5920AF, 5830AF, 5820, 5800, 5500-HI, and 5500-EI Switch Series.

Figure 6. Network transformation with HP IRF technology



Role

Member switches form an IRF virtual device. Each of them performs one of the following two roles:

- Master—manages the IRF virtual device
- Subordinate—members that are backups of the master

If the master fails, the IRF virtual device automatically elects a new master from one of the subordinates. Masters and subordinates are elected through the role election mechanism. An IRF virtual device has only one master at a time.

IRF port

An IRF port is a logical port dedicated to the internal connection of an IRF virtual device. An IRF port can be numbered as IRF-port 1 or IRF-port 2. An IRF port is effective only after it is bound to a physical port.

Important:

An IRF-port 1 on one device can only be connected to the physical port bound to the IRF-port 2 of a neighboring device; otherwise, an IRF virtual device cannot be formed.

Figure 7. IRF virtual device

Physical IRF port

Physical IRF ports are physical (copper or fiber) ports bound to an IRF port. They perform the following functions:

- Connect IRF member switches
- Forward IRF protocol packets and data packets between IRF member switches

Priority	Member priority determines the role of a member during a role election process. A member with a higher priority is more likely to be a master. The priority of a switch defaults to 1.
Member ID	An IRF virtual device uses member IDs to identify its members uniquely. Configuration information such as port (physical or logical) numbers, port configurations, and member priorities relate to member IDs.
Domain ID	Each switch belongs to one IRF domain. By default, the domain ID is 0. Although switches with different domain IDs can form an IRF virtual device, HP recommends assigning the same domain ID to the members of the same IRF virtual device. Otherwise, the LACP MAD detection cannot function properly.

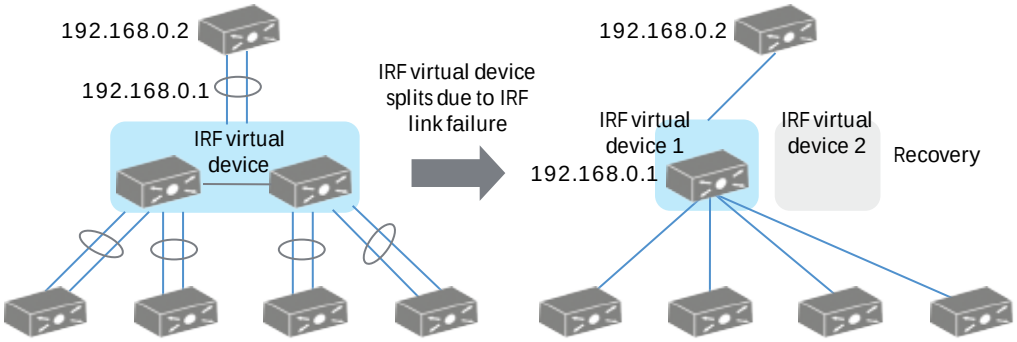
MAD

MAD

MAD protects IRF link failure when both switches with the same configuration meet the criteria for master switch. In this case, MAD shuts down one of the switches according to role election. The switch with a higher priority becomes the master, and then the local interfaces for switch 2 are shut down.

When an IRF link is down as a result of MAD, switch 1 continues to run. Switch 2 inactivates all local interfaces.

Figure 8. HP IRF split due to link failure



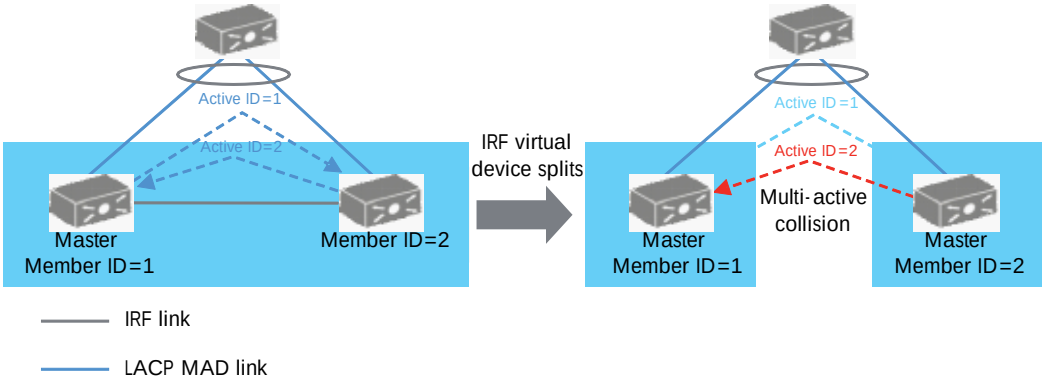
MAD detects multiple active IRFs using one of three methods:

- LACP
- BFD
- ARP

LACP MAD

- Most widely deployed
- Fastest convergence time
- Needs only one CLI MAD enabled under bridge-aggregation interface
- Needs a third switch (typically HP Networking) to understand extended LACPDU packets

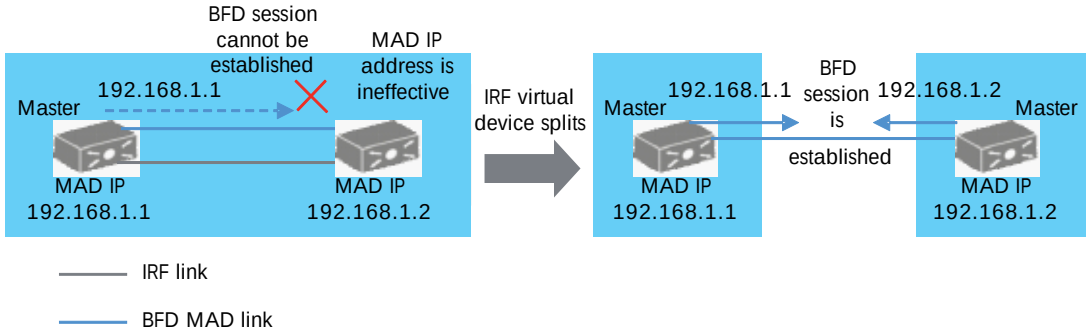
Figure 9. HP IRF split with LACP MAD protection



BFD MAD

- Provides fast convergence time
- Needs a separate link between two switches to act as a BFD MAD link
- Does not require switches outside the IRF domain

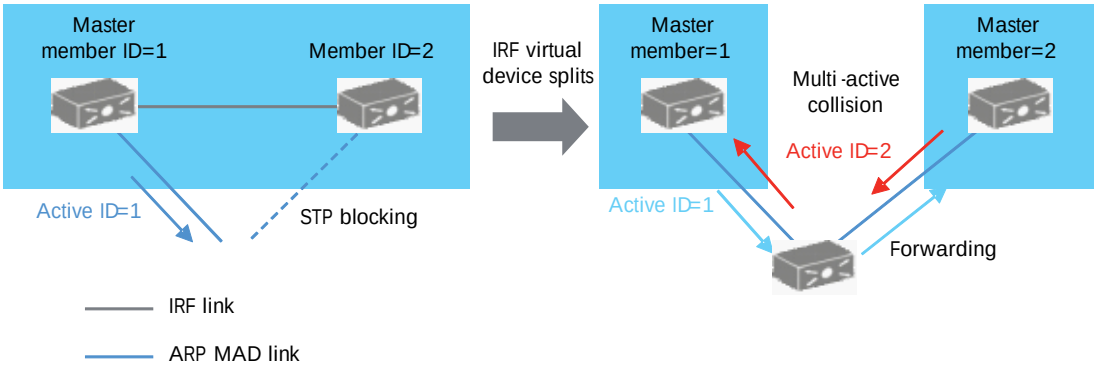
Figure 10. HP IRF split with BFD MAD protection



ARP MAD

This MAD is not widely deployed.

Figure 11. HP IRF split with ARP MAD protection .



For more information on IRF and MAD, see the HP 5820 & 5800 Series Ethernet Switches IRF Configuration Guide.

HP IRF and VC setup configurations

Quick CLI reference table

HP Networking Comware CLI is similar to the Cisco IOS/NX-OS format. The following table gives a quick comparison of Comware CLI and Cisco CLI, related to this setup.

HP Comware	Cisco
system	config terminal
undo	no
quit	exit
save force	wr mem
reset saved-config	wr erase
reboot	reload
display current	show run
display saved-configuration	show startup
display int brief	show ip int brief
display logbuffer	show log
display link-aggregation	show etherchannel/port-channel
display this (show current interface config)	
sysname	hostname
port link-mode bridge	switchport
port link-mode route	no switchport
port link-type access	switchport mode access
port link-type trunk	switchport mode trunk
port access vlan x	switchport access vlan x
port trunk permit vlan x	switchport trunk allowed vlan x
port link-aggregation group x	channel-group x
interface Bridge-Aggregation x	int port-channel x

HP 5820 Switch Series: convert standalone switches to IRF logical switch

This conversion procedure assumes that two standalone switches start from a clean factory-default startup configuration. If not, enter **reset saved-config (write erase** on Cisco) to reset startup configuration to factory default.

5820 (switch 1)

1. Change switch 1 IRF priority to 10. The default value is 1, and the higher priority is selected to be the IRF master and active switch when MAD is detected.

irf member 1 priority 10

2. Shut down the IRF physical ports to prepare them to be included under the IRF logical port "irf-port 1/2" configuration. Otherwise, when trying to include these interfaces later under IRF-port, Comware will indicate that the physical interfaces are not shut down.

int ten1/0/23

[Ten-GigabitEthernet1/0/23]**shut**

[Ten-GigabitEthernet1/0/23]**int ten1/0/24**

[Ten-GigabitEthernet1/0/24]**shut**

3. Create logical port "irf-port 1/2" and include ten1/0/23 and ten1/0/24.

Note: If you create "irf-port 1/2" on switch 1, you must use "irf-port 2/1" on switch 2. Alternatively, create local "irf-port 1/1" and use "irf-port 2/2" on switch 2. The following two scenarios do not work:

"irf-port 1/1"-- "irf-port 2/1"

"irf-port 1/2"--"irf-port 2/2"

irf-port 1/2

[irf-port1/2]**port group interface ten1/0/23**

[irf-port1/2]**port group interface ten1/0/24**

4. While ten1/0/23 and ten1/0/24 are shut down, go to switch 2 (on page 11) to configure it to peer with switch 1. Then, complete the remaining steps in this procedure.

5. Undo the shutdown of ten1/0/23 and ten1/0/24 to bring up the irf-link. After the links and interfaces appear, proceed to the next step. Nothing happens until step 6 is executed.

int ten1/0/23

[Ten-GigabitEthernet1/0/23]**undo shut**

[Ten-GigabitEthernet1/0/23]**int ten1/0/24**

[Ten-GigabitEthernet1/0/24]**undo shut**

6. Activate the irf-port configuration to start IRF peering between the two switches.

irf-port-configuration active

After several seconds, switch 2 reloads. When switch 2 comes back on, two switches are merged into one virtual IRF switch. You can use the three IRF commands to verify the running status for this virtual IRF switch. See the output following 5800 (switch 2).

5800 (switch 2)

1. Change switch 2 member ID from default 1 to 2.

irf member 1 renumber 2

2. Before continuing with the following steps, reboot the switch to make all interface numbering changes from 1/x/y to 2/x/y. This command is executed when the switch is not in system mode.

reboot

After rebooting

3. Shut down the IRF physical ports to prepare them to be included under the IRF logical port "irf-port 2/1" configuration. Otherwise, when trying to include these interfaces later under IRF-port, Comware will indicate that the physical interfaces are not shut down.

int ten2/0/27

[Ten-GigabitEthernet2/0/27]**shut**

[Ten-GigabitEthernet2/0/27]**int ten2/0/28**

[Ten-GigabitEthernet2/0/28]**shut**

4. Create logical port "irf-port 2/1" and include ten2/0/27 and ten2/0/28.

irf-port 2/1

[irf-port2/1]**port group interface ten2/0/27**

[irf-port2/1]**port group interface ten2/0/28**

5. Undo the shutdown of ten2/0/27 and ten2/0/28 to bring up the irf-link. After the links and interfaces appear, proceed to the next step. Nothing happens until step 6 is executed.

int ten2/0/27

[Ten-GigabitEthernet2/0/27]**undo shut**

[Ten-GigabitEthernet2/0/27]**int ten2/0/28**

[Ten-GigabitEthernet2/0/28]**undo shut**

6. Activate irf port configuration to start IRF peering between two switches. At this moment, nothing happens because both switch 1 IRF physical links are still shut down.

irf-port-configuration active

7. Go to switch 1 (on page 10) to start IRF physical links and activate the IRF-link configuration. Several seconds later, switch 2 reloads itself with the message below (only part of the booting message is shown here for reference).

IRF port 1 is up.

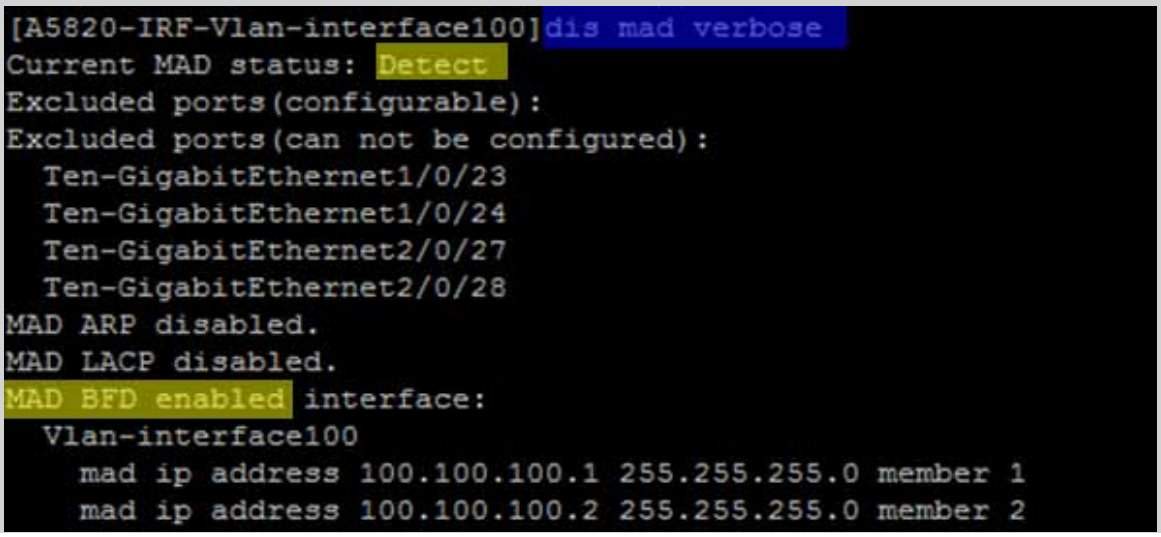
Starting.....

```
**************************************************************************
*                                                                        *
*       5800-32C BOOTROM, Version 205                                    *
*                                                                        *
**************************************************************************
```

After merging, IRF status checks the output. For the complete logical switch configuration, see Appendix 2 (on page 40).

```
[A5820-IRF]dis irf
Switch   Role    Priority  CPU-Mac         Description
*+1      Master  10        0023-8943-7525  -----
  2      Slave   1         0023-893c-45d6  -----
-------------------------------------------------

* indicates the device is the master.
+ indicates the device through which the user logs in.

The Bridge MAC of the IRF is: 0023-8943-7524
Auto upgrade                : yes
Mac persistent              : 6 min
Domain ID                   : 0
[A5820-IRF]
[A5820-IRF]
[A5820-IRF]dis irf top
                          Topology Info
------------------------------------------------------------------------
               IRF-Port1                 IRF-Port2
Switch    Link       neighbor      Link       neighbor    Belong To
1         DIS        --            UP         2           0023-8943-7525
2         UP         1             DIS        --          0023-8943-7525

[A5820-IRF]
[A5820-IRF]
[A5820-IRF]dis irf config
MemberID NewID    IRF-Port1                IRF-Port2
 1       1        disable                  Ten-GigabitEthernet1/0/23
                                           Ten-GigabitEthernet1/0/24
 2       2        Ten-GigabitEthernet2/0/27 disable
                  Ten-GigabitEthernet2/0/28
```

```
#
vlan 100
#
interface vlan-interface100
 mad bfd enable
 mad ip address 100.100.100.1 255.255.255.0 member 1
 mad ip address 100.100.100.2 255.255.255.0 member 2
#
interface GigabitEthernet1/0/25
 port link-mode bridge
 port access vlan 100
 stp disable
#
interface GigabitEthernet2/0/3
 port link-mode bridge
 port access vlan 100
 stp disable
```



To STP for the BFD MAD interface, issue the command **`stp disable`**. The BFD MAD interface is a dedicated interface and should not run any other services or features.

HP 5820 Switch Series: LLDP

LLDP is the equivalent of Cisco Discovery Protocol.
LLDP transmits and receives are enabled by default on 5820 interfaces. No configuration is required.

```
[A5820-IRF]dis lldp neigh list

System Name          Local Interface  Chassis ID      Port ID
A5820-IRF            GE1/0/25         0023-8943-7524  GigabitEthernet2/0/3
Procurve for GW      GE1/0/28         000a-5774-5f00  1
Procurve for GW      GE2/0/24         000a-5774-5f00  2
VcD_dbfd80bb5d2c     XGE1/0/1         d485-64ce-f015  X1
VcD_dbfd80bb5d2c     XGE1/0/2         d485-64ce-f033  X1
VcD_dbfd80bb5d2c     XGE2/0/25        d485-64ce-f015  X2
VcD_dbfd80bb5d2c     XGE2/0/26        d485-64ce-f033  X2
```

The “VcD_xyz” string is the unique VC domain ID generated internally when creating VC. VC1 and VC2 share the same LLDP “System Name” because they are in the same VC domain. To determine which physical VC module is the LLDP neighbor, use the “Chassid ID” field. This is the VC module system MAC address. To determine the system MAC address for a particular VC module, log into VC by SSH and use the **show interconnect** command.

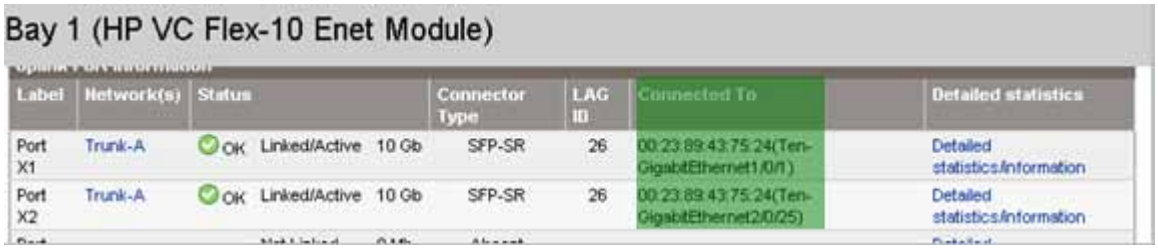
```
->show interconnect enc0:1
ID                  : enc0:1
Enclosure           : oa8
Bay                 : 1
Type                : VC-ENET
Product Name        : HP VC Flex-10 Enet Module
Role                : Primary
Status              : OK
Comm Status         : OK
OA Status           : OK
Power State         : On
MAC Address         : d4:85:64:ce:f0:15
Node WWN            : -- --
Firmware Version    : 3.15 2010-10-09T07:18:16Z
Manufacturer        : HP
Part Number         : 455880-B21
Spare Part Number   : 456095-001
Rack Name           : R8-9-10
Serial Number       : 3C4031000B
UID                 : Off
```

HP Flex-10: LLDP

LLDP transmits and receives are enabled by default on all VC modules interfaces, including HP Flex-10 and FlexFabric. No configuration is required.

Trunk-A and Trunk-B are defined in the following LACP sections. All links will show as active only after finishing the LACP configuration on the switch and VC.

VC1 connects with IRF logical switch ten1/0/1 and ten2/0/25.



Bay 1 (HP VC Flex-10 Enet Module)

Label	Network(s)	Status	Connector Type	LAG ID	Connected To	Detailed statistics
Port X1	Trunk-A	OK Linked/Active 10 Gb	SFP-SR	26	00:23:89:43:75:24(Ten-GigabitEthernet1/0/1)	Detailed statistics/information
Port X2	Trunk-A	OK Linked/Active 10 Gb	SFP-SR	26	00:23:89:43:75:24(Ten-GigabitEthernet2/0/25)	Detailed statistics/information

VC2 connects with IRF logical switch ten1/0/2 and ten2/0/26.

Bay 2 (HP VC Flex-10 Enet Module)

Label	Network(s)	Status	Connector Type	LAG ID	Connected To	Detailed statistics
Port X1	Trunk-B	OK Linked/Active 10 Gb	SFP-SR	26	00:23:89:43:75:24(Ten-GigabitEthernet1/0/2)	Detailed statistics/information
Port X2	Trunk-B	OK Linked/Active 10 Gb	SFP-SR	26	00:23:89:43:75:24(Ten-GigabitEthernet2/0/26)	Detailed statistics/information

HP 5820 Switch Series: LACP

The bridge-aggregation interface is equal to the port channel interface on Cisco to bundle multiple physical links.

```
interface Bridge-Aggregation2
 port link-type trunk
 port trunk permit vlan 1 to 2
 link-aggregation mode dynamic
 stp edged-port enable
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 to 2
 link-aggregation mode dynamic
 stp edged-port enable
#
interface Ten-GigabitEthernet1/0/1
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 2
#
interface Ten-GigabitEthernet1/0/2
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 3
#
interface Ten-GigabitEthernet2/0/25
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 2
#
interface Ten-GigabitEthernet2/0/26
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 3
#
```

When connecting with VC, the Spanning Tree edge ports (Cisco PortFast) feature should be enabled because VC does not communicate STP with any network device. The command is **stp edged-port enable** under the interface. This can speed up network convergence time, especially when links come up.

The BPDU guard feature can be enabled for more security to protect edge ports. The global command is **stp bpdu-protection.**

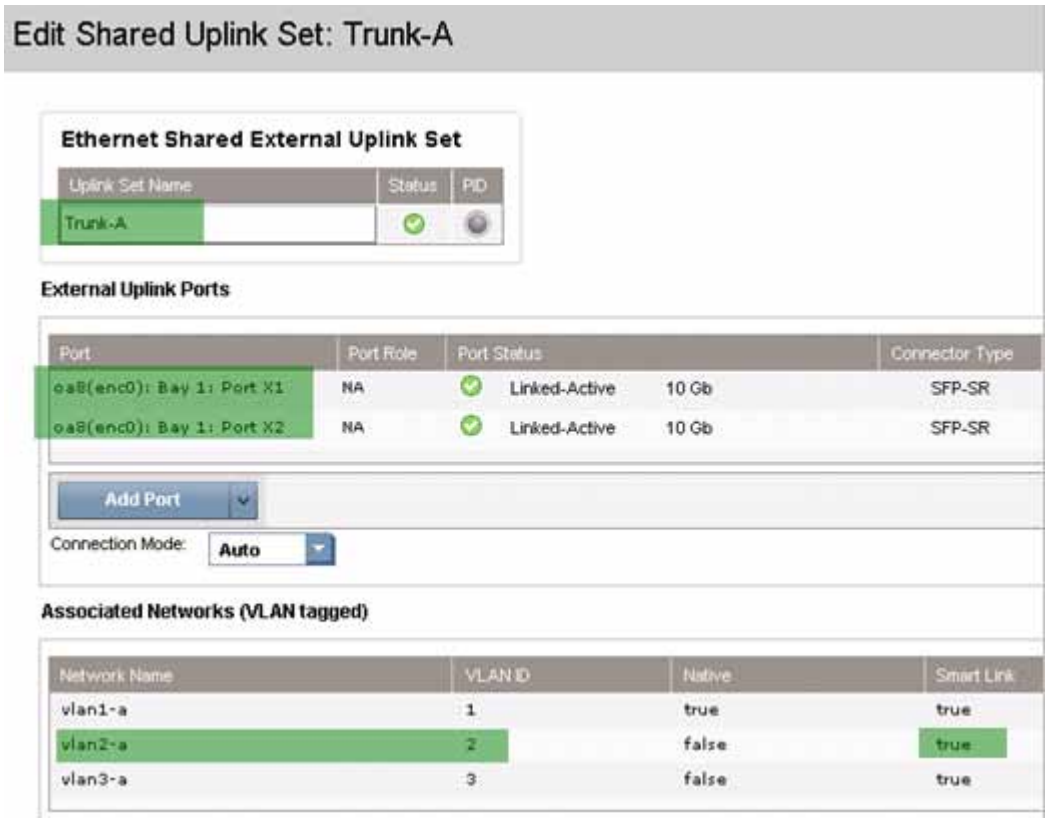
These practices are in line with networking best design when connecting with host NICs. Networking switches should treat any ports connecting with VC as the ports connecting with regular servers.

Bridge-aggregation interfaces commands

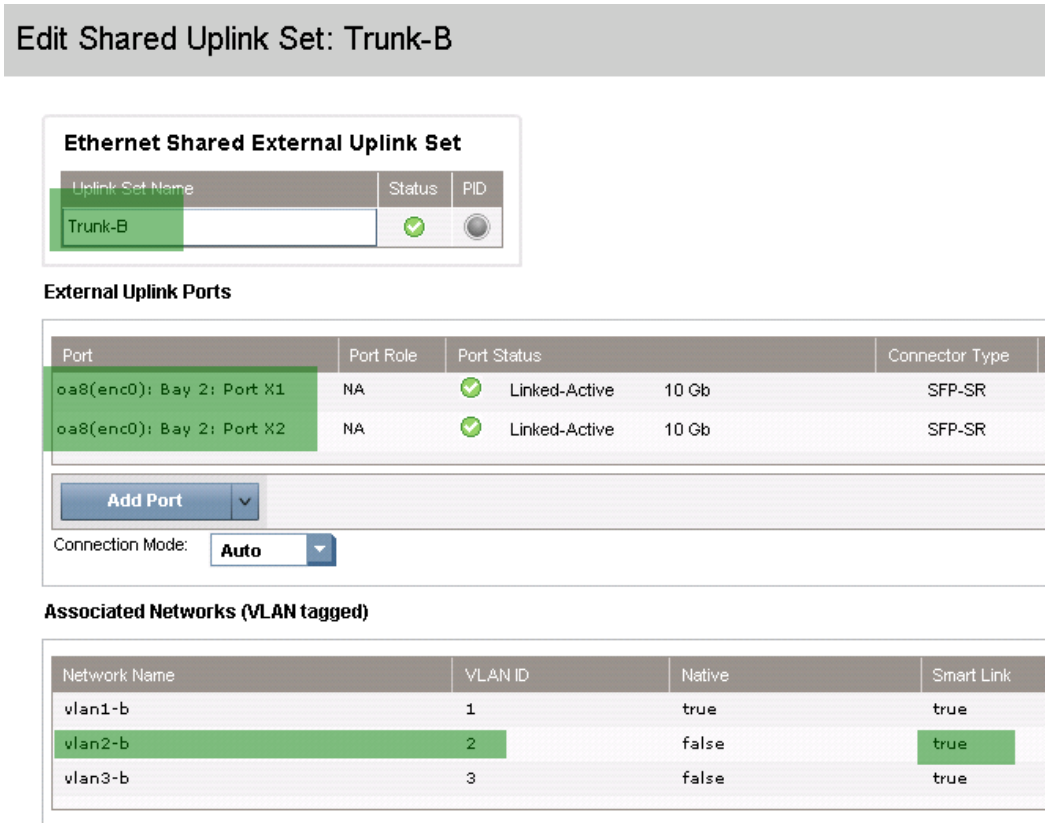
```
[A5820-IRF]dis link-aggregation verbose b2

Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregation Interface: Bridge-Aggregation2
Aggregation Mode: Dynamic
Loadsharing Type: Shar
System ID: 0x8000, 0023-8943-7524
Local:
  Port            Status  Priority Oper-Key  Flag
--------------------------------------------------------------------------------
  XGE1/0/1        S       32768    1         {ACDEF}
  XGE2/0/25       S       32768    1         {ACDEF}
Remote:
  Actor           Partner Priority Oper-Key  SystemID               Flag
--------------------------------------------------------------------------------
  XGE1/0/1        17      1        3         0x1   , d485-64ce-f015 {ABCDEF}
  XGE2/0/25       18      1        3         0x1   , d485-64ce-f015 {ABCDEF}
[A5820-IRF]dis link-aggregation verbose b3

Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregation Interface: Bridge-Aggregation3
Aggregation Mode: Dynamic
Loadsharing Type: Shar
System ID: 0x8000, 0023-8943-7524
Local:
  Port            Status  Priority Oper-Key  Flag
--------------------------------------------------------------------------------
  XGE1/0/2        S       32768    2         {ACDEF}
  XGE2/0/26       S       32768    2         {ACDEF}
Remote:
  Actor           Partner Priority Oper-Key  SystemID               Flag
--------------------------------------------------------------------------------
  XGE1/0/2        17      1        7         0x1   , d485-64ce-f033 {ABCDEF}
  XGE2/0/26       18      1        7         0x1   , d485-64ce-f033 {ABCDEF}
```

HP Flex-10: LACP

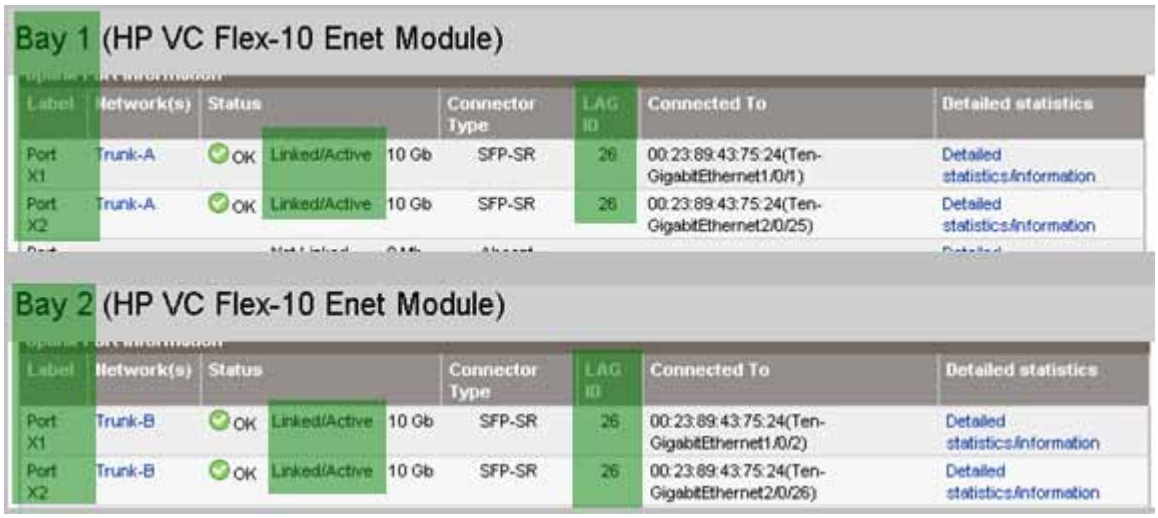
Trunk uplink configuration on VC1



Trunk uplink configuration on VC2



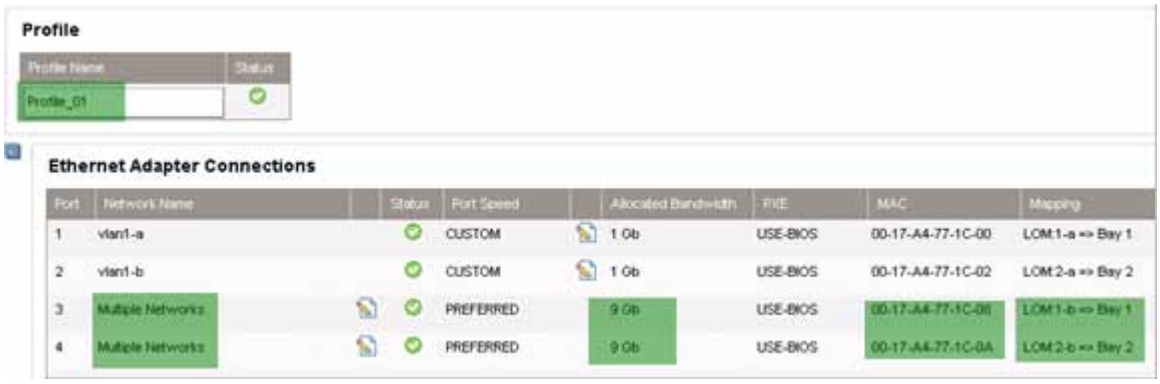
Trunk uplinks monitoring on VC



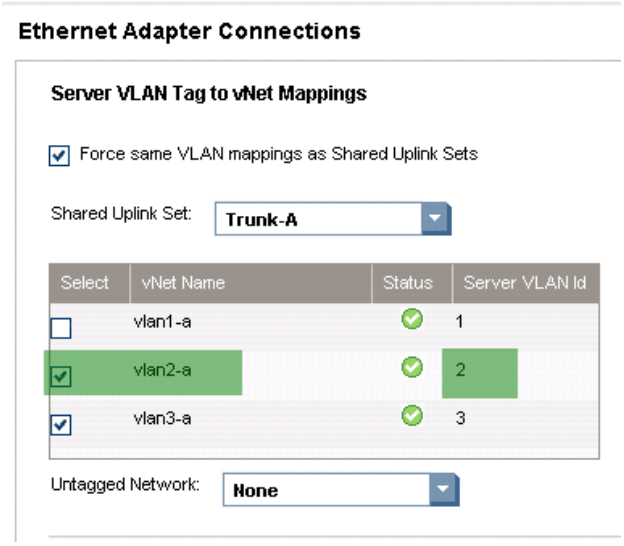
Both trunks show active/active. Also LLAG ID shows that an LACP bundle has been established with IRF virtual switch. Both channels use LAG 26. As they are on different modules, VC can uniquely identify them.

HP Flex-10: server profile

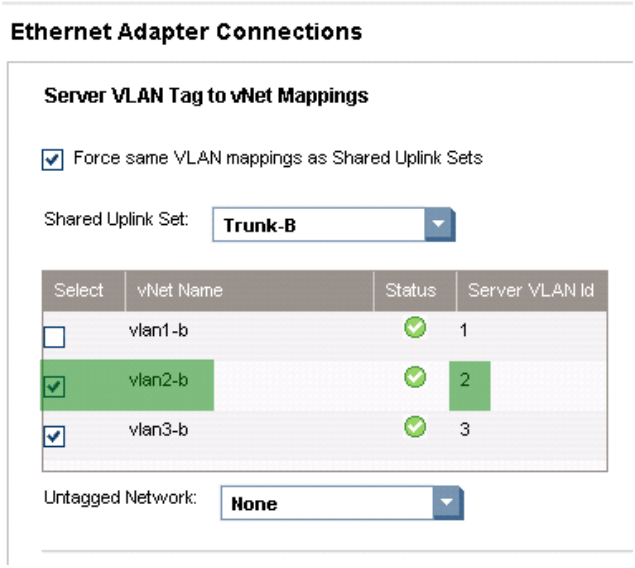
Server profile configuration



Port 3 Multiple Networks configuration



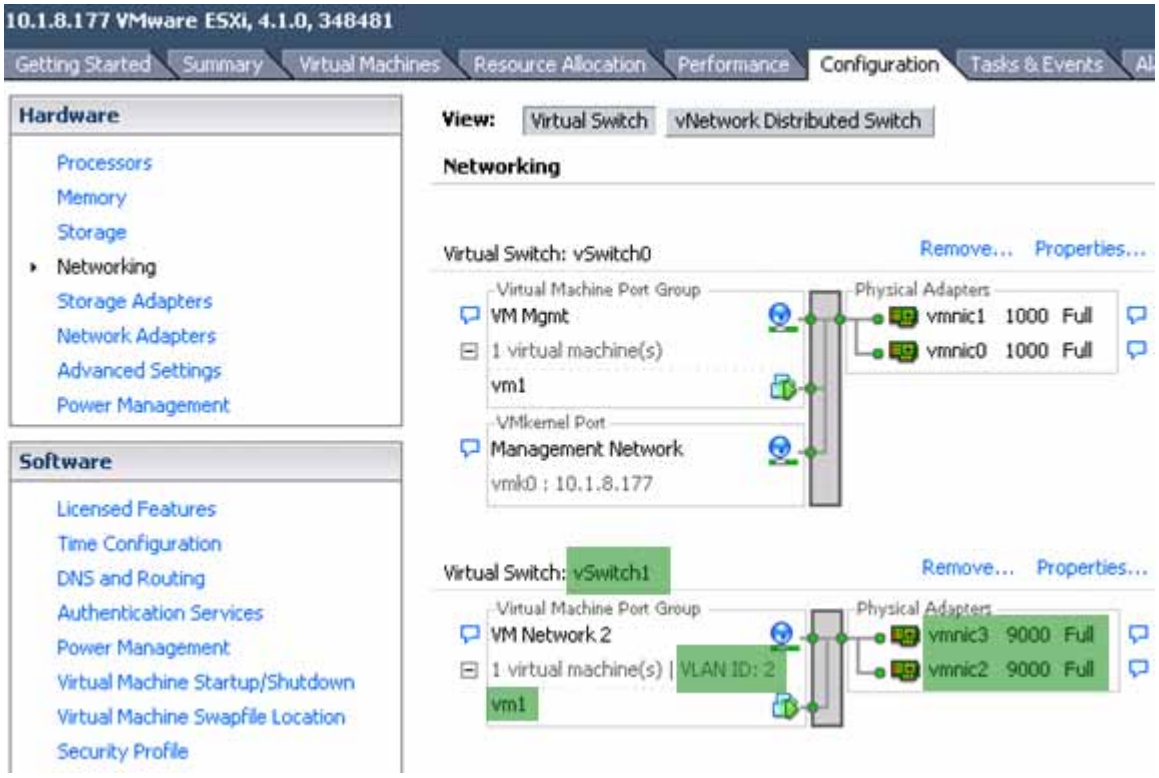
Port 4 Multiple Networks configuration



VMware ESXi configuration

Host adapter

Switch 1 configuration



VM1 network adapter configuration for VLAN2

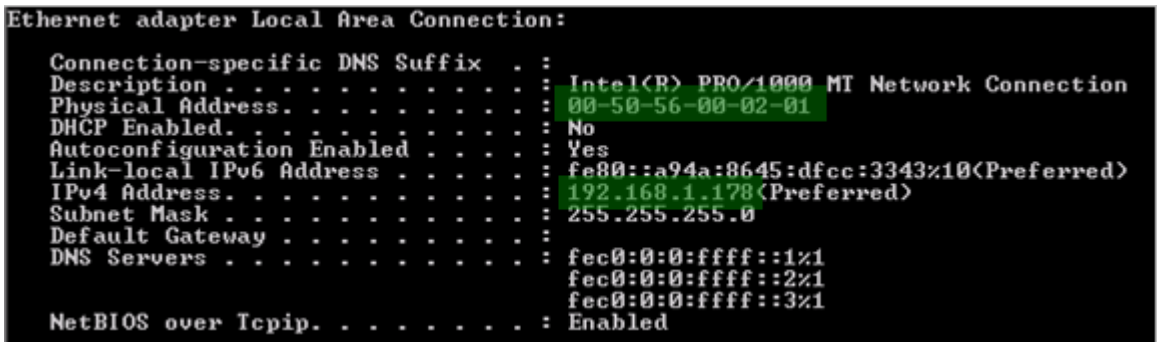
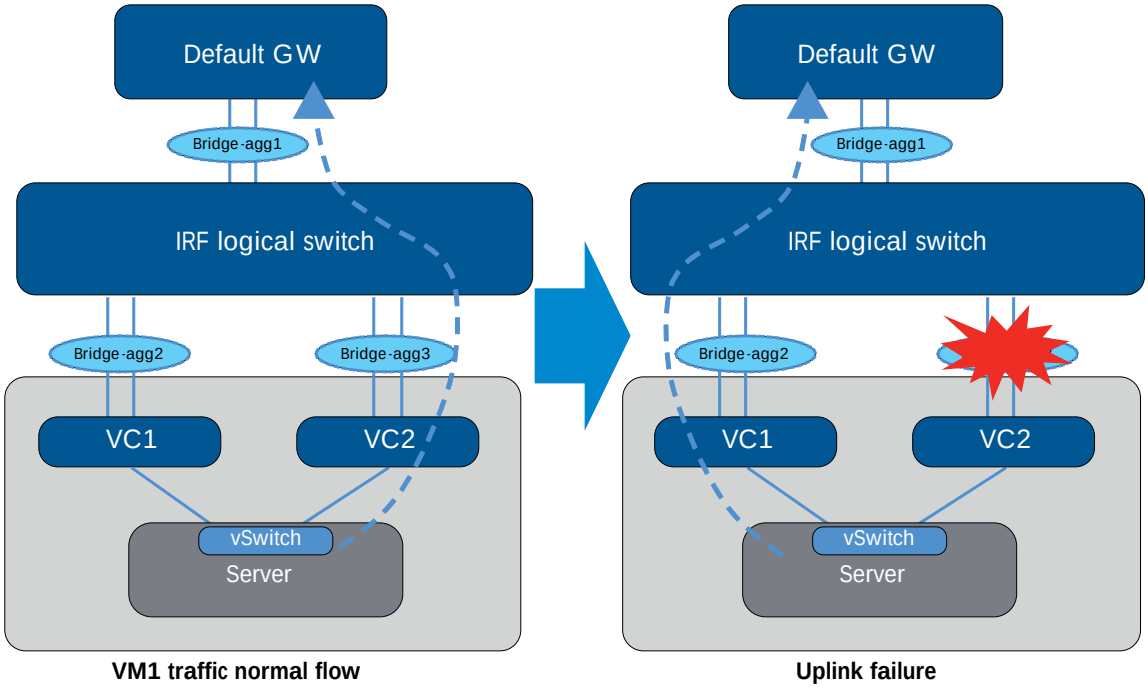


Figure 12. Uplink failure



VM1 has a continuous ping to its default GW 192.168.1.1. Under normal conditions, vSwitch hashes the traffic from this VM to the VM NIC3, which is mapped to the VC2 and then enters the bridge-aggregate 3 interface in the IRF logical switch.

The test issued a `shutdown` command under interface bridge-aggregation 3. From the `display MAC address` command, we can see the traffic failed over to the other path.

Test result:

- Shut down interface bridge-aggregation 3: about 3–4 seconds packets loss
- Undo the shutdown of interface bridge-aggregation 3: about 1–2 seconds packets loss with "stp edged-port enable" Without it, about 30 seconds of packet loss occurs due to the regular STP learning stage

Note:
IRF convergence time is much faster than three seconds, typically less than 50 microseconds. The overall three second convergence time is related to VC convergence around the smartlink to notify the server link in the event of uplink downtime, which then triggers vSwitch to converge the packet flow. Even with a regular switch without IRF (verified in the lab), three seconds is the expected VC / vSwitch convergence time in similar topology.

Shut interface bridge-aggregation 3

```
<A5820-IRF>
<A5820-IRF>dis mac-address dynamic vlan 2
MAC ADDR       VLAN ID  STATE          PORT INDEX               AGING TIME(s)
000a-5774-5f01 2        Learned        Bridge-Aggregation1      AGING
0050-5600-0201 2        Learned        Bridge-Aggregation3      AGING

  ---  2 mac address(es) found  ---

<A5820-IRF>int b3
           ^
 % Unrecognized command found at '^' position.
<A5820-IRF>sys
System View: return to User View with Ctrl+Z.
[A5820-IRF]int b3
[A5820-IRF-Bridge-Aggregation3]shut
[A5820-IRF-Bridge-Aggregation3]dis mac-add dynamic vlan 2
MAC ADDR       VLAN ID  STATE          PORT INDEX               AGING TIME(s)
000a-5774-5f01 2        Learned        Bridge-Aggregation1      AGING
0050-5600-0201 2        Learned        Bridge-Aggregation2      AGING

  ---  2 mac address(es) found  ---
```

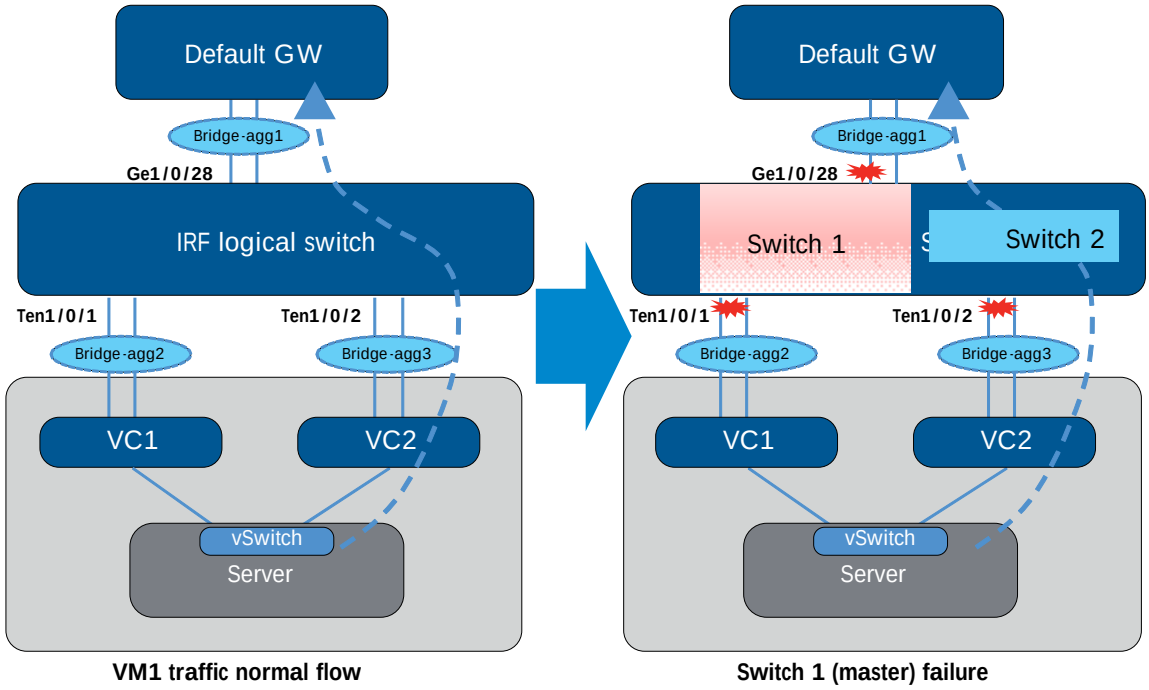
```
2011/03/11 12:18:55 : Reply[1785] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 12:18:55 : Reply[1786] from 192.168.1.1: bytes=32 time=1.6 ms TTL=64
2011/03/11 12:18:55 : Reply[1787] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 12:18:55 : Reply[1788] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 12:18:56 : Reply[1789] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 12:18:57 : 192.168.1.1: request timed out
2011/03/11 12:18:58 : 192.168.1.1: request timed out
2011/03/11 12:18:59 : 192.168.1.1: request timed out
2011/03/11 12:19:00 : 192.168.1.1: request timed out
2011/03/11 12:19:00 : Reply[1794] from 192.168.1.1: bytes=32 time=1.6 ms TTL=64
2011/03/11 12:19:00 : Reply[1795] from 192.168.1.1: bytes=32 time=1.9 ms TTL=64
2011/03/11 12:19:00 : Reply[1796] from 192.168.1.1: bytes=32 time=1.4 ms TTL=64
2011/03/11 12:19:00 : Reply[1797] from 192.168.1.1: bytes=32 time=1.4 ms TTL=64
2011/03/11 12:19:00 : Reply[1798] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
```

Undo the shutdown of interface bridge-aggregation 3

```
[A5820-IRF-Bridge-Aggregation3]dis mac-add dynamic vlan 2
MAC ADDR       VLAN ID  STATE          PORT INDEX               AGING TIME(s)
000a-5774-5f01 2        Learned        Bridge-Aggregation1      AGING
0050-5600-0201 2        Learned        Bridge-Aggregation3      AGING

  ---  2 mac address(es) found  ---

[A5820-IRF-Bridge-Aggregation3]int b3
[A5820-IRF-Bridge-Aggregation3] shut
[A5820-IRF-Bridge-Aggregation3]dis mac-add dynamic vlan 2
MAC ADDR       VLAN ID  STATE          PORT INDEX               AGING TIME(s)
000a-5774-5f01 2        Learned        Bridge-Aggregation1      AGING
0050-5600-0201 2        Learned        Bridge-Aggregation2      AGING

  ---  2 mac address(es) found  ---

[A5820-IRF-Bridge-Aggregation3]undo shut
[A5820-IRF-Bridge-Aggregation3]dis mac-add dynamic vlan 2
MAC ADDR       VLAN ID  STATE          PORT INDEX               AGING TIME(s)
000a-5774-5f01 2        Learned        Bridge-Aggregation1      AGING
0050-5600-0201 2        Learned        Bridge-Aggregation3      AGING
```

```
2011/03/11 12:26:45 : Reply[482] from 192.168.1.1: bytes=32 time=1.4 ms TTL=64
2011/03/11 12:26:46 : 192.168.1.1: request timed out
2011/03/11 12:26:47 : 192.168.1.1: request timed out
2011/03/11 12:26:47 : Reply[485] from 192.168.1.1: bytes=32 time=1.8 ms TTL=64
2011/03/11 12:26:47 : Reply[486] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
```

Figure 13. Switch failure



VM1 has a continuous ping to its default GW 192.168.1.1. Under normal conditions, vSwitch hashes the traffic from this VM to the VM NIC3, which is mapped to VC2 and then enters the bridge-aggregation 3 interface in the IRF logical switch.

The test issues a `reboot` command on switch 1 5820. Switch 2 takes over as the new master and any interface related to `1/y/z` is shut down.

Test result:

Switch 1 down: Ping packet loss did not occur, so the convergence time was less than one second.

Switch 1 up: Ping packet loss did not occur, so the convergence time was less than one second.

Note:

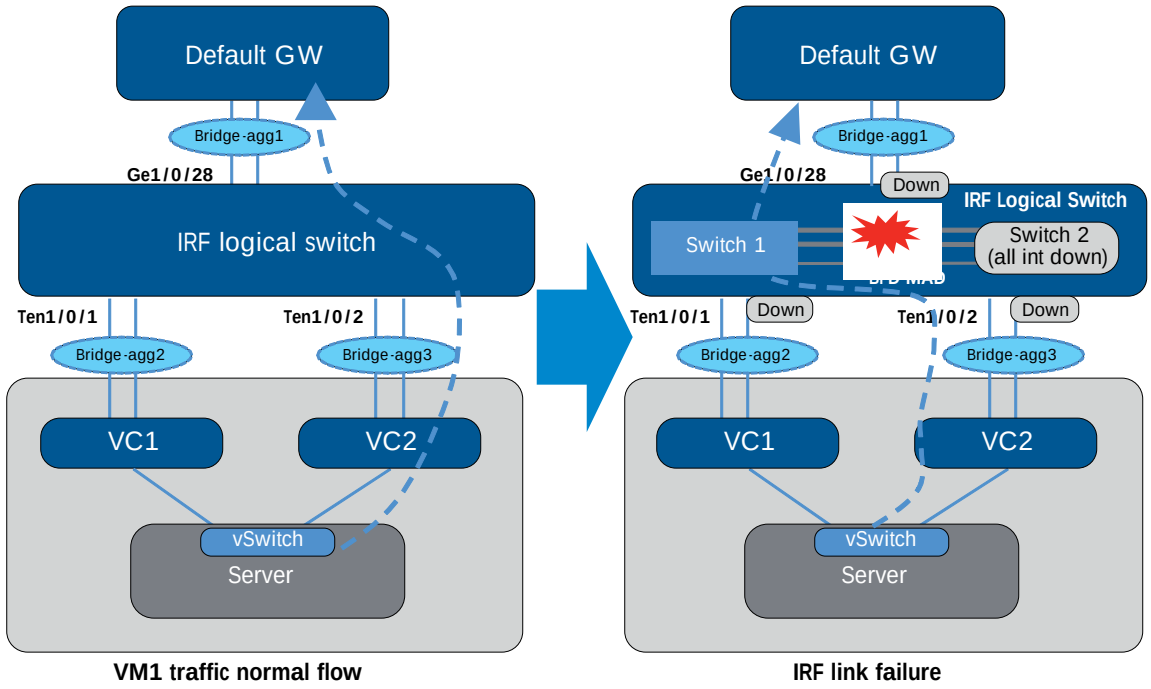
The convergence time remained less than one second because the traffic flow did not switch over to the other path. It's still used interface bridge-aggregation 3 because even with switch 1 and all `1/y/z` interfaces down, interface bridge-aggregation 3 still had the other interface ten`2/0/26` up. So, the convergence time is the result of LACP rehashing the traffic to the other remaining link, which is typically less than one second.

For this scenario, IRF does not change the traffic flow path, even when losing one switch. The two uplinks operate at 10 GB/s each.

After switch 1 comes back up, it remains the slave to prevent traffic switchover again, even though it has higher priority.

```
<A5820-IRF>dis irf
Switch  Role    Priority  CPU-Mac         Description
   1    Slave   10        0023-8943-7525  -----
*+2    Master  1         0023-893c-45d6  -----
--------------------------------------------------

* indicates the device is the master.
+ indicates the device through which the user logs in.

The Bridge MAC of the IRF is: 0023-8943-7524
Auto upgrade                : yes
Mac persistent              : 6 min
Domain ID                   : 0
```

Figure 14. IRF link failure



VM1 has a continuous ping to its default GW 192.168.1.1. Under normal conditions, the vSwitch hashes the traffic from this VM to the VM NIC3, which is mapped to the VC2, and then enters the bridge-aggregation 3 interface in the IRF logical switch.

The test issued a shut down command under switch 1 5820 IRF1/2 to simulate IRF link failure.

Test result:

- Shut irf-port 1/2: Ping packet loss did not occur, so the convergence time was less than one second.
- Do unshut of the irf-port 1/2: About one second packet loss after switch 2 rebooted and came back up to join the IRF domain

Note:
Upon losing the IRF link, MAD initiates and elects one master for the domain, and the other switch (switch 2 with lower IRF priority) shuts down all its local interfaces to prevent a dual active (split brain) scenario. When the IRF link is restored, switch 2 reboots itself and rejoins the IRF domain.

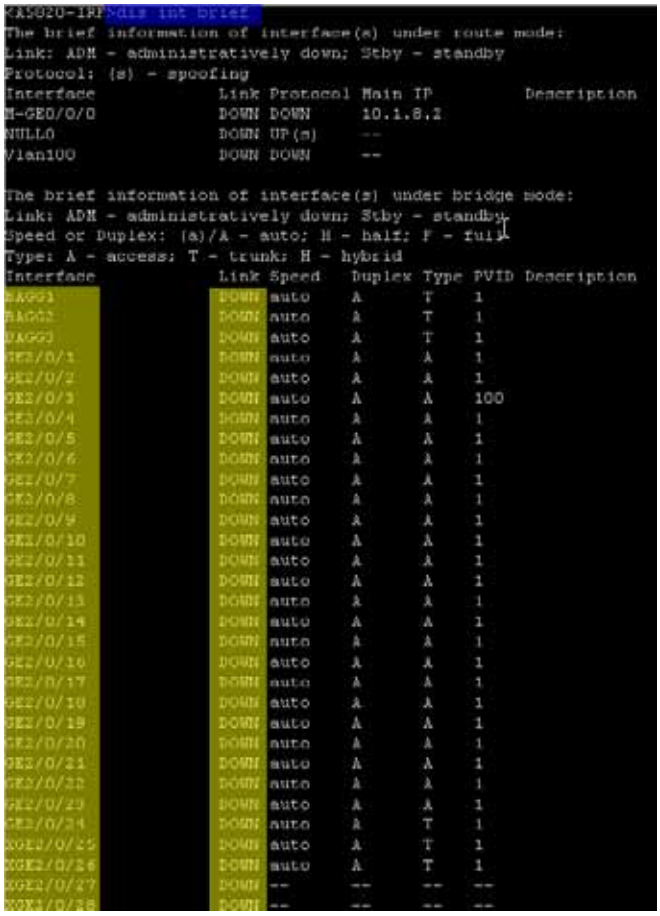
Packet loss when switch 2 (5800) came back and joined IRF domain:

```
2011/03/11 14:14:19 : Reply[1694] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 14:14:19 : Reply[1695] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/11 14:14:20 : 192.168.1.1: request timed out
2011/03/11 14:14:20 : Reply[1697] from 192.168.1.1: bytes=32 time=2.6 ms TTL=64
2011/03/11 14:14:20 : Reply[1698] from 192.168.1.1: bytes=32 time=1.6 ms TTL=64
```

Switch 2 (5800) view after IRF link failure with BFD MAD protection

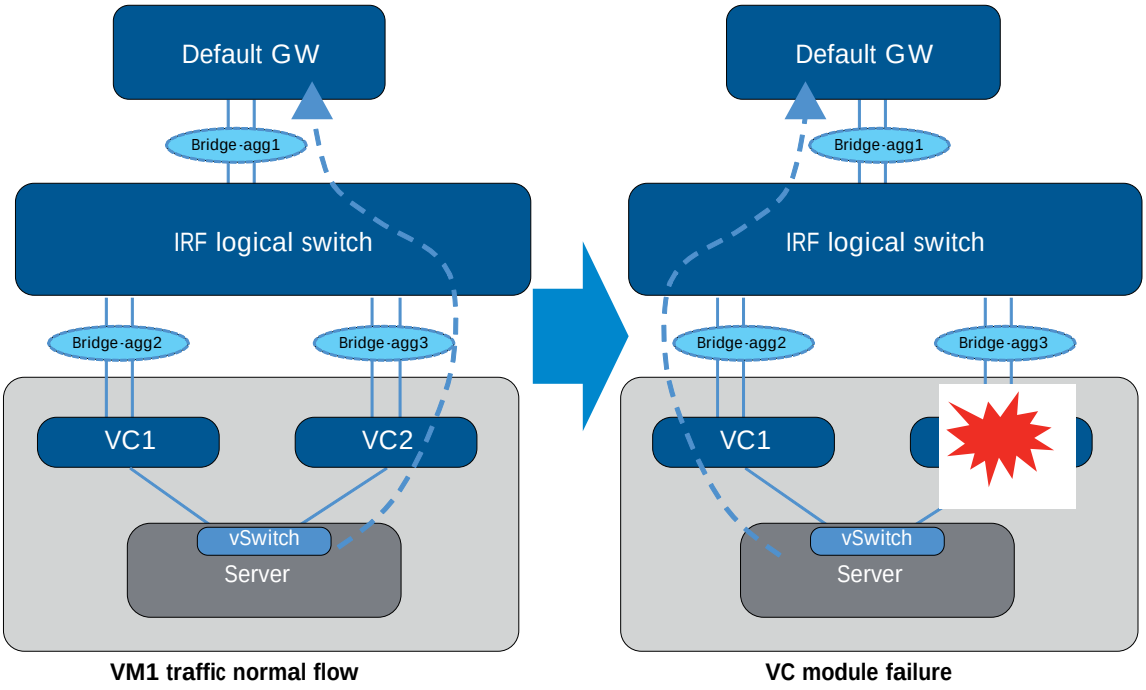
```
<A5820-IRF>dis irf
Switch   Role    Priority  CPU-Mac         Description
 *+2     Master  1         0023-893c-45d6  -----
-------------------------------------------------

 * indicates the device is the master.
 + indicates the device through which the user logs in.

 The Bridge MAC of the IRF is: 0023-8943-7524
 Auto upgrade                : yes
 Mac persistent              : 6 min
 Domain ID                   : 0
<A5820-IRF>dis mad ver
Current MAD status: Recovery
Excluded ports(configurable):
Excluded ports(can not be configured):
  Ten-GigabitEthernet2/0/27
  Ten-GigabitEthernet2/0/28
MAD ARP disabled.
MAD LACP disabled.
MAD BFD enabled interface:
  Vlan-interface100
    mad ip address 100.100.100.1 255.255.255.0 member 1
    mad ip address 100.100.100.2 255.255.255.0 member 2
<A5820-IRF>
```

Switch 2 (5800) view after all local interfaces were shut down to prevent a dual active scenario



Switch 1 (5820) console log after the IRF link failed

Figure 15. VC module failure



VM1 traffic normal flow **VC module failure**

VM1 has a continuous ping to its default GW 192.168.1.1. Under normal conditions, the vSwitch hashes the traffic from this VM to the VM NIC3, which is mapped to the VC2 and enters the bridge-aggregation 3 interface in IRF the logical switch.

The test uses the power-off button on HP Onboard Administrator (OA) to shut down VC2 to simulate module failure.

Test result:

- VC2 down: about one second packet loss
- VC2 up: about six seconds packet loss

Note:
The VC2 up event had more convergence time because the VM NIC3, which is mapped to VC2, was up. Therefore, the vSwitch started to send traffic to VC2 before VC2 was ready internally for switching traffic.

VC2 down

```
2011/03/16 11:31:52 : Reply[986] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:31:53 : Reply[987] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:31:54 : Reply[988] from 192.168.1.1: bytes=32 time=1.4 ms TTL=64
2011/03/16 11:31:55 : Reply[989] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:31:57 : 192.168.1.1: request timed out
2011/03/16 11:31:57 : Reply[991] from 192.168.1.1: bytes=32 time=1.4 ms TTL=64
2011/03/16 11:31:58 : Reply[992] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
```

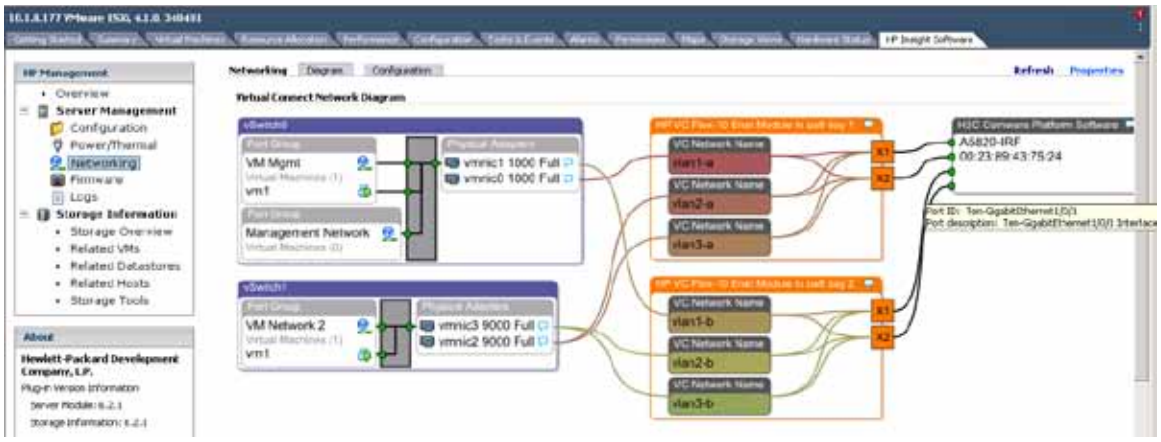
VC2 up

```
2011/03/16 11:34:06 : Reply[1119] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:34:07 : Reply[1120] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:34:08 : Reply[1121] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:34:10 : 192.168.1.1: request timed out
2011/03/16 11:34:11 : 192.168.1.1: request timed out
2011/03/16 11:34:12 : 192.168.1.1: request timed out
2011/03/16 11:34:13 : 192.168.1.1: request timed out
2011/03/16 11:34:14 : 192.168.1.1: request timed out
2011/03/16 11:34:15 : 192.168.1.1: request timed out
2011/03/16 11:34:15 : Reply[1128] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:34:16 : Reply[1129] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
2011/03/16 11:34:17 : Reply[1130] from 192.168.1.1: bytes=32 time=1.5 ms TTL=64
```

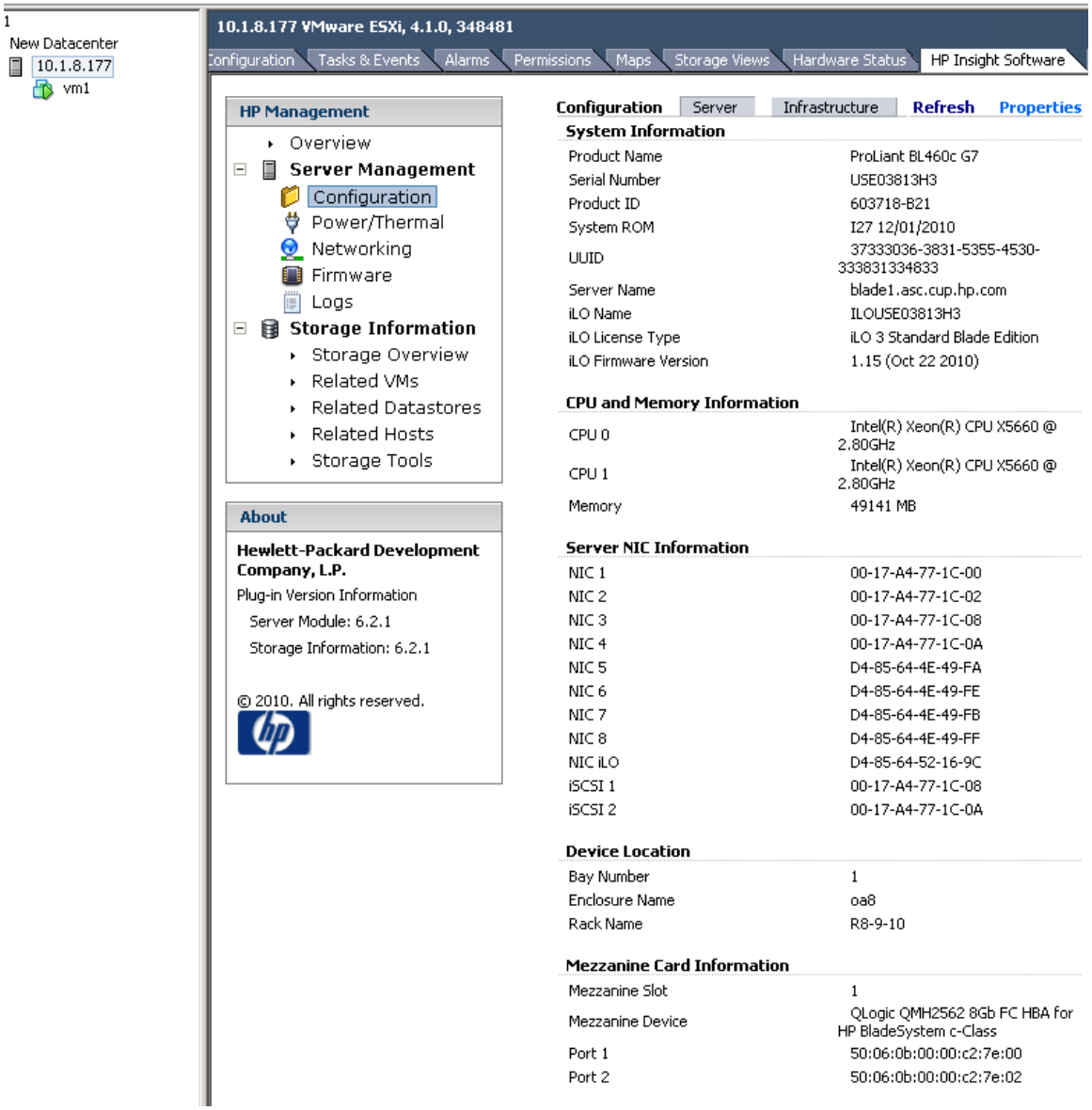
HP Insight Control for VMware vCenter monitoring

HP Insight Control for VMware vCenter utilizes a visual networking view from vSwitch to VC to physical access switch. The following images provide examples of its appearance and functionality.

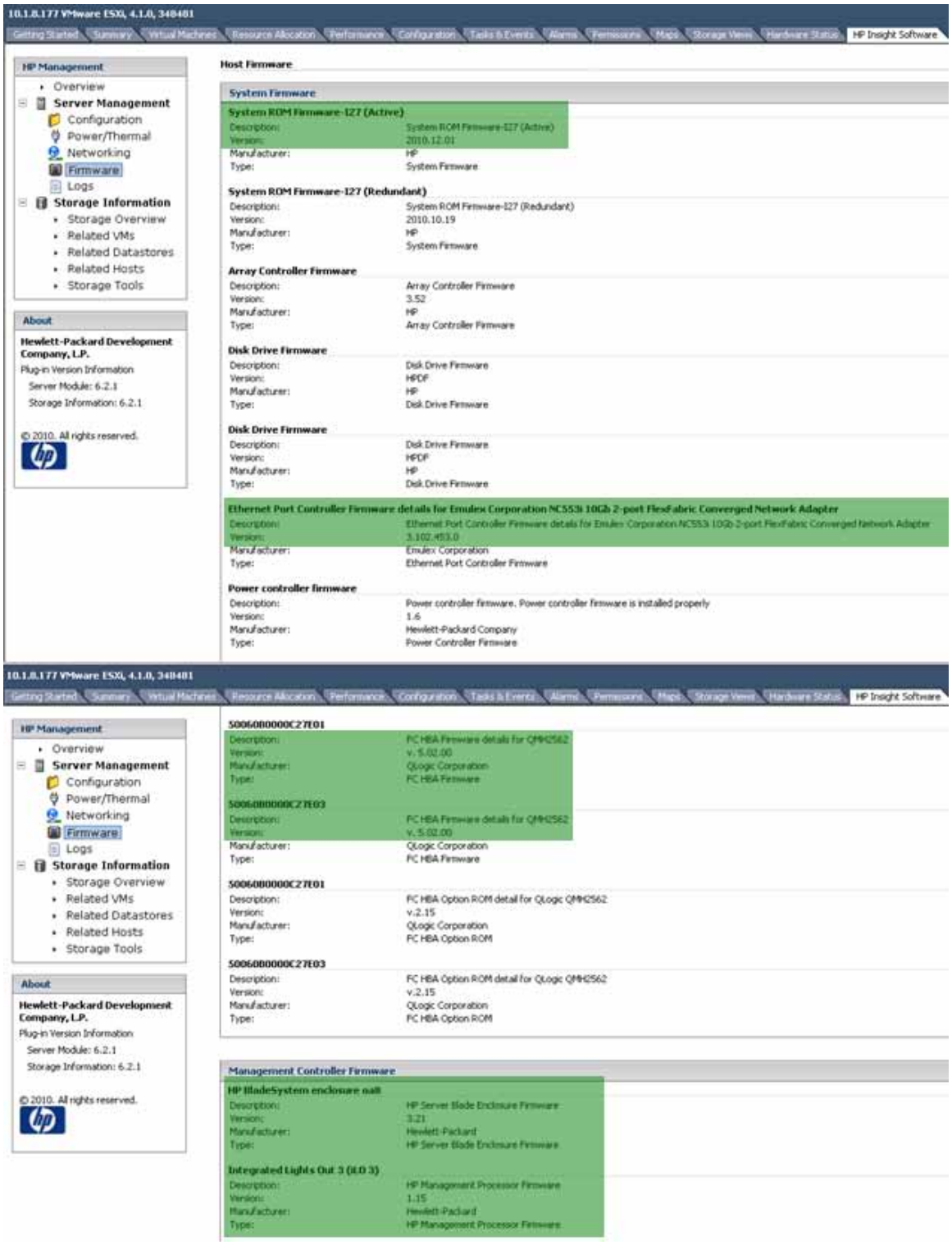
VM1 uses vSwitch 1, which has two uplinks (VM NIC2 and VM NIC3). The uplinks carry tagged packets for VLAN2 and VLAN3. VLAN3 is not in used in the testing but is provided to show the concept of tagged trunking between VC and vSwitch. The graphic also displays the physical uplink ports used to connect to the access switch. The host name and MAC address of that switch are also provided, and are obtained through the use of LLDP between VC and the network switch.



Host hardware inventory details



Host and enclosure firmware version report



HP IMC network management

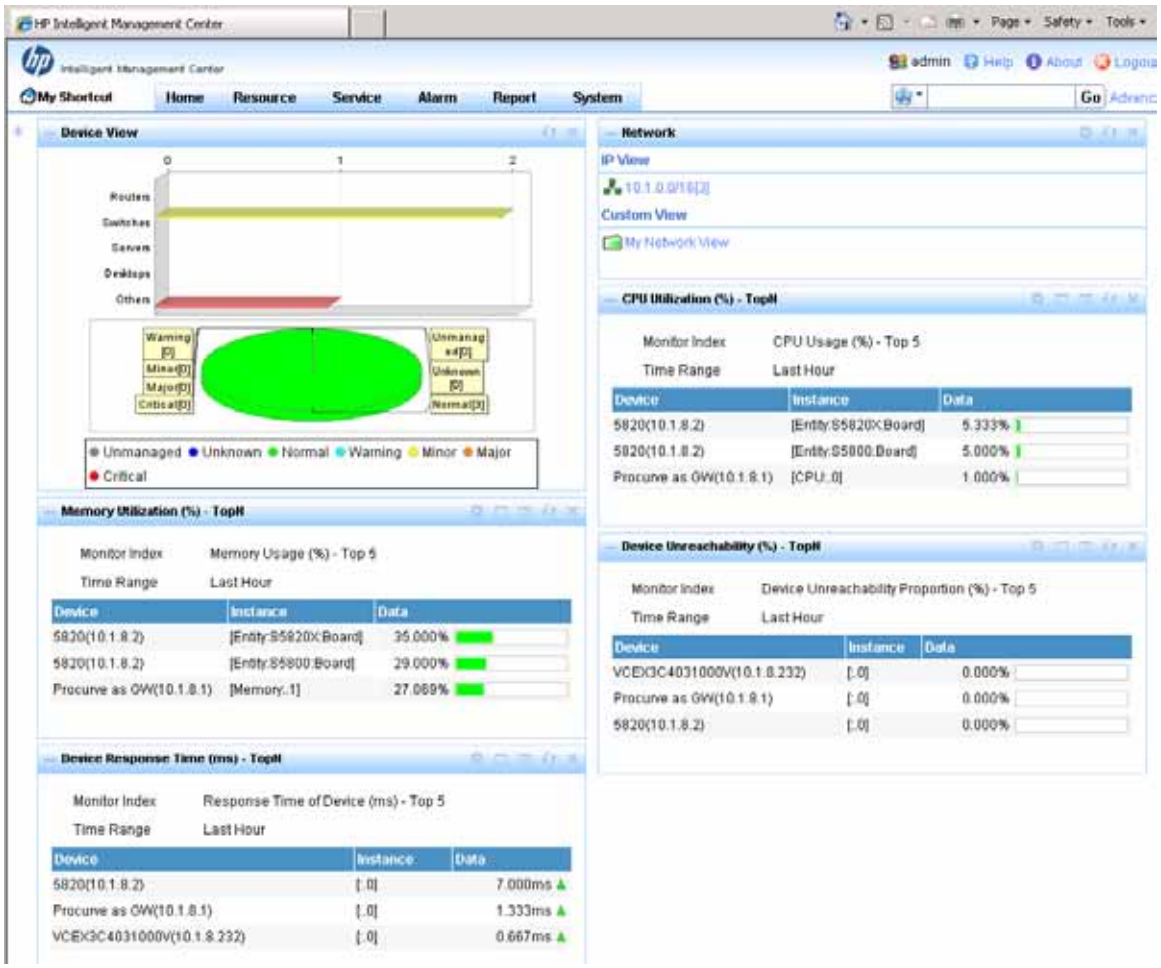
IMC is an HP Networking management software that supports network device configuration, accounting, performance, security management, and monitoring. It can manage HP network devices, as well as routers and switches from other vendors.

The following images corresponding to this setup provide an overview of the appearance and functionality of IMC. It does not represent the full functionality of IMC.

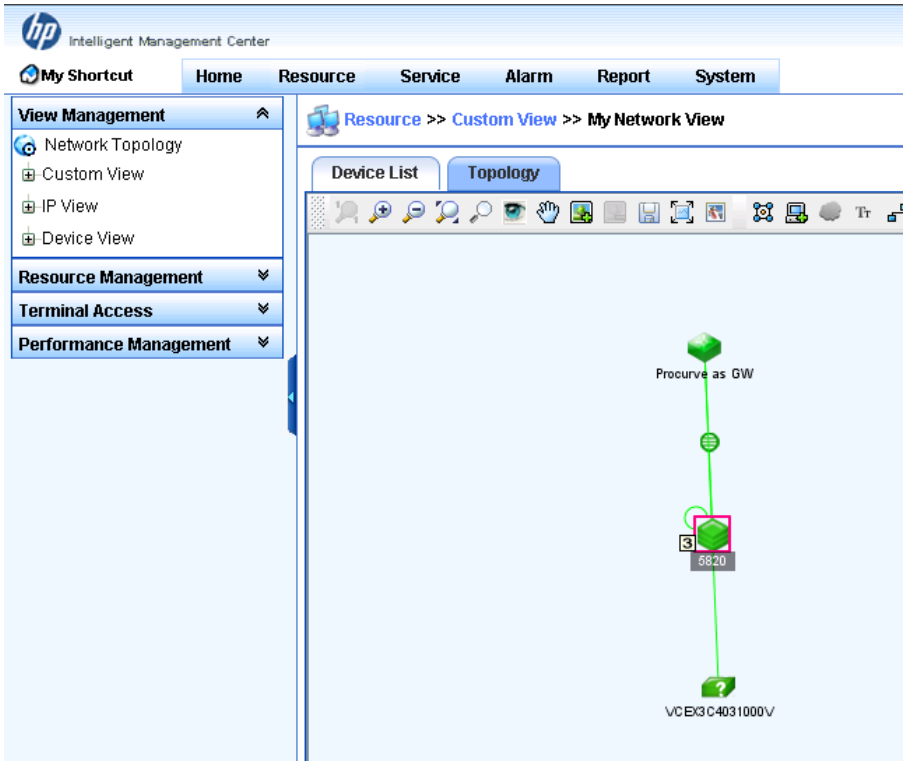
For more information on HP IMC, visit:
http://h17007.www1.hp.com/us/en/products/network-management/index.aspx

To the download full-featured evaluation software, visit:
https://h10145.www1.hp.com/downloads/SoftwareReleases.aspx?ProductNumber=JF377A&lang=en&cc=us&prodSeriesId=4176535

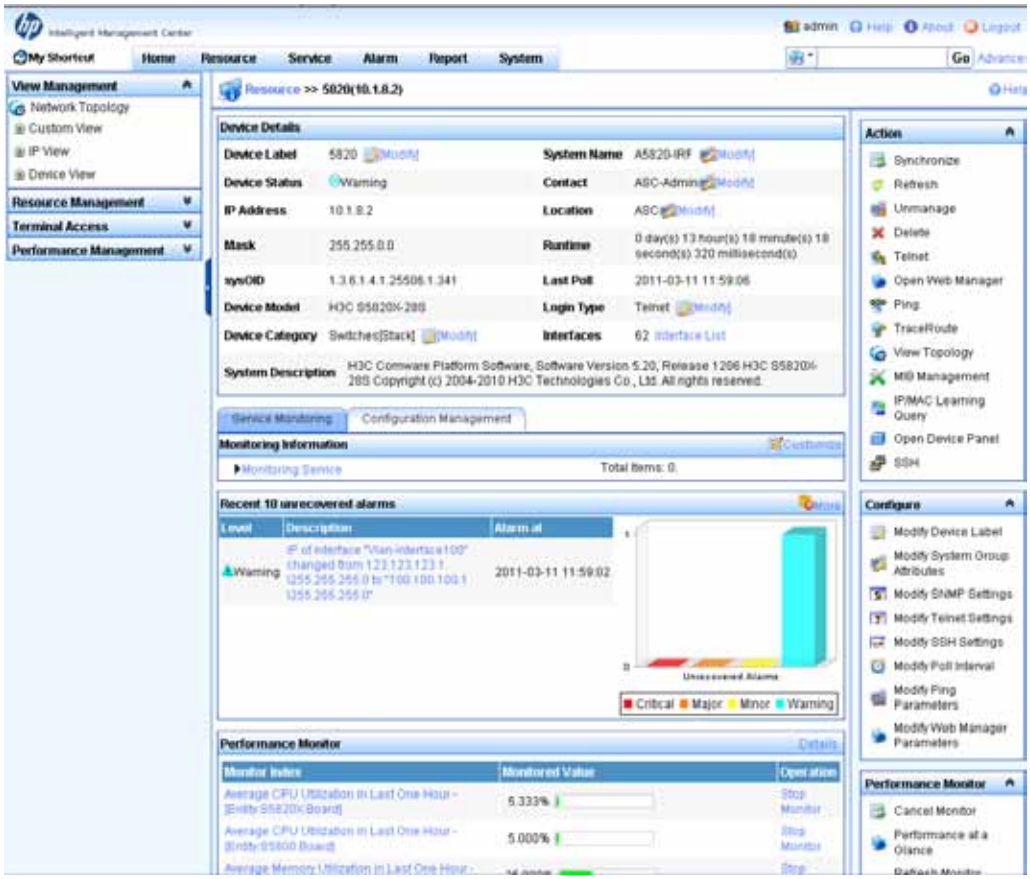
Overview page (can customize layout)



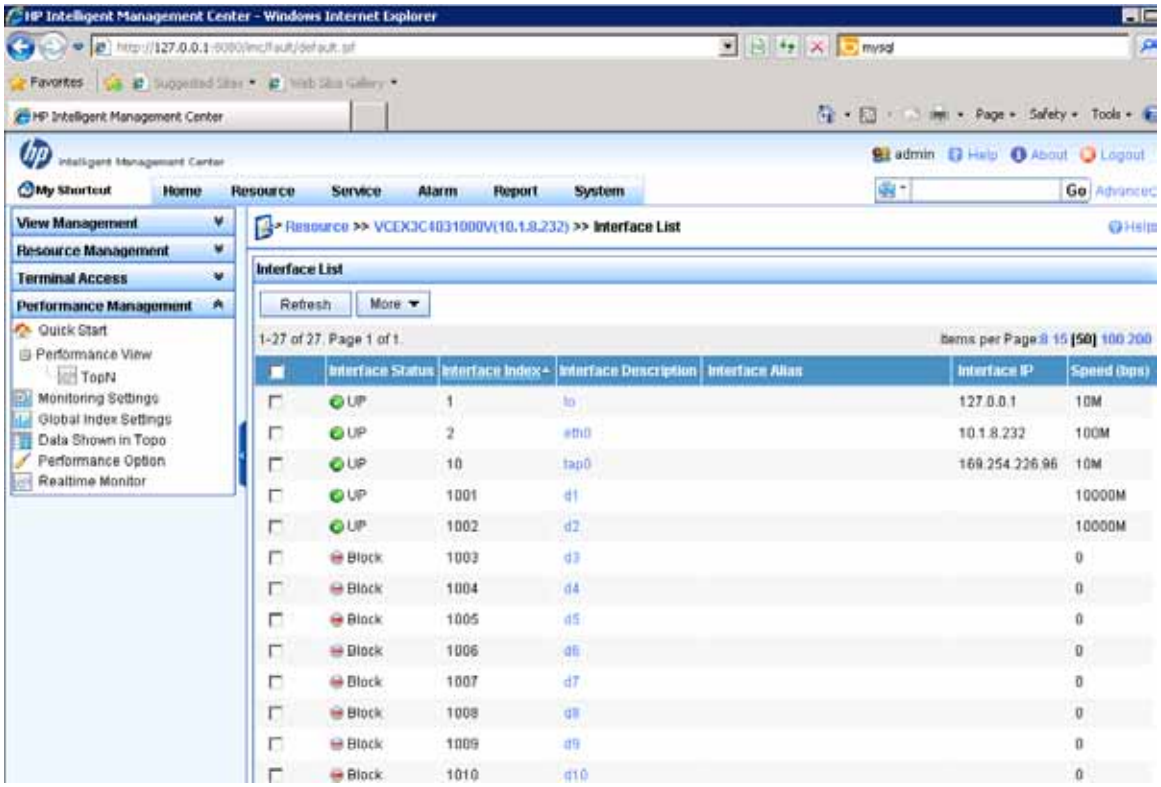
Network topology



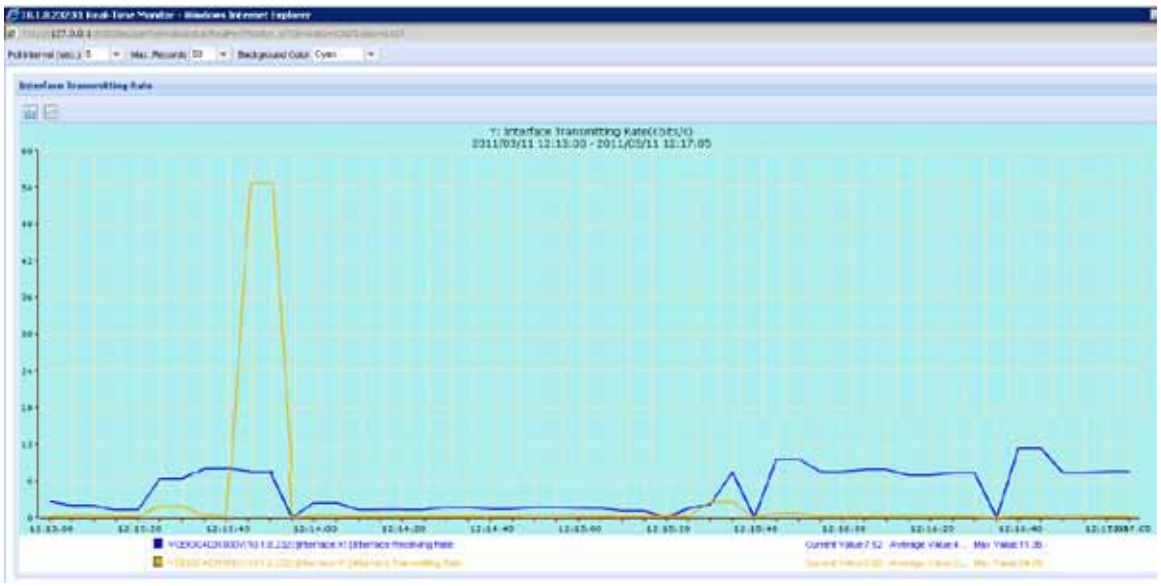
5820 IRF logical switch



VC interface list view



Interface traffic rate realtime monitoring



Summary

Today's data center networks must be designed to deliver much higher levels of performance, scalability, and availability than before to meet service-level agreements and maintain continuity of operations. Beyond sheer performance, these data center networks must quickly recover from hardware- or software-related faults to minimize service disruptions.

As discussed in this white paper, the IRF-based HP switch and HP VC server edge solution can easily address this requirement to recover from failure quickly with, in many cases, a loss of zero packets.

Appendix 1: HP 5820 Switch Series IRF configuration

```
[5820-IRF]dis current-configuration
#
 version 5.20, Release 1206
#
 sysname 5820-IRF
#
 irf mac-address persistent timer
 irf auto-update enable
 undo irf link-delay
 irf member 1 priority 10
#
 domain default enable system
#
 telnet server enable
#
vlan 1
#
vlan 2
#
vlan 100
#
radius scheme system
 server-type extended
 primary authentication 127.0.0.1 1645
 primary accounting 127.0.0.1 1646
 user-name-format without-domain
#
domain system
 access-limit disable
 state active
 idle-cut disable
 self-service-url disable
#
user-group system
#
 stp mode rstp
 stp enable
#
interface Bridge-Aggregation1
 port link-type trunk
 port trunk permit vlan 1 to 2
 link-aggregation mode dynamic
#
interface Bridge-Aggregation2
 port link-type trunk
 port trunk permit vlan 1 to 2
 link-aggregation mode dynamic
 stp edged-port enable
#
interface Bridge-Aggregation3
 port link-type trunk
 port trunk permit vlan 1 to 2
 link-aggregation mode dynamic
 stp edged-port enable
#
interface NULL0
```

```
#
interface Vlan-interface100
 mad bfd enable
 mad ip address 100.100.100.1 255.255.255.0 member 1
 mad ip address 100.100.100.2 255.255.255.0 member 2
#
interface GigabitEthernet1/0/25
 port link-mode bridge
 port access vlan 100
 stp disable
#
interface GigabitEthernet1/0/26
 port link-mode bridge
#
interface GigabitEthernet1/0/27
 port link-mode bridge
#
interface GigabitEthernet1/0/28
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 1
#
interface GigabitEthernet2/0/1
 port link-mode bridge
#
interface GigabitEthernet2/0/2
 port link-mode bridge
#
interface GigabitEthernet2/0/3
 port link-mode bridge
 port access vlan 100
 stp disable
#
interface GigabitEthernet2/0/4
 port link-mode bridge
#
interface GigabitEthernet2/0/5
 port link-mode bridge
#
interface GigabitEthernet2/0/6
 port link-mode bridge
#
interface GigabitEthernet2/0/7
 port link-mode bridge
#
interface GigabitEthernet2/0/8
 port link-mode bridge
#
interface GigabitEthernet2/0/9
 port link-mode bridge
#
interface GigabitEthernet2/0/10
 port link-mode bridge
#
interface GigabitEthernet2/0/11
 port link-mode bridge
#
interface GigabitEthernet2/0/12
```

```
 port link-mode bridge
#
interface GigabitEthernet2/0/13
 port link-mode bridge
#
interface GigabitEthernet2/0/14
 port link-mode bridge
#
interface GigabitEthernet2/0/15
 port link-mode bridge
#
interface GigabitEthernet2/0/16
 port link-mode bridge
#
interface GigabitEthernet2/0/17
 port link-mode bridge
#
interface GigabitEthernet2/0/18
 port link-mode bridge
#
interface GigabitEthernet2/0/19
 port link-mode bridge
#
interface GigabitEthernet2/0/20
 port link-mode bridge
#
interface GigabitEthernet2/0/21
 port link-mode bridge
#
interface GigabitEthernet2/0/22
 port link-mode bridge
#
interface GigabitEthernet2/0/23
 port link-mode bridge
#
interface GigabitEthernet2/0/24
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 1
#
interface M-GigabitEthernet0/0/0
 ip address 10.1.8.2 255.255.0.0
#
interface Ten-GigabitEthernet1/0/1
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 2
#
interface Ten-GigabitEthernet1/0/2
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 3
#
interface Ten-GigabitEthernet1/0/3
 port link-mode bridge
#
```

```
interface Ten-GigabitEthernet1/0/4
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/5
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/6
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/7
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/8
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/9
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/10
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/11
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/12
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/13
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/14
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/15
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/16
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/17
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/18
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/19
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/20
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/21
 port link-mode bridge
#
interface Ten-GigabitEthernet1/0/22
 port link-mode bridge
#
interface Ten-GigabitEthernet2/0/25
 port link-mode bridge
```

```
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 2
#
interface Ten-GigabitEthernet2/0/26
 port link-mode bridge
 port link-type trunk
 port trunk permit vlan 1 to 2
 port link-aggregation group 3
#
interface Ten-GigabitEthernet1/0/23
#
interface Ten-GigabitEthernet1/0/24
#
interface Ten-GigabitEthernet2/0/27
#
interface Ten-GigabitEthernet2/0/28
#
 ip route-static 0.0.0.0 0.0.0.0 10.1.0.1
#
 snmp-agent
 snmp-agent local-engineid 800063A203002389437528
 snmp-agent community read public
 snmp-agent sys-info contact ASC-Admin
 snmp-agent sys-info location ASC
 snmp-agent sys-info version all
 snmp-agent target-host trap address udp-domain 10.1.220.178 udp-port 161 params securityname
public
 snmp-agent trap enable default-route
#
 load xml-configuration
#
user-interface aux 0 1
user-interface vty 0 15
 authentication-mode none
 user privilege level 3
#
irf-port 1/2
 port group interface Ten-GigabitEthernet1/0/23 mode enhanced
 port group interface Ten-GigabitEthernet1/0/24 mode enhanced
#
irf-port 2/1
 port group interface Ten-GigabitEthernet2/0/27 mode enhanced
 port group interface Ten-GigabitEthernet2/0/28 mode enhanced
#
return
[5820-IRF]
```

Appendix 2: running status of design three

Using design option three, if one port channel interface is configured on a 5820 switch and one SUS is configured on a VC to bundle four links on both sides, the 5820 switch does not select two out of four links as active LACP links.

```
[A5820-IRF-Ten-GigabitEthernet2/0/26]dis link-aggregation ver b2

Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Port Status: S -- Selected, U -- Unselected
Flags:  A -- LACP_Activity, B -- LACP_Timeout, C -- Aggregation,
        D -- Synchronization, E -- Collecting, F -- Distributing,
        G -- Defaulted, H -- Expired

Aggregation Interface: Bridge-Aggregation2
Aggregation Mode: Dynamic
Loadsharing Type: Shar
System ID: 0x8000, 0023-8943-7524
Local:
  Port            Status  Priority Oper-Key  Flag
--------------------------------------------------------------------------------
  XGE1/0/1        S       32768    1         {ACDEF}
  XGE1/0/2        U       32768    1         {AC}
  XGE2/0/25       S       32768    1         {ACDEF}
  XGE2/0/26       U       32768    1         {AC}
Remote:
  Actor           Partner Priority Oper-Key  SystemID               Flag
--------------------------------------------------------------------------------
  XGE1/0/1        17      1        3         0x1   , d485-64ce-f015 {ABCDEF}
  XGE1/0/2        17      1        3         0x1   , d485-64ce-f033 {ABCD}
  XGE2/0/25       18      1        3         0x1   , d485-64ce-f015 {ABCDEF}
  XGE2/0/26       18      1        3         0x1   , d485-64ce-f033 {ABCD}
```

VC1 status is OK.

Bay 1 (HP VC Flex-10 Enet Module)

Manufacturer:	HP
Firmware Rev.	3.15 2010-10-09T07:18:16Z

Uplink Port Information

Label	Network (s)	Status		Connector Type	LAG ID	Connected To	Detailed statistics
Port X1	Trunk-A	OK	Linked/Active 10 Gb	SFP-SR	25	00:23:89:43:75:24(Ten-GigabitEthernet1/0/1)	Detailed statistics/information
Port X2	Trunk-A	OK	Linked/Active 10 Gb	SFP-SR	25	00:23:89:43:75:24(Ten-GigabitEthernet2/0/25)	Detailed statistics/information

VC2 status is not OK. The LAG ID column is empty, which means no LACP bundle is established. Both links are put into standby as individual links for this SUS.

Bay 2 (HP VC Flex-10 Enet Module)

Port X1	Trunk-A	OK	Linked/Standby 10 Gb	SFP-SR		00:23:89:43:75:24(Ten-GigabitEthernet1/0/2)	Detailed statistics/information
Port X2	Trunk-A	OK	Linked/Standby 10 Gb	SFP-SR		00:23:89:43:75:24(Ten-GigabitEthernet2/0/26)	Detailed statistics/information

Acronyms

ARP—Address Resolution Protocol

BFD—Bidirectional Forwarding Detection

BPDU—Bridge Protocol Data Unit

GW—Gateway

IMC—Intelligent Management Center

IRF—Intelligent Resilient Framework

LACP—Link Aggression Control Protocol

LACPDU—Link Aggression Control Protocol Data Unit

LLDP—Link Layer Discovery Protocol

MAC—media access control

MAD—multi-active detection

OA—Onboard Administrator

SSH—secure shell

STP—Spanning Tree Protocol

SUS—Shared Uplink Set

VC—Virtual Connect

VLAN—virtual local area network

VM—Virtual Machine

vPC—Virtual Port Channel

VSS—Virtual Switching System

LAG—Link Aggregation Group

Respond to your evolving enterprise needs quickly and cost-effectively through the integrated solution of HP VC with HP switches and HP IRF, while reducing cost. Visit, http://h17007.www1.hp.com/us/en/solutions/datacenter/index.aspx

4AA3-7337ENW, Created April 2012