Preface

This preface describes the audience, organization, and conventions of the *Cisco Nexus 5000 Series NX-OS Interfaces Operations Guide, Release 5.0(3)N2(1)*. It also provides information on how to obtain related documentation.

This chapter includes the following topics:

Audience

This publication is for experienced network administrators who configure and maintain Cisco NX-OS on Cisco Nexus 5000 Platform switches and Cisco Nexus 5500 Platform switches.

Document Conventions

Command descriptions use the following conventions:

Convention	Description
bold	Bold text indicates the commands and keywords that you enter literally as shown.
Italic	Italic text indicates arguments for which the user supplies the values.
[x]	Square brackets enclose an optional element(keyword or argument).
[x | y]	Square brackets enclosing keywords or arguments separated by a vertical bar indicate an optional choice.
{x | y}	Braces enclosing keywords or arguments separated by a vertical bar indicate a required choice.

[x {y | z}]	Nested set of square brackets or braces indicate optional or required choices within optional or required elements. Braces and a vertical bar within square brackets indicate a required choice within an optional element.
variable	Indicates a variable for which you supply values, in context where italics cannot be used.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.

Screen examples use the following conventions::

Convention	Description
screen font	Terminal sessions and information the switch displays are in screen font.
boldface screen font	Information you must enter is in boldface screen font.
italic screen font	Arguments for which you supply values are in italic screen font.
< >	Nonprinting characters, such as passwords, are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the manual.

Caution Means *reader be careful*. In this situation, you might do something that could result in equipment damage or loss of data.

Related Documentation

Documentation for Cisco Nexus 5000 Series Switches and Cisco Nexus 2000 Series Fabric Extenders is available at the following URL:

http://www.cisco.com/en/US/products/ps9670/tsd_products_support_series_home.html

Obtaining Documentation and Submitting a Service Request

For information on obtaining documentation, submitting a service request, and gathering additional information, see the monthly *What's New in Cisco Product Documentation*, which also lists all new and revised Cisco technical documentation:

http://www.cisco.com/en/US/docs/general/whatsnew/whatsnew.html

Subscribe to the *What's New in Cisco Product Documentation* as an RSS feed and set content to be delivered directly to your desktop using a reader application. The RSS feeds are a free service. Cisco currently supports RSS Version 2.0.

CHAPTER 1

Virtual Port Channel Operations

This chapter describes the best practices and operational procedures for the virtual port channel (vPC) feature on Cisco Nexus 5000 Series switches that run Cisco NX-OS Release 5.0(2)N2(1) and earlier releases.

This chapter includes the following sections:

Information About vPC Operations

A vPC allows links that are physically connected to two different Cisco Nexus 5000 Series switches to appear as a single port channel to a third switch. The third switch can be a Cisco Nexus 2000 Series Fabric Extender or a switch, server, or any other networking device. A vPC can provide Layer 2 multipath capability which allows you to create redundancy by increasing bandwidth, enabling multiple parallel paths between nodes, and load-balancing traffic where alternative paths exist.

For a quick overview of vPC configurations, see the *Virtual PortChannel Quick Configuration Guide* at the following URL:
http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/configuration_guide_c07-543563.html

vPC Consistency Checks

This section includes the following topics:

- Bypassing a vPC Consistency Check When a Peer Link is Lost, page 1-8

Type 1 and Type 2 Consistency Check Parameters

Before a Cisco Nexus 5000 Series switch brings up a vPC, the two Cisco Nexus 5000 Series switches in the same vPC domain exchange configuration information to verify if both switches have compatible configurations for a vPC topology. Depending on the severity of the impact of possible mismatched configurations, some configuration parameters are considered as Type 1 consistency check parameters while others are considered as Type 2.

When a mismatch in Type 1 parameters occur, the following applies:

- If a graceful consistency check is enabled (default), the primary switch keeps the vPC up while the secondary switch brings it down
- If a graceful consistency check is disabled, both peer switches suspend VLANs on the vPC ports.

Note The graceful consistency check is a new feature introduced in Cisco NX-OS Release 5.0(2)N2(1) and is enabled by default. For more details, see the "Graceful Consistency Check" section on page 1-3.

When Type 2 parameters exist, a configuration mismatch generates a warning syslog message. The vPC on the Cisco Nexus 5000 Series switch remains up and running. The global configuration, such as Spanning Tree Protocol (STP), and the interface-level configurations are included in the consistency check.

The **show vpc consistency-parameters global** command lists all global consistency check parameters. Beginning with Cisco NX-OS Release 5.0(2)N1(1), QoS parameters have been downgraded from Type 1 to Type 2.

This example shows how to display all global consistency check parameters:

```
switch# show vpc consistency-parameters global
Legend:
        Type 1 : vPC will be suspended in case of mismatch
Name                        Type  Local Value            Peer Value
-------------               ----  ---------------------- -----------------------
QoS                         2     ([], [3], [], [], [],  ([], [3], [], [], [],
                                  [])                    [])
Network QoS (MTU)           2     (1538, 2240, 0, 0, 0,  (1538, 2240, 0, 0, 0,
                                  0)                     0)
Network Qos (Pause)         2     (T, F, F, F, F, F)     (T, F, F, F, F, F)
Input Queuing (Bandwidth)   2     (50, 50, 0, 0, 0, 0)   (50, 50, 0, 0, 0, 0)
Input Queuing (Absolute     2     (F, F, F, F, F, F)     (F, F, F, F, F, F)
Priority)
Output Queuing (Bandwidth)  2     (50, 50, 0, 0, 0, 0)   (50, 50, 0, 0, 0, 0)
Output Queuing (Absolute    2     (F, F, F, F, F, F)     (F, F, F, F, F, F)
Priority)
STP Mode                    1     MST                    MST
STP Disabled                1     None                   None
STP MST Region Name         1     ""                     ""
STP MST Region Revision     1     0                      0
STP MST Region Instance to  1
 VLAN Mapping
STP Loopguard               1     Disabled               Disabled
STP Bridge Assurance        1     Enabled                Enabled
STP Port Type, Edge         1     Normal, Enabled,       Normal, Enabled,
BPDUFilter, Edge BPDUGuard        Disabled               Disabled
STP MST Simulate PVST       1     Enabled                Enabled
Allowed VLANs               -     1,10,100-101,200-201   1,10,100-101,200-201,2
```

```
                                                          000
Local suspended VLANs        -     -                      -
```

Use the **show vpc consistency-parameters interface port-channel** *number* command to display the interface-level consistency parameters.

This example shows how to display the interface-level consistency parameters:

```
n5k-1# show vpc consistency-parameters interface port-channel 200

    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value            Peer Value
-------------               ----  ---------------------- -----------------------
STP Port Type               1     Default                Default
STP Port Guard              1     None                   None
STP MST Simulate PVST       1     Default                Default
lag-id                      1     [(7f9b,                [(7f9b,
                                  0-23-4-ee-be-64, 80c8, 0-23-4-ee-be-64, 80c8,
                                   0, 0), (8000,          0, 0), (8000,
                                  0-1e-13-15-7-40, 1, 0, 0-1e-13-15-7-40, 1, 0,
                                   0)]                    0)]
mode                        1     active                 active
Speed                       1     10 Gb/s                10 Gb/s
Duplex                      1     full                   full
Port Mode                   1     trunk                  trunk
Native Vlan                 1     1                      1
Shut Lan                    1     No                     No
Allowed VLANs               -     1-999,1001-3967,4048-4 1-3967,4048-4093
                                  093
```

The Cisco Nexus 5000 Series switch conducts vPC consistency checks when it attempts to bring up a vPC or when you make a configuration change.

In the interface consistency parameters shown in the above output, all configurations except the Allowed VLANs are considered as Type 1 consistency check parameters. The Allowed VLAN (under the trunk interface) is considered as a Type 2 consistency check parameter. If the Allowed VLAN ranges are different on both VLANs that means that only common VLANs are active and trunked for the vPC while the remaining VLANs are suspended for this port channel.

Graceful Consistency Check

Beginning with Cisco NX-OS Release 5.0(2)N2(1) and later releases, when a Type 1 mismatch occurs, by default, the primary vPC links are not suspended. Instead, the vPC remains up on the primary switch and the Cisco Nexus 5000 Series switch performs Type 1 configurations without completely disrupting the traffic flow. The secondary switch brings down its vPC until the inconsistency is cleared.

However, in Cisco NX-OS Release 5.0(2)N2(1) and earlier releases, this feature is not enabled for dual-homed FEX ports. When Type-1 mismatches occur in this topology, the VLANs are suspended on both switches. The traffic is disrupted on these ports for the duration of the inconsistency.

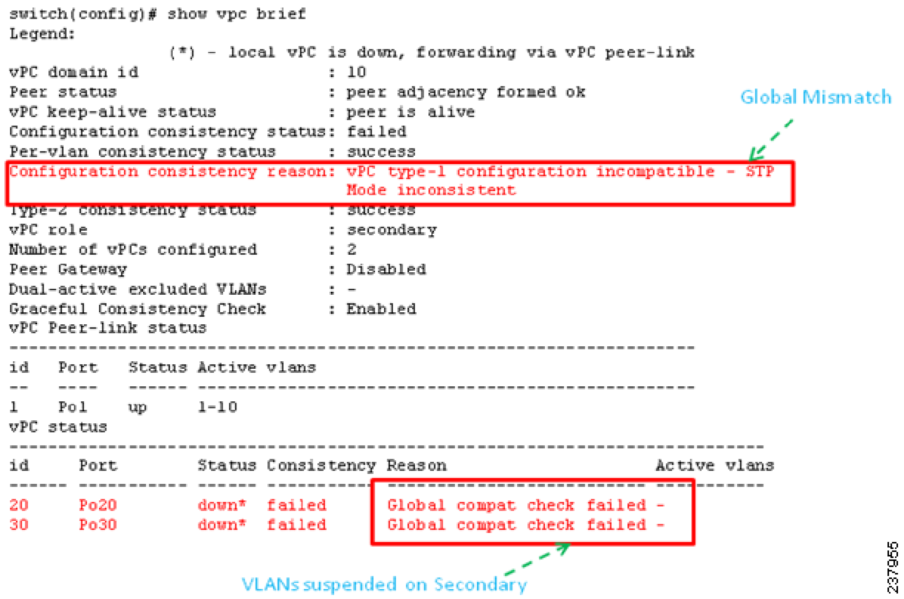
To minimize disruption, we recommend that you use the configuration synchronization feature for making configuration changes on these ports.

To enable a graceful consistency check, use the **graceful consistency-check** command. Use the **no** form of this command to disable the feature. The graceful consistency check feature is enabled by default.

This example shows how to enable a graceful consistency check:

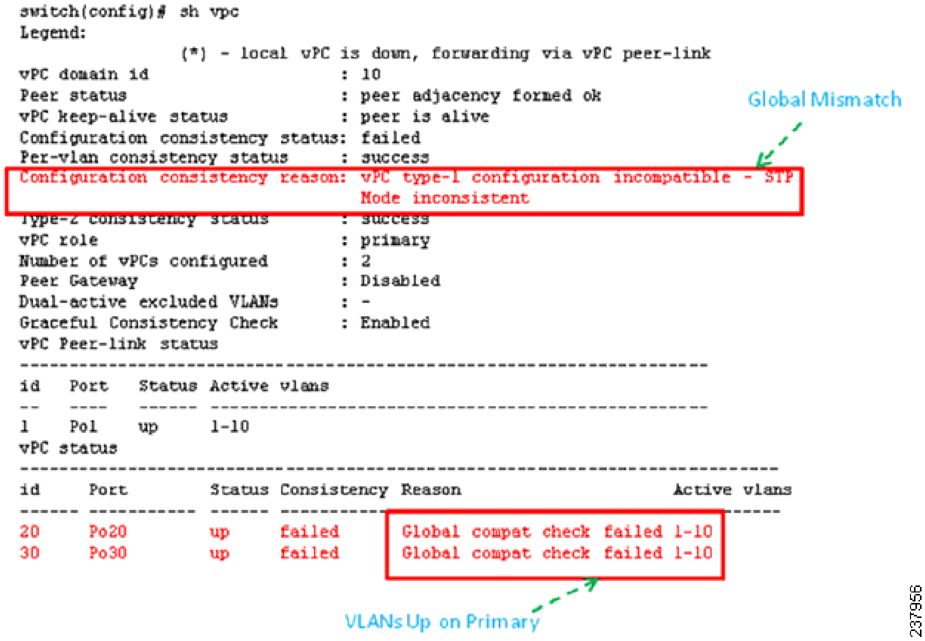
```
switch(config)# vpc domain 10
switch(config-vpc-domain)# [no] graceful consistency-check
```

This example shows that the vPC ports are down on a secondary switch when an STP mode mismatch occurs:



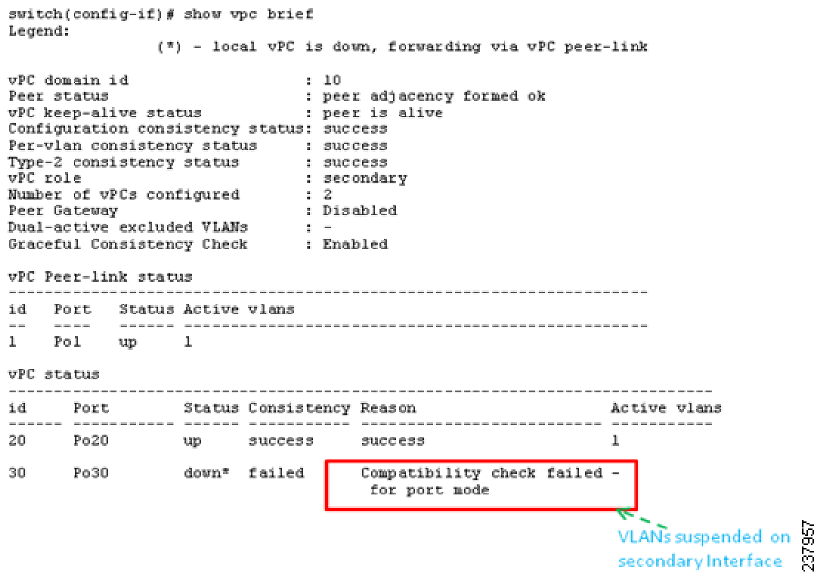
```
switch(config)# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link
vPC domain id                   : 10
Peer status                     : peer adjacency formed ok
vPC keep-alive status           : peer is alive
Configuration consistency status: failed
Per-vlan consistency status     : success
Configuration consistency reason: vPC type-1 configuration incompatible - STP
                                  Mode inconsistent
Type-2 consistency status       : success
vPC role                        : secondary
Number of vPCs configured       : 2
Peer Gateway                    : Disabled
Dual-active excluded VLANs      : -
Graceful Consistency Check      : Enabled
vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ --------------------------------------------------
1    Po1    up     1-10
vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- ------------
20     Po20        down*  failed      Global compat check failed -
30     Po30        down*  failed      Global compat check failed -
```

This example shows that the vPC ports and the VLANs remain up on the primary switch when an STP mode mismatch occurs:



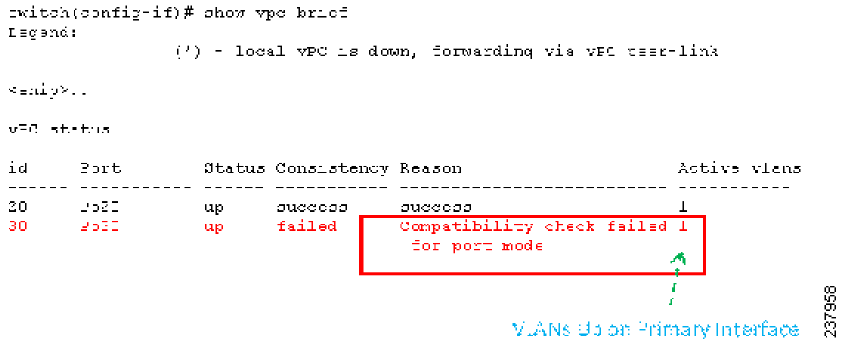
```
switch(config)# sh vpc
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link
vPC domain id                   : 10
Peer status                     : peer adjacency formed ok
vPC keep-alive status           : peer is alive
Configuration consistency status: failed
Per-vlan consistency status     : success
Configuration consistency reason: vPC type-1 configuration incompatible - STP
                                  Mode inconsistent
Type-2 consistency status       : success
vPC role                        : primary
Number of vPCs configured       : 2
Peer Gateway                    : Disabled
Dual-active excluded VLANs      : -
Graceful Consistency Check      : Enabled
vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ --------------------------------------------------
1    Po1    up     1-10
vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- ------------
20     Po20        up     failed      Global compat check failed 1-10
30     Po30        up     failed      Global compat check failed 1-10
```

This example shows that the vPC ports are down on a secondary switch when an interface-level Type 1 inconsistency occurs:



```
switch(config-if)# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

vPC domain id                   : 10
Peer status                     : peer adjacency formed ok
vPC keep-alive status           : peer is alive
Configuration consistency status: success
Per-vlan consistency status     : success
Type-2 consistency status       : success
vPC role                        : secondary
Number of vPCs configured       : 2
Peer Gateway                    : Disabled
Dual-active excluded VLANs      : -
Graceful Consistency Check      : Enabled

vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ --------------------------------------------------
1    Po1    up     1

vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- -----------
20     Po20        up     success     success                    1

30     Po30        down*  failed      Compatibility check failed -
                                       for port mode
```

This example shows that the vPC ports and the VLANs remain up on the primary switch when an interface-level Type 1 inconsistency occurs:



```
switch(config-if)# show vpc brief
Legend:
                 (*) - local vPC is down, forwarding via vPC peer-link

<snip>..

vPC status

id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- -----------
20     Po20        up     success     success                    1
30     Po30        up     failed      Compatibility check failed 1
                                       for port mode
```

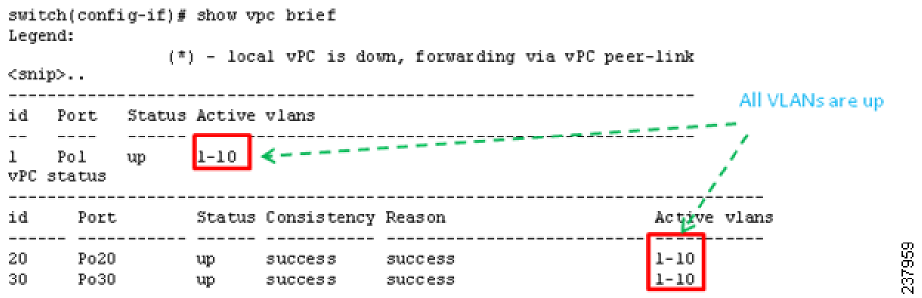
Configuring Per-VLAN Consistency Checks

Beginning with Cisco NX-OS Release 5.0(2)N2(1), the Cisco Nexus 5000 Series switch performs Type-1 consistency checks on a per-VLAN basis when you enable or disable STP on a VLAN. VLANs that do not pass this consistency check are brought down on the primary and secondary switches while other VLANs are not affected.

When you enter the **no spanning-tree vlan** *number* command on one peer switch, only the specified VLAN is suspended on both peer switches; the other VLANs remain up.

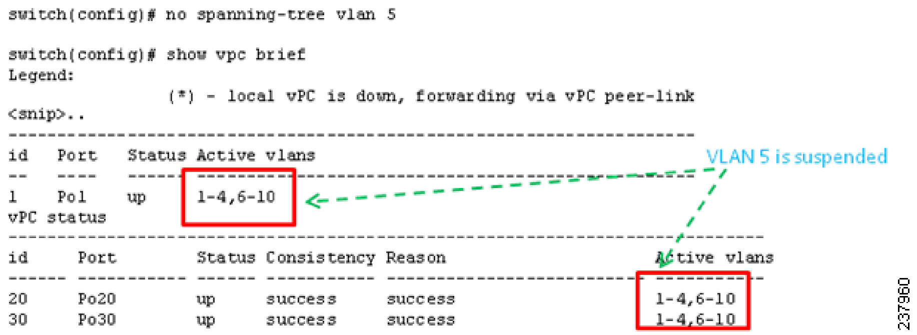
Note Per-VLAN consistency checks are not dependent on whether graceful consistency checks are enabled.

This example shows the active VLANs before suspending a specified VLAN:



```
switch(config-if)# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link
<snip>..
----------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ ---------------------------------------------------
1    Po1    up     1-10
vPC status
-----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- ------------
20     Po20        up     success     success                    1-10
30     Po30        up     success     success                    1-10
```

This example shows that VLAN 5 is suspended but the remaining VLANs are up:

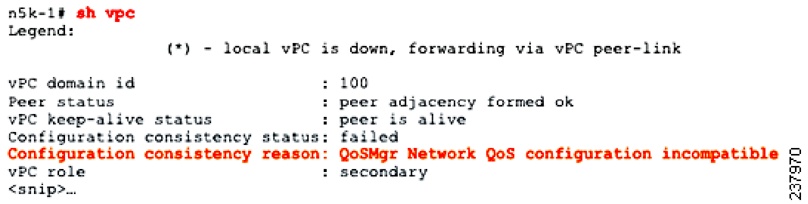


```
switch(config)# no spanning-tree vlan 5

switch(config)# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link
<snip>..
----------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ ---------------------------------------------------
1    Po1    up     1-4,6-10
vPC status
-----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- ------------
20     Po20        up     success     success                    1-4,6-10
30     Po30        up     success     success                    1-4,6-10
```

Identifying Inconsistent vPC Configurations

The **show vpc** command displays the vPC status and the vPC consistency check result for the global consistency check and the interface-specific consistency check.

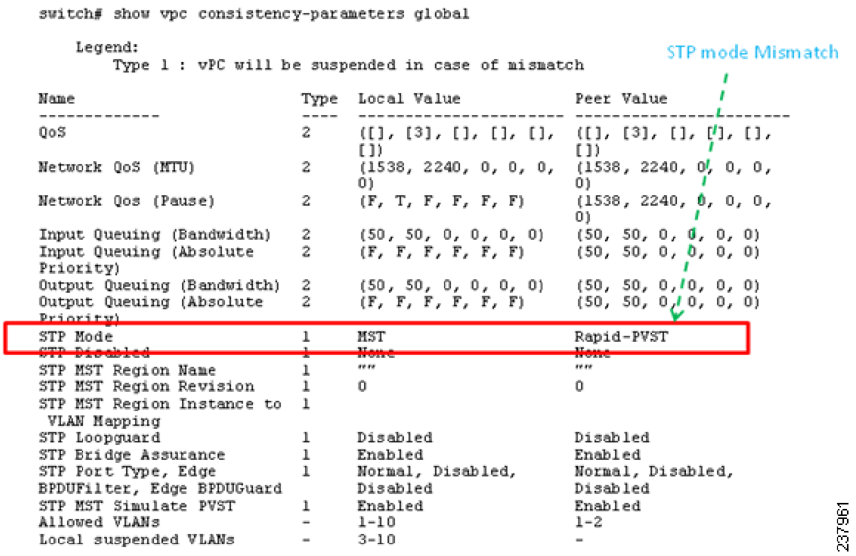
This example shows the global vPC consistency check failed because of the mismatched Network QoS configuration:



```
n5k-1# sh vpc
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

vPC domain id                    : 100
Peer status                      : peer adjacency formed ok
vPC keep-alive status            : peer is alive
Configuration consistency status: failed
Configuration consistency reason: QoSMgr Network QoS configuration incompatible
vPC role                         : secondary
<snip>…
```

You can use the **show vpc consistency-parameters global** command to identify the configuration difference between two vPC peer switches.

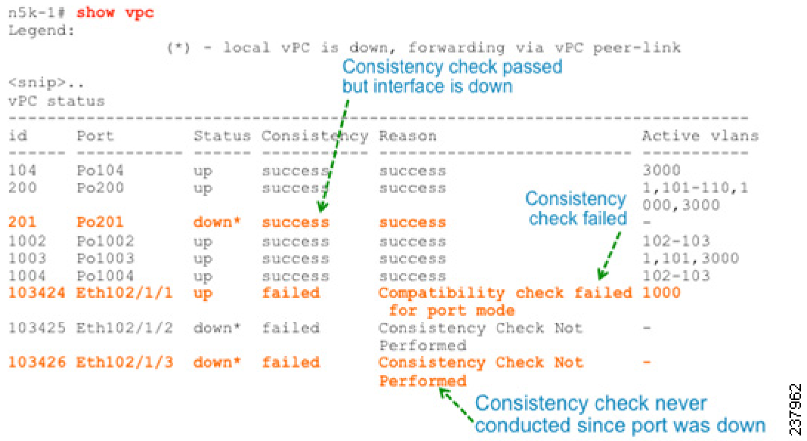
This example shows the global consistency check failed because the STP mode was configured differently on the two vPC switches:



```
switch# show vpc consistency-parameters global

    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                         Type  Local Value              Peer Value
-------------                ----  ----------------------   -----------------------
QoS                          2     ([], [3], [], [], [],    ([], [3], [], [], [],
                                   [])                      [])
Network QoS (MTU)            2     (1538, 2240, 0, 0, 0,    (1538, 2240, 0, 0, 0,
                                   0)                       0)
Network Qos (Pause)          2     (F, T, F, F, F, F)       (1538, 2240, 0, 0, 0,
                                                            0)
Input Queuing (Bandwidth)    2     (50, 50, 0, 0, 0, 0)     (50, 50, 0, 0, 0, 0)
Input Queuing (Absolute      2     (F, F, F, F, F, F)       (50, 50, 0, 0, 0, 0)
Priority)
Output Queuing (Bandwidth)   2     (50, 50, 0, 0, 0, 0)     (50, 50, 0, 0, 0, 0)
Output Queuing (Absolute     2     (F, F, F, F, F, F)       (50, 50, 0, 0, 0, 0)
Priority)
STP Mode                     1     MST                      Rapid-PVST
STP Disabled                 1     None                     None
STP MST Region Name          1     ""                       ""
STP MST Region Revision      1     0                        0
STP MST Region Instance to   1
 VLAN Mapping
STP Loopguard                1     Disabled                 Disabled
STP Bridge Assurance         1     Enabled                  Enabled
STP Port Type, Edge          1     Normal, Disabled,        Normal, Disabled,
BPDUFilter, Edge BPDUGuard         Disabled                 Disabled
STP MST Simulate PVST        1     Enabled                  Enabled
Allowed VLANs                -     1-10                     1-2
Local suspended VLANs        -     3-10                     -
```

You can use the **show vpc** command also shows the vPC consistency check result for each vPC and the reason for the consistency check failure.

This example shows how to display the vPC consistency check status:



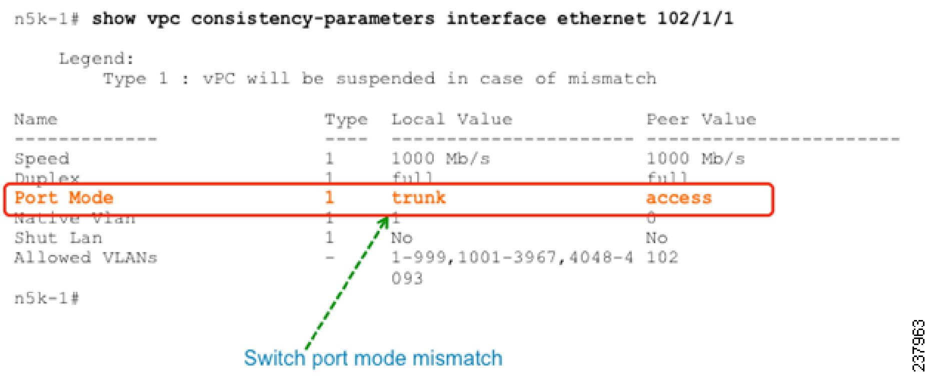
```
n5k-1# show vpc
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

<snip>..
vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                      Active vlans
------ ----------- ------ ----------- --------------------------- -----------
104    Po104       up     success     success                     3000
200    Po200       up     success     success                     1,101-110,1
                                                                  000,3000
201    Po201       down*  success     success                     -
1002   Po1002      up     success     success                     102-103
1003   Po1003      up     success     success                     1,101,3000
1004   Po1004      up     success     success                     102-103
103424 Eth102/1/1  up     failed      Compatibility check failed 1000
                                       for port mode
103425 Eth102/1/2  down*  failed      Consistency Check Not       -
                                      Performed
103426 Eth102/1/3  down*  failed      Consistency Check Not       -
                                      Performed
```

If the consistency check fails, the consistency check is not performed on vPC member ports that are down.

If the consistency check has succeeded and the port is brought down, the consistency check shows that it was successful.

You can use the **show vpc consistency-parameters interface ethernet** *slot/port* command to identify the configuration difference that leads to a consistency check failure for a specified interface or port channel.

This example shows how to display configuration differences that lead to consistency check failures.



```
n5k-1# show vpc consistency-parameters interface ethernet 102/1/1

    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value             Peer Value
-------------               ----  ----------------------  -----------------------
Speed                       1     1000 Mb/s               1000 Mb/s
Duplex                      1     full                    full
Port Mode                   1     trunk                   access
Native Vlan                 1     1                       0
Shut Lan                    1     No                      No
Allowed VLANs               -     1-999,1001-3967,4048-4  102
                                  093
n5k-1#
```

Switch port mode mismatch

Bypassing a vPC Consistency Check When a Peer Link is Lost

The vPC consistency check message is sent by the vPC peer link. The vPC consistency check cannot be performed when the peer link is lost. When the vPC peer link is lost, the operational secondary switch suspends all of its vPC member ports while the vPC member ports remain on the operational primary switch. If the vPC member ports on the primary switch flaps afterwards (for example, when the switch or server that connects to the vPC primary switch is reloaded), the ports remain down due to the vPC consistency check and you cannot add or bring up more vPCs.

Beginning with Cisco NX-OS Release 5.0(2)N2(1), the auto-recovery feature brings up the vPC links when one peer is down. This feature performs two operations:

- If both switches reload, and only one switch boots up, auto-recovery allows that switch to assume the role of the primary switch. The vPC links come up after a configurable period of time if the vPC peer-link and the peer-keepalive fail to become operational within that time. If the peer-link comes up but the peer-keepalive does not come up, both peer switches keep the vPC links down. This feature is similar to the reload restore feature in Cisco NX-OS Release 5.0(2)N1(1) and earlier releases. The reload delay period can range from 240 to 3600 seconds.
- When you disable vPCs on a secondary vPC switch because of a peer-link failure and then the primary vPC switch fails, the secondary switch reenables the vPCs. In this scenario, the vPC waits for three consecutive keepalive failures before recovering the vPC links.

Note The auto-recovery feature in Cisco NX-OS Release 5.0(2)N2(1) and later releases replaces the reload restore feature in Cisco NX-OS Release 5.0(2)N1(1) and earlier releases.

The auto-recovery feature is disabled by default. To enable auto-recovery, enter the **auto-recovery** command in the vPC domain mode.

This example shows how to enable the auto-recovery feature and to set the reload delay period:

```
switch(config)# vpc domain 10
switch(config-vpc-domain)# auto-recovery ?
  <CR>
  reload-delay  Duration to wait after reload to recover vPCs

switch(config-vpc-domain)# auto-recovery reload-delay ?
  <240-3600>  Time-out for restoring vPC links (in seconds)
switch(config-vpc-domain)# auto-recovery reload-delay 240
Warning:
 Enables restoring of vPCs in a peer-detached state after reload, will wait for 240
seconds (by default) to determine if peer is un-reachable
```

This example shows how to display the status of the auto-recovery feature:

```
switch(config-vpc-domain)# show running-config vpc
!Command: show running-config vpc
!Time: Tue Dec  7 02:38:44 2010

version 5.0(2)N2(1)
feature vpc
vpc domain 10
  peer-keepalive destination 10.193.51.170
  auto-recovery
```

Configuring Changes in vPC Topologies

One of the challenges with vPC topologies is how to make configuration changes with minimum traffic disruption. Due to the consistency check, the configuration made on one vPC switch could potentially lead to consistency check failure and traffic disruption.

Beginning with Cisco NX-OS Release 5.0(2)N2(1), you can use the following procedure to make configuration changes for Type 1 consistency check parameters on a Cisco Nexus 5000 Series switch. We recommend that you perform the following procedure during a maintenance window because it might reduce the vPC bandwidth by half for a short duration.

Note A graceful consistency-check does not apply to dual-homed FEX ports. As a result, both switches keep the port down for the duration of an inconsistency. Using the configuration synchronization feature reduces the duration of the inconsistency.

To make configuration changes for Type 1 consistency-check parameters, follow these steps:

Step 1 Enable graceful consistency-check in a vPC domain.

```
switch# config term
switch(config)# vpc domain 10
switch(config-vpc-domain)# graceful consistency-check
```

Step 2 Enable the configuration synchronization feature on both vPC peer switches.

For details on using the configuration synchronization feature, see the “Configuration Synchronization Operations” chapter.

Step 3 Perform all configuration changes in the switch profile.

```
switch# config sync
switch(config-sync)# switch-profile abc
switch(config-sync-sp)# interface Port-channel 100
switch(config-sync-sp-if)# switchport mode trunk
switch(config-sync-sp-if)# commit
```

When you commit switch profile configurations on the local switch, the configuration is also sent to the vPC peer switch to reduce misconfigurations when changes are made on only one vPC switch and to reduce the downtime because the configuration is applied rapidly. When there is a short mismatch duration, a graceful consistency-check keeps the primary side forwarding traffic.

Note When you are making a configuration change for a Type 2 consistency check parameter, such as Allowed VLAN for trunk ports, you do not need to follow this procedure.

Replacing a Cisco Nexus 5000 Series Switch or Cisco Nexus 2000 Fabric Extender

This section describes how to replace a Cisco Nexus 5000 Series switch or Cisco Nexus 2000 Series Fabric Extender in a vPC topology with minimal disruption.

This section include the following topics:

- Replacing a Cisco Nexus 5000 Series Switch, page 1-11
- Replacing a Cisco Nexus 2000 Series Fabric Extender, page 1-13

Replacing a Cisco Nexus 5000 Series Switch

When you replace a Cisco Nexus 5000 Series switch, you must perform the following procedure on the replacement switch to synchronize the configuration with the existing Cisco Nexus 5000 Series switch. The procedure can be done in a hybrid single/dual-homed Fabric Extender vPC topology.

Note Do not connect a peer-link, vPC, or single/dual homed Fabric Extender topology fabric port to the replacement switch.

Note For a vPC+ topology, ensure that you wait for twenty minutes before you replace a vPC+ switch. Otherwise, vPC legs in the primary switch will get suspended due to switch-id conflict.

Before You Begin

- Power up replacement switch with no cables other than mgmt0 and console cable connected to the switch.
- Copy the required Cisco NX-OS kickstart/system files into the switch bootflash.
- If you have a backup of the switch configuration, copy it to new switch bootflash.
- Enable the FEX pre-provisioning feature on the switch in the vPC topology.
- Enable the configuration synchronization feature on the switch and apply all the switch profile configurations except for the sync peer destination IP address.

To replace a Cisco Nexus 5000 Series switch in a vPC topology, follow these steps:

Step 1 Boot the replacement switch.

The new switch comes up without a configuration. Ensure the software version is upgraded to match the existing switch.

Step 2 Enable FEX pre-provisioning for all single or dual homed Fabric Extender modules on the replacement switch.

Note Ensure that you unconfigure the **system default switchport shutdown** command on the replacement switch. Otherwise, when Fabric Extender Modules are coming online on the replacement switch, dual-homed FEX ports on the primary switch will flap causing traffic disruption.

Step 3 Configure the replacement switch as follows:

Note Before you configure the replacement switch using any of the following method, disable the vPC auto-recovery feature on both the vPC peers using the **no auto-recovery** command under the vPC domain. This is to ensure that there is no vPC role change because of the sticky bit feature, when the replacement switch is brought up. vPC auto-recovery feature is enabled by default in Cisco NX-OS release 7.x and later.

- If the running configuration was saved offline, go to Step 4 to Step 10 to apply the configuration.

- If the running configuration was not saved offline, you can obtain it from the peer switch if the configuration synchronization feature is enabled. (Create a switch profile and then go to Step 11).
- If neither condition is met, manually add the configuration and then go to Step 11.

Step 4 Edit the configuration file to remove the sync-peer command if using the configuration synchronization feature.

Step 5 Configure the mgmt0 port IP address and download the configuration file.

Step 6 Copy the saved configuration file to the running configuration.

Step 7 Edit the saved configuration file and delete all commands between the **configure sync** command and the **commit** command, including these two commands.

Step 8 Copy the new, edited configuration file to the running configuration again.

Step 9 Verify that the configuration is correct by entering the **show running-config** command and the **show provision failed-config** *slot* command.

Step 10 If switch profile configuration changes were made on the peer switch while the replacement switch was out of service, apply those configurations in the switch profile and then enter the **commit** command.

Step 11 Shut down all single-homed Fabric Extender vPC host ports.

Step 12 Connect the single-homed Fabric Extender topology fabric ports.

Step 13 Wait for single-homed Fabric Extenders to come online.

Step 14 Ensure the vPC role priority of the existing switch is better than the replacement switch.

Step 15 Connect the vPC peer keepalive link to the peer switch. Ensure that vPC peer keepalive link is operational by entering the **show vpc** command.

Warning **If the auto-recovery feature was not disabled in Step 3, ensure that either the vPC peer-keepalive or the vPC peer-link comes up before the auto-recovery timer expires (default 240 seconds). If this does not happen, the replacement switch will assume the vPC primary role (dual active). If the vPC peer-link is restored in this state, there will be a vPC role change causing vPCs on the peer switch to go down as the switch transitions to vPC secondary role. If required, with just peer-keepalive link operational, reload the replacement switch one more time with all the other interfaces still in the shutdown state.**

Step 16 Ensure that the vPC role field is **none established**. Use the **show vpc** or **show vpc role** command to view the vPC role. If the vPC role field displays **Primary**, then do not proceed with the replacement procedure. Reload the switch to get the vPC role field to **none established**.

Step 17 Connect the vPC peer-link ports to the peer switch. Ensure that the vPC peer link is operational by entering the **show vpc** command.

Step 18 Connect the dual-homed Fabric Extender topology fabric ports.

Step 19 Connect the switch vPC ports.

Step 20 Enter the **no shutdown** command on all single-homed Fabric Extender vPC ports.

Step 21 Verify that the replaced vPC switch and the Fabric Extenders on the replacement switch are online and there is no traffic disruption.

Step 22 If you are using the configuration synchronization feature, add the sync-peer configuration to the switch profile if this wasn't enabled in Step 3.

Step 23 If you are using the configuration synchronization feature, enter the **show switch-profile** *name* **status** command to ensure both switches are synchronized.

Step 24 If vPC auto recovery was disabled, enable auto recovery using the **auto-recovery** command under vPC domain on both switches.

Replacing a Cisco Nexus 2000 Series Fabric Extender

This section describes how to replace a Cisco Nexus 2000 Series Fabric Extender with minimal disruption. This section includes the following topics:

- Replacing a Fabric Extender in a Dual-Homed Fabric Extender vPC Topology, page 1-13
- Replacing a Fabric Extender in a Single-Homed Fabric Extender vPC Topology, page 1-13
- Installing a New Cisco Nexus 2000 Series Fabric Extender, page 1-14

Replacing a Fabric Extender in a Dual-Homed Fabric Extender vPC Topology

Because the hosts behind a Fabric Extender in a dual-homed Fabric Extender vPC topology are by definition singly-connected, traffic disruption will occur for those hosts.

If the replacement Fabric Extender is a different model, the Cisco Nexus 5000 Series switch does not allow you to pre-provision a new type until you disconnect the old Fabric Extender.

To retain the configuration on both Cisco Nexus 5000 Series peer switches in the vPC topology, follow these steps.

Step 1 Save the configuration for the Fabric Extender interfaces to a file.

Step 2 Disconnect the Fabric Extender fabric ports and wait until the Fabric Extender is offline.

Step 3 Pre-provision the slot with the new Fabric Extender model.

Step 4 Modify the configuration file if necessary for the new Fabric Extender if the configurations are incompatible.

Note For vPC ports, this step might affect consistency.

Step 5 Copy the file to the running configuration.

Step 6 Connect the Fabric Extender fabric and host ports and then wait for the Fabric Extender to come online.

Step 7 Verify that all ports are up with the correct configuration.

Replacing a Fabric Extender in a Single-Homed Fabric Extender vPC Topology

If the replacement Fabric Extender is the same model as the original Fabric Extender, then there is no disruption; the configuration on the Fabric Extender interfaces remain unchanged.

If the replacement Fabric Extender is a different model, the Cisco Nexus 5000 Series switch does not allow you to pre-provision a new type until you disconnect the old Fabric Extender.

To replace a Fabric Extender in a single homed Fabric Extender vPC topology, follow the procedure described in "Replacing a Fabric Extender in a Dual-Homed Fabric Extender vPC Topology" section on page 1-13.

Installing a New Cisco Nexus 2000 Series Fabric Extender

With pre-provisioning, you can fully configure the new Fabric Extender before the Fabric Extender is connected to a Cisco Nexus 5000 Series switch.

To install a new Cisco Nexus 2000 Series Fabric Extender, follow these steps:

Step 1 Pre-provision the slot with the Fabric Extender model.

Step 2 Configure the interfaces as though the Fabric Extender is connected.

Step 3 Connect the Fabric Extender and wait for it to come online.

Step 4 Verify that all configurations are applied correctly

Note The switch applies all configurations serially in a best-effort fashion when the Fabric Extender comes online.

vPC Failure Recovery

This section describes different vPC failure scenarios and how to recover from them. This section includes the following topics:

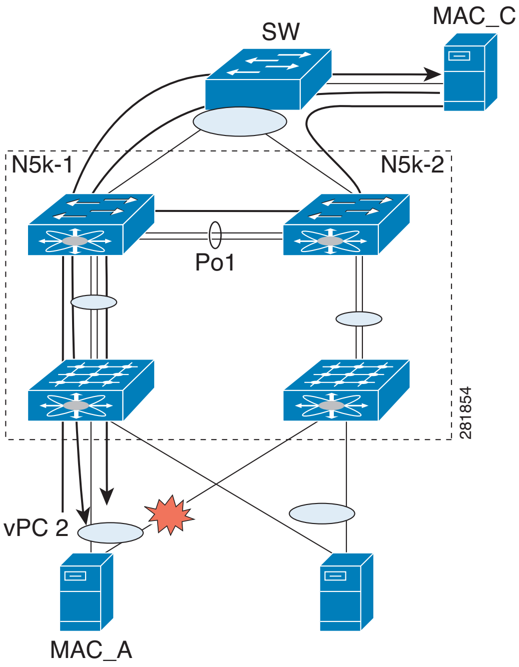
- vPC Member Port Failure, page 1-14
- vPC Peer Link Failure, page 1-15
- vPC Peer Keepalive Link Failure, page 1-16
- vPC Peer Switch Failure, page 1-17
- vPC Peer Link Failure Followed by a Peer Keepalive Link Failure, page 1-17
- vPC Keepalive Link Failure Followed by a Peer Link Failure, page 1-17

vPC Member Port Failure

Figure 1-1 shows the traffic flow when one vPC member port fails. Once the host MAC_A detects a link failure on one of the port-channel members, it redistributes the affected flows to the remaining port channel members. The return flow from MAC_C to MAC_A could take the path of the left- or the right-side Cisco Nexus 5000 Series switch, depending on the port-channel hash algorithm of the top switch. For those flows that traverse the right-side Cisco Nexus 5000 Series switch (the red line), the Cisco Nexus 5000 Series switch passes the traffic to the left-side Cisco Nexus 5000 Series switch, because it no longer has the local connection to host MAC_A. This is one of the scenarios where a vPC peer link is used to carry data traffic.

We recommend that you provision enough bandwidth for peer links to accommodate the bandwidth needed for link failure scenarios.

Figure 1-1 vPC Response to a Member Port Failure

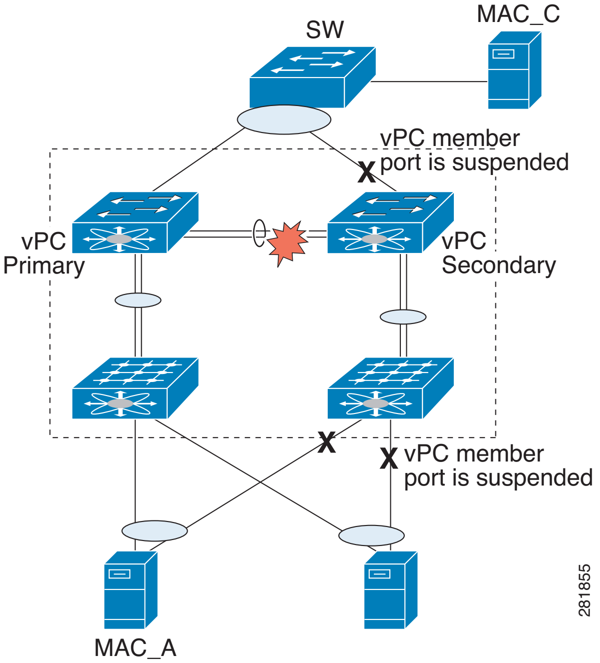


vPC Peer Link Failure

Figure 1-2 shows the vPC response to a peer link failure. In a vPC topology, one vPC peer switch is elected as the vPC primary switch and the other switch is elected as the vPC secondary switch, based on the configured role priority for the switch. In the unlikely scenario where the vPC peer link goes down, the vPC secondary switch shuts down all of its vPC member ports if it can still receive keepalive messages from the vPC primary switch (which indicates that the vPC primary switch is still alive). The vPC primary switch keeps all of its interfaces up. As a result, the hosts or switches that are connected to the Cisco Nexus 5000 Series switch or Cisco Nexus 2000 Series Fabric Extender vPC pair redistributes all the flows to the vPC member ports that are connected to the vPC primary switch.

As a best practice, we recommend that you configure a physical port channel that has at least two 10 Gigabit-Ethernet ports as the vPC peer link.

Figure 1-2 vPC Response to a Peer Link Failure



A vPC consistency check cannot be done when a vPC peer-link is down either due to a link failure or when the peer switch is completely down. In either case, any newly configured vPC does not come up because the vPC consistency check cannot proceed, or the existing vPC remains disabled after the link flaps.

Use the reload restore feature that was introduced in Cisco NX-OS Release 5.0(2)N1(1) to fix this problem. The reload restore feature allows a switch to bypass the vPC consistency check and bring up vPC ports when the peer-link or peer switch fails. The reload restore feature has been replaced with the auto-recovery feature in Cisco NX-OS Release 5.0(2)N2(1).

vPC Peer Keepalive Link Failure

The vPC keepalive link carries the heartbeat message between two vPC peer switches. The failure of the vPC keepalive link alone does not impact the vPC operation or data forwarding. Although it has no impact on data forwarding, we recommend that you fix the keepalive as soon as possible to avoid a double failure scenario that could impact the data traffic.

When both switches come up together (such as after power gets restored following a power outage) and only the mgmt/keepalive link fails, the peers are unreachable. However, all other links, including vPC peer links, are up. In this scenario, reaching the vpc-peers through keepalives are achieved through keepalive links while the primary and secondary role election is established through the vpc-peer link. You must establish the first keepalive for the role election to occur in the case when a switch comes up and the vPC-peer link is up.

When keepalives fail to reach the peer switches, role election does not proceed and the primary or secondary role is not established on either vPC peer switch and all vPC interfaces are kept down on both switches.

Note If this scenario occurs again or if the keepalive link goes down after vPC peers are established, the roles do not change and all vPCs remain up.

vPC Peer Switch Failure

When one peer switch fails, half of the network bandwidth is lost and the remaining vPC switch maintains the network connectivity. If the failure occurs on a primary switch, the secondary switch becomes the primary switch.

When one peer switch fails, the remaining peer switch maintains network connectivity for the vPC until it is reloaded. This situation could happen if both vPC peer switches are reloaded and only one switch comes up or both switches loose power and then the power is restored only on one switch. In either case, since the vPC primary election cannot proceed, the Cisco Nexus 5000 Series switch keeps the vPC ports in suspend mode.

To fix these problems, use the reload restore feature and the auto recovery feature as follows:

In NX-OS Release 5.0(2)N1(1), enter the **reload restore** command:

```
switch(config-vpc-domain)# reload restore <timeout in second>
```

In NX-OS Release 5.0(2)N2(1), enter the **auto-recovery reload-delay** command:

```
switch(config-vpc-domain)# auto-recovery reload-delay ?
  <240-3600>  Time-out for restoring vPC links (in seconds)
```

These commands allow the vPC peer switch to bypass the vPC consistency check and bring up vPC ports after the delay timer expires.

vPC Peer Link Failure Followed by a Peer Keepalive Link Failure

If a peer link failure occurs, the vPC secondary switch checks if the primary switch is alive. The secondary switch suspends its vPC member ports after it confirms that the primary switch is up.

If the vPC primary switch goes down, the vPC secondary switch stops receiving Keepalive messages on the vPC Peer Keepalive link. After three consecutive Keepalive message timeouts, the vPC secondary switch changes its role to be the vPC primary switch and brings up its vPC member ports.

In Cisco NX-OS Release 5.0(2)N2(1), if you enable the auto-recovery feature and if the vPC primary switch goes down, the vPC secondary switch does not receive messages on the vPC peer keepalive link. Then, after three consecutive keepalive timeouts, the vPC secondary switch changes its role to primary and brings up the vPC member ports.

vPC Keepalive Link Failure Followed by a Peer Link Failure

If the vPC keepalive link fails first and then a peer link fails, the vPC secondary switch assumes the primary switch role and keeps its vPC member ports up.

If the peer link and keepalive link fails, there could be a chance that both vPC switches are healthy and the failure occurs because of a connectivity issue between the switches. In this situation, both vPC switches claim the primary switch role and keep the vPC member ports up. This situation is known as a split-brain scenario. Because the peer link is no longer available, the two vPC switches cannot synchronize the unicast MAC address and the IGMP group and therefore they cannot maintain the complete unicast and multicast forwarding table. This situation is rare.

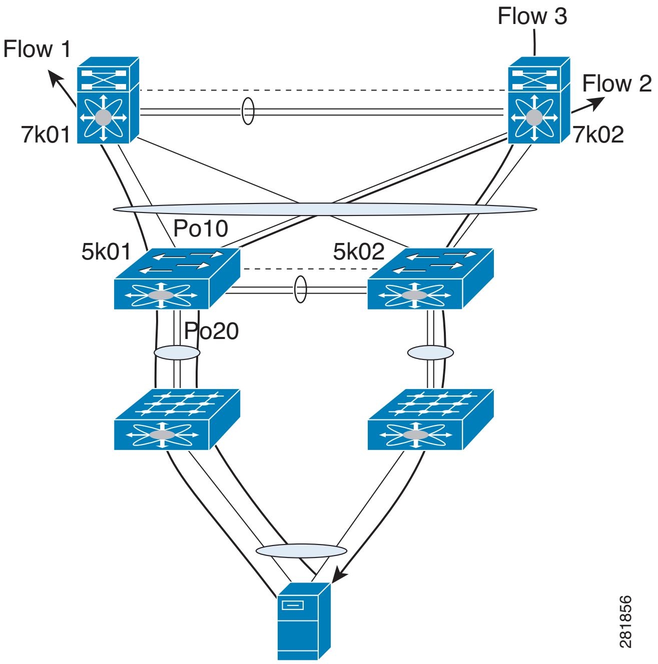
We recommend that you have a well-planned network design that includes spreading peer links and keepalive links to multiple ASICs or multiple modules and different cabling routes for keepalive and peer links to avoid a double failure.

Tracing Traffic Flow in a vPC Topology

This section describes how to trace a traffic flow in a vPC topology that is similar to a port-channel environment.

Figure 1-3 shows that each hop in the network chooses one vPC member port to carry the traffic flow independently.

Figure 1-3 Traffic Flow in a vPC Topology



In this example, for flow 1, the host makes a decision whether the traffic flow is sent to the FEX on left or the right side. The FEX runs its hash algorithm to choose one uplink to carry the flow. The N5k determines if the flow should be sent to N7k1 or N7k2. When the egress port for a traffic flow is a vPC, the vPC switch always prefers to use its own vPC member port to carry the traffic in order to minimize the utilization of peer links.

The Cisco NX-OS and Cisco IOS software includes commands to identify the port channel member that carries a particular flow.

This example assumes that the default hash algorithm is used which is src-mac, dst-mac, src-ip and dst-ip. If the hash algorithm also includes the Layer 4 UDP/TCP port, the port information also needs to be provided in the command. The port channel in the command should be the egress port channel.

```
switch# show port-channel load-balance forwarding-path interface Po3 src-interface
ethernet 1/1 vlan 1 src-mac 0000.0000.1111 src-ip 1.1.1.1 dst-mac 001e.1324.4dc0 dst-ip
2.2.2.2
Missing params will be substituted by 0's.
Load-balance Algorithm on switch: source-dest-ip
crc8_hash: 14   Outgoing port id: Ethernet1/31
Param(s) used to calculate load-balance:
        dst-ip:   2.2.2.2
        src-ip:   1.1.1.1
        dst-mac:  001e.1324.4dc0
        src-mac:  0000.0000.1111
switch#
```

The commands do not show how flows are distributed on the FEX uplink from the FEX to the N5k.

While using the SPAN feature to monitor the traffic flow, the communications between two hosts can be split between two vPC switches. Therefore, you may need to enable SPAN on both vPC switches to obtain a complete trace.

CHAPTER 2

Cisco Nexus 5500 Platform Layer 3 and vPC Operations

This chapter describes virtual port channel (vPC) operations when Layer 3 routing features are enabled on the Cisco Nexus 5500 Platform.

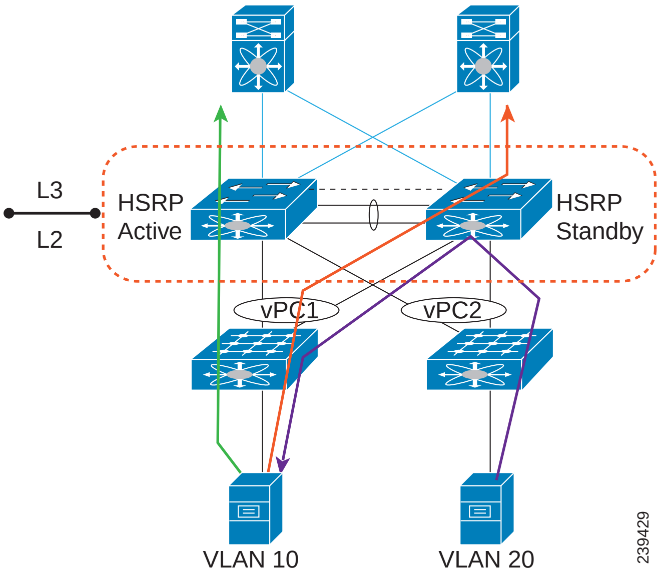
This chapter includes the following sections:

vPC and First Hop Redundancy Protocol

When you use a Cisco Nexus 5548 switch or Cisco Nexus 5596UP switch as a default gateway for hosts, you can deploy the First Hop Redundancy Protocol (FHRP) to provide default gateway redundancy. Beginning with Cisco NX-OS Release 5.0(3)N1(1b), an active FHRP peer and a standby peer can perform Layer 3 forwarding when you enable vPC. This optimization improves bandwidth, avoids sending the Layer 3 traffic over the vPC peer link, and requires no configuration or protocol change. Only the FHRP active peer answers ARP requests. Because both active and standby FHRP peers can forward Layer 3 traffic, you do not need to configure an aggressive timer for FHRP to provide faster failover and convergence time if an active FHRP peer fails.

Figure 2-1 shows that the Layer 3 traffic that originated from the host and is destined to a host several hops away can be routed by both the Host Standby Router Protocol (HSRP) active and the HSRP standby switch..

Figure 2-1 vPC and FHRP



ARP Processing with vPC

When the host connects to a Cisco Nexus 5500 Platform switch and Cisco Nexus 2000 Fabric Extenders in a vPC topology, the host can send an ARP request to the FHRP standby peer due to a hashing algorithm. The ARP request that is received by the standby peer is forwarded to the active peer and the active peer can answer it with an ARP reply.

Similarly, when traffic is moving from north to south, such as when one Cisco Nexus 5500 Platform switch sends an ARP request to a host, the ARP reply might be sent to another switch. In such a case, the ARP reply is forwarded as a Layer 2 frame to the Cisco Nexus 5500 Platform switch that originated the ARP request.

As of Cisco NX-OS Release 5.0(3)N1(1b), ARP synchronization does not occur between two Cisco Nexus 5500 Platform switches. The two switches resolve and maintain their ARP table independently. When one vPC peer switch is reloaded, the switch needs to resolve the ARP by sending ARP requests to the hosts.

Layer 3 Forwarding for Packets to a Peer Switch MAC Address

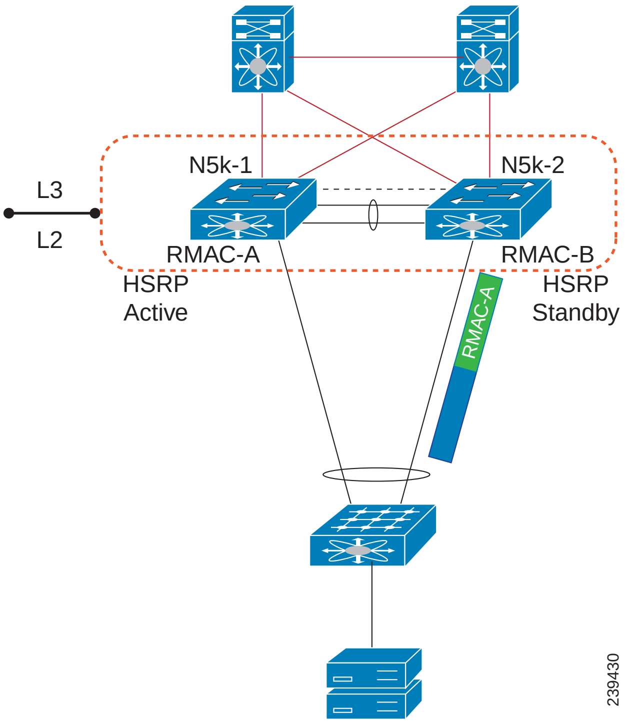
Typically, a router performs a Layer 3 route table lookup and Layer 3 forwarding when the destination MAC in the Ethernet frame matches its own MAC address. Otherwise, the packets are switched (if Layer 2 functionality is enabled) or dropped. In a topology with Layer 3 and vPC enabled, a vPC peer switch could receive IP packets with the peer's MAC address as the destination MAC rather than the virtual

MAC address (when FHRP is enabled) or its own MAC address. In this scenario, a Cisco Nexus 5500 Platform switch can forward the traffic to the peer using a peer link and the peer switch performs the Layer 3 forwarding.

The above scenario often happens with some filers or with Layer 3 peering over vPC. In the case of filers, they may achieves improved load balance and better performance by forwarding traffic to the Burnt-in-Address (BIA) of the routers instead of the HSRP MAC.

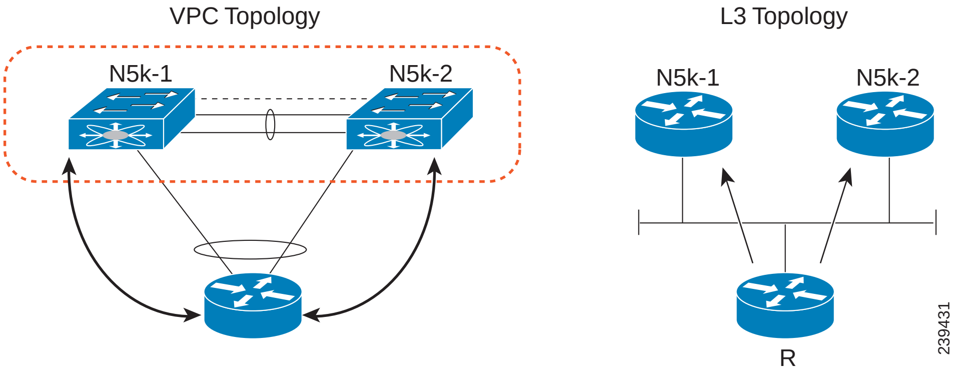
Figure 2-2 shows that when the NAS filer sends out packets with N5k-1's MAC RMAC-A as the destination MAC, the packets can be sent over to the N5k-2 switch due to the port channel hashing.

Figure 2-2 vPC and Peer-Gateway



Another scenario that could lead to this situation is when a router is connected to a Cisco Nexus 5500 Platform in a vPC topology.

Figure 2-3 Connecting to a Router in a vPC Topology



In Figure 2-3, router R considers N5k-1 and N5k-2 as two Layer 3 ECMP next-hop routers and runs ECMP hashing to choose which router to use as the actual next hop for a given flow. Router R connects to N5k-1 and N5k-2 via a vPC. This port channel has an IP address on router R, and Router R performs Layer 3 peering with N5k-1 and N5k-2 over this port channel. It runs the port channel hash algorithm to choose one physical link to reach the Layer 3 next hop. Because the Layer 3 ECMP and port channel run independent hash calculations there is a possibility that when the Layer 3 ECMP chooses N5k-1 as the Layer 3 next hop for a destination address while the port channel hashing chooses the physical link toward N5k-2. In this scenario,N5k-2 receives packets from R with the N5k-1 MAC as the destination MAC.

Sending traffic over the peer-link to the correct gateway is acceptable for data forwarding, but it is suboptimal because it makes traffic cross the peer link when the traffic could be routed directly.

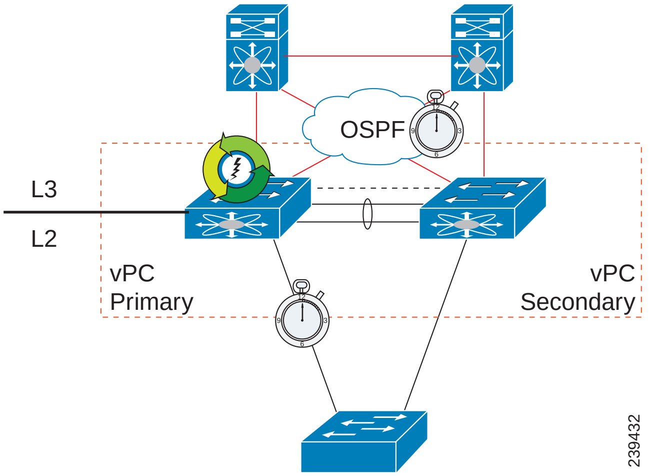
Beginning in Cisco NX-OS Release 5.0(3)N1(1b), you can use the **peer-gateway** command to allow Cisco Nexus 5500 Platform switches to perform Layer 3 forwarding if the destination MAC of the incoming packet is the MAC of its vPC peer switch. The **peer-gateway** command avoids forwarding such packets to the vPC peer link.

Note You must configure the **peer-gateway** command on both vPC peer switches.

Improved Convergence with a vPC Topology and Layer 3 Routing

Beginning in Cisco NX-OS Release 5.0(3)N1(1b), a delay timer was introduced to avoid the situation where a vPC member port is brought up before the Layer 3 is converged. For example, when one Cisco Nexus 5500 Platform switch is reloaded, the switch starts to receive traffic from hosts once the vPC member ports are up. A delay might occur before the switch establishes a routing protocol adjacency and learns all routes. During this period of the time, received traffic is dropped due to the lack of a route-to-destination address. Figure 2-4 shows an example of where the delay can be used to avoid black hole traffic when a Cisco Nexus 5000 Platform switch is configured for Layer 3 with vPC.

Figure 2-4 vPC Delay Restore



The delay restore feature allows you to configure a timed delay before vPC member ports are brought online. The delay allows the switch to learn all routes, to bring up the vPC member ports, and to forward traffic from hosts. The following example shows how to configure a timed delay of 120 seconds:

```
layer3-switch(config-vpc-domain)# delay restore ?
  <1-3600>  Delay in bringing up the vPC links (in seconds)
layer3-switch(config-vpc-domain)# delay restore 120
layer3-switch(config-vpc-domain)#
```

vPC Peer Link Failure

In addition to suspending vPC member ports, the vPC secondary switch also suspends its switched virtual interface (SVIs) when a vPC peer link is lost. When this occurs, the vPC secondary switch stops advertising the local subnets, which prevents traffic blackholing.

Layer 3 Module Failure

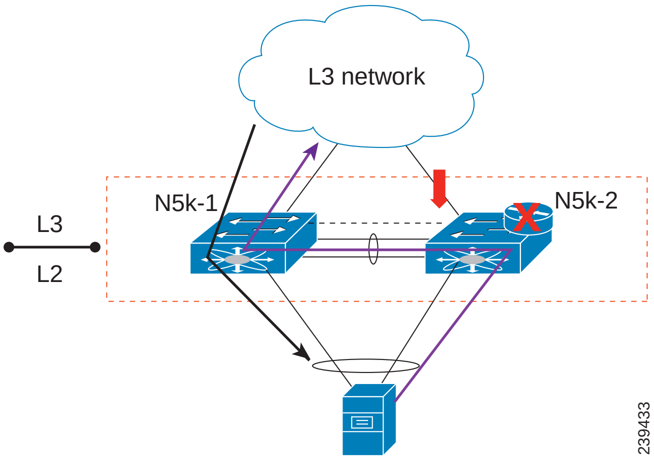
When a Layer 3 module fails on a Cisco Nexus 5500 Platform switch all Layer 3 interfaces are suspended, including Layer 3 port channel and SVI interfaces. As a result, the Layer 3 routing table on the neighboring routers is updated which results in the north to south traffic to be directed towards the peer Nexus 5500 Platform switch. The Layer 2 interfaces, including the Layer 2 port channel and out-of-band management interfaces, remain up.

In a non-vPC topology, when the Layer 3 and SVI interfaces are down, the redundant Cisco Nexus 5500 Platform switch becomes the active peer for all FHRP groups and it continues to forward traffic.

In a vPC topology, although the SVI interfaces are suspended, the vPC member ports are still up on the Cisco Nexus 5500 Platform switch. Even if the switch has a faulty Layer 3 module, Layer 2 traffic forwarding continues.

Figure 2-5 shows a topology where the Layer 3 module on N5k-2 fails. In this scenario, the Layer 3 connection toward the Layer 3 network and all SVI interfaces are suspended. However, the traffic from the hosts can still be sent to N5k-2 depending on the hash results. With the failure of the Layer 3 module, N5k-2 functions as a Layer 2 switch. It forwards the traffic to N5k-1, which forwards the traffic to the Layer 3 network. The return traffic is sent to N5k-1, which sends the traffic directly to the hosts.

Figure 2-5 Layer 3 Module Failure



Note Only the Layer 3 traffic needs to cross the peer link. The VLAN traffic is switched by N5k-2 locally.

The peer gateway is disabled on both vPC switches if the Layer 3 module fails on one switch.

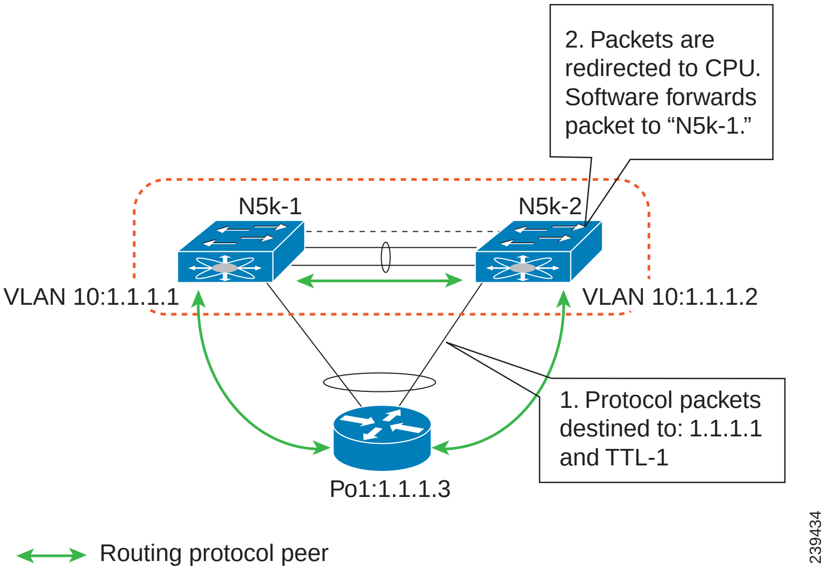
For topologies with in-band management, the failure of a Layer 3 module means that the connectivity to the management network and the management system is also lost.

Connecting to a Router in a vPC Topology

When you connect a router to a pair of Cisco Nexus 5500 Platform switches in a vPC topology and enable routing, traffic forwarding may result in suboptimal traffic paths crossing the peer link similar to the situation described in the "Layer 3 Forwarding for Packets to a Peer Switch MAC Address" section on page 2-2. We recommend that you use Layer 3 links for connections between the router and the Nexus 5500 switch, instead of a port channel with an IP address.

Figure 2-6 illustrates the topology that is not recommended. In this topology, control protocol packets may be hashed by the port channel to the wrong Cisco Nexus 5500 Platform switch, which would then forward the control packets to the correct routing peer (1.1.1.1) in the picture.

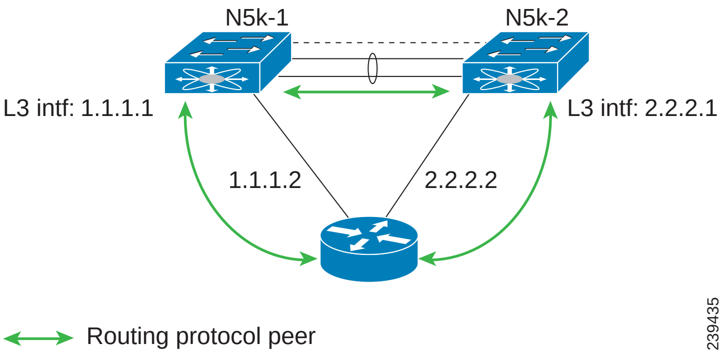
Figure 2-6 Control Traffic Forwarding in a vPC Topology


2. Packets are redirected to CPU. Software forwards packet to "N5k-1."
N5k-1
N5k-2
VLAN 10:1.1.1.1
VLAN 10:1.1.1.2
1. Protocol packets destined to: 1.1.1.1 and TTL-1
Po1:1.1.1.3
Routing protocol peer
239434

This topology is supported for unicast traffic but not for multicast traffic. In this topology, we recommend that you use Layer 3 interfaces instead of vPC interfaces to connect routers to Cisco Nexus 5500 Platform switches whenever possible.

Figure 2-7, shows the recommended topology for connectivity of routers to a vPC domain. The router connects with Layer 3 interfaces 1.1.1.2 and 2.2.2.2 to the two vPC peers and these interfaces are not part of a vPC port channel.

Figure 2-7 Connecting a Router to a vPC Domain Using Layer 3 Interfaces



Dedicated VRF For a Keepalive Interface

Beginning in Cisco NX-OS Release 5.0(3)N1(1b), the Cisco Nexus 5500 Platform switch supports VRF lite with a Layer 3 module and Enterprise license and you can create a VRF and assign the interface to a VRF. Prior to this release, two VRFs were created by default: the VRF management and VRF default. The management interface(mgmt0) and all SVI interfaces resided in the VRF management and VRF default respectively.

We recommend that you use an out-of-band management interface (mgmt0) as a vPC keepalive interface although you have the option to use the front-panel data port as a vPC keepalive interface. When you choose to use the front panel 10-Gigabit Ethernet port as the vPC keepalive interface, you should create a separate VRF for vPC keepalive packets when Layer 3 is enabled with vPC. This process eliminates the possibility of disrupting the vPC keepalive link by the wrong routes learned by a dynamic routing protocol.

This example shows how to configure a new VRF named vpc_keepalive for the vPC keepalive link and how to display the vPC peer keepalive configuration:

```
vrf context vpc_keepalive
interface Ethernet1/31
  switchport access vlan 123
interface Vlan123
  vrf member vpc_keepalive
  ip address 123.1.1.2/30
  no shutdown
vpc domain 1
  peer-keepalive destination 123.1.1.1 source 123.1.1.2 vrf vpc_keepalive

layer3-switch# show vpc peer-keepalive

vPC keep-alive status           : peer is alive
--Peer is alive for             : (154477) seconds, (908) msec
--Send status                   : Success
--Last send at                  : 2011.01.14 19:02:50 100 ms
--Sent on interface             : Vlan123
--Receive status                : Success
--Last receive at               : 2011.01.14 19:02:50 103 ms
--Received on interface         : Vlan123
--Last update from peer         : (0) seconds, (524) msec

vPC Keep-alive parameters
--Destination                   : 123.1.1.1
--Keepalive interval            : 1000 msec
--Keepalive timeout             : 5 seconds
--Keepalive hold timeout        : 3 seconds
--Keepalive vrf                 : vpc_keepalive
--Keepalive udp port            : 3200
--Keepalive tos                 : 192
```

The services provided by the Cisco Nexus 5500 Platform switch, such as Ping, SSH, Telnet, and RADIUS, are VRF-aware. You must specify the VRF name in the CLI in order to use the correct routing table.

```
layer3-switch# ping 123.1.1.1 vrf vpc_keepalive
PING 123.1.1.1 (123.1.1.1): 56 data bytes
64 bytes from 123.1.1.1: icmp_seq=0 ttl=254 time=3.234 ms
64 bytes from 123.1.1.1: icmp_seq=1 ttl=254 time=4.931 ms
64 bytes from 123.1.1.1: icmp_seq=2 ttl=254 time=4.965 ms
64 bytes from 123.1.1.1: icmp_seq=3 ttl=254 time=4.971 ms
64 bytes from 123.1.1.1: icmp_seq=4 ttl=254 time=4.915 ms

--- 123.1.1.1 ping statistics ---
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min/avg/max = 3.234/4.603/4.971 ms
```

vPC Consistency Check for Layer 3 Parameters

In a vPC topology, vPC peer switches run routing protocols independently and they maintain the routing table independently. Consistency checks are not performed to verify that Layer 3 configurations in the vPC domain are configured symmetrically.

For example, if you configure a router ACL (RACL) on one SVI and you do not configure the router on the corresponding SVI on the vPC peer, a syslog message is not displayed. You must configure the RACL on both devices. This is consistent with the operation of independent routing devices.

Similarly, if you configure peer gateway on one vPC peer and you want the same peer gateway configuration on the other vPC peer, you must configure the peer gateway on the vPC peer.

To confirm that a vPC domain is correctly configured for Layer 3 operations, the following configurations must be consistent:

- SVI configurations
- RACLs
- Routing protocol configurations

Multicast Interaction in a vPC Topology

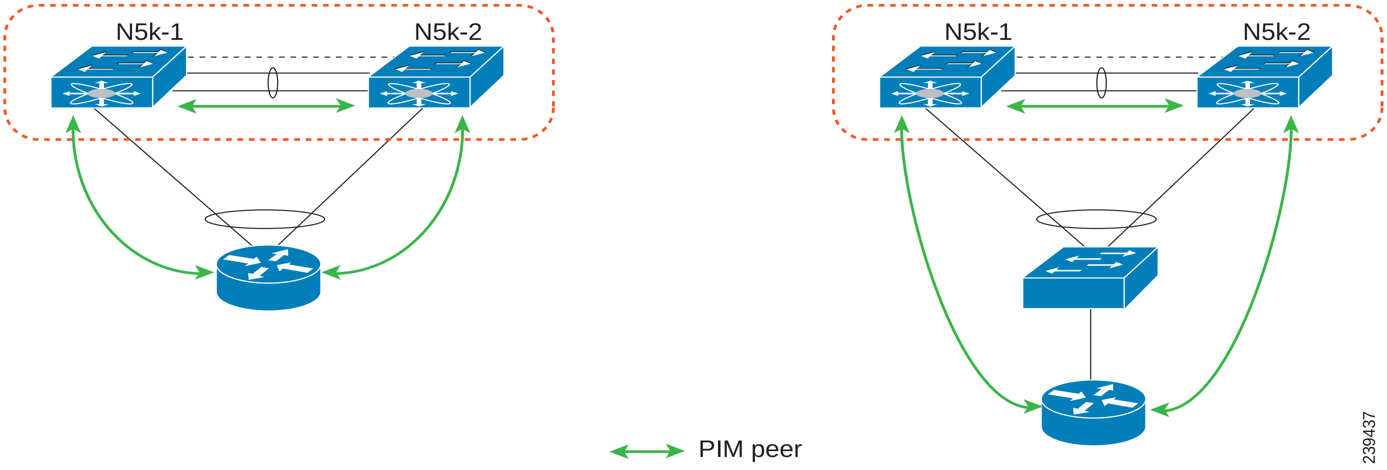
This section includes the following topics:

- Unsupported Multicast Topology, page 2-9
- Multicast Routing Table Size, page 2-10

Unsupported Multicast Topology

Figure 2-8 shows an unsupported multicast topology in a vPC configuration.

Figure 2-8 Unsupported Multicast Topology with a vPC



When a PIM router is connected to Cisco Nexus 5500 Platform switches in a vPC topology, the PIM join messages are received only by one switch. The multicast data might be received by the other switch.

Note Multicast forwarding in this topology does not work.

Multicast Routing Table Size

When you enable a vPC on a Nexus 5500 Platform switch, one multicast route (*,G) or (S,G) requires two entries in the routing table; therefore, the multicast routing table size is half the size of what is supported in topologies where vPC is not enabled.

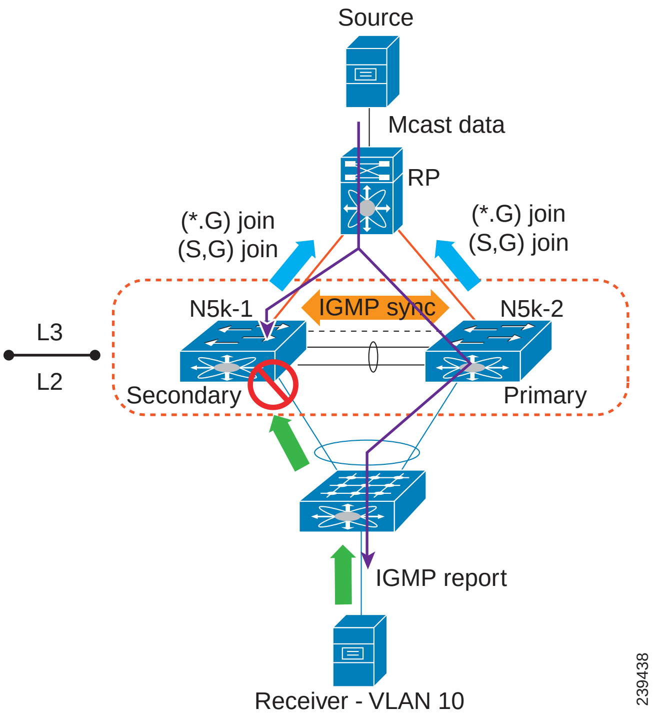
Beginning with Cisco NX-OS Release 5.0(3)N1(1b), the Cisco Nexus 5500 Platform multicast routing table size is 2000 entries in non-vPC topologies and 1000 entries in vPC topologies.

Faster Convergence with the Prebuilt Source Tree

In a non-vPC topology, only the designated router (DR) can join the source tree. In a vPC topology, when a receiver is connected to a Cisco Nexus 5500 Platform switch or Fabric Extender (FEX) via vPC, both peer switches initiate a PIM (S,G) join toward the source DR. In a topology where both vPC peer switches have equal costs to the source, the vPC primary switch wins the assert and forwards multicast traffic for receivers connected to the Nexus 5500 Platform switch or FEX using the vPC. The vPC secondary switch also joins the source tree and pulls the multicast data. To prevent data duplication, the vPC secondary switch drops the data due to an empty outgoing interface (OIF) list. Once the vPC secondary switch detects the failure of the vPC primary switch, it adds the receiver VLAN to the OIF list and starts to forward the multicast traffic immediately. Because the vPC secondary switch joins the source tree before the failure, it does not need to initiate the (S,G) join and waits for the tree to be built. As a result, it improves the convergence time in the case of a failure with the active multicast traffic forwarder.

Figure 2-9 shows one receiver that is connected to a dual-homed FEX. The source and Rendezvous Point (RP) are in the Layer 3 network. N5k-2, which is the VPC primary switch, is the multicast traffic forwarder for receivers in VLAN 10.

Figure 2-9 vPC Switch as the Receiver Designated Router



This example shows the output of the multicast routing table and VLAN 10 appears in the OIF list of (S,G) entry on N5k-2. N5k-1 joins the source tree but its OIF list remains empty.

```
N5k-1# show ip mroute 224.1.1.1
IP Multicast Routing Table for VRF "default"

(*, 224.1.1.1/32), uptime: 03:03:31, pim ip igmp
  Incoming interface: Ethernet1/6, RPF nbr: 155.1.2.2
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 03:01:16, igmp

(155.1.3.100/32, 224.1.1.1/32), uptime: 02:13:32, ip pim mrib
  Incoming interface: Ethernet1/6, RPF nbr: 155.1.2.2
  Outgoing interface list: (count: 0)


N5k-2# show ip mroute 224.1.1.1
IP Multicast Routing Table for VRF "default"

(*, 224.1.1.1/32), uptime: 01:48:07, igmp pim ip
  Incoming interface: Ethernet1/6, RPF nbr: 155.1.2.6
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 01:48:07, igmp

(155.1.3.100/32, 224.1.1.1/32), uptime: 01:00:24, ip pim mrib
  Incoming interface: Ethernet1/6, RPF nbr: 155.1.2.6
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 00:55:14, mrib
```

The multicast forwarding algorithm applies to all hosts that are connected to the Cisco Nexus 5500 Platform switch or the FEX in a VPC topology, including hosts directly connected to the switch or hosts connected to straight-through FEX topology.

Using a vPC Switch as a Designated Router (PIM DR)

This section includes the following topics:

DR Election and Source Registration

In vPC topologies, a DR election occurs based on the DR priority and the IP address. The elected DR is responsible for sending the source registration toward the RP. When multicast traffic from a directly connected source is received by the non-DR peer switch, the peer switch notifies the DR switch using a Cisco Fabric Services (CFS) message about the source and group address. The DR generates source registration packets to the rendezvous point (RP).

Multicast Data Forwarding

The Cisco Nexus 5500 Platform switch implements a dual-DR mechanism where both vPC peer switches can forward multicast traffic from directly connected sources. The data forwarding rules are as follows:

- The peer switch receives multicast packets from a directly connected source, performs an mroute lookup, and replicates packets for each interface in the OIF list.
- If the OIF is a VLAN trunked over a vPC peer link, one copy is sent over to the peer link for each VLAN that is present in the OIF list. By default, the vPC peer link is considered an mrouter port. Therefore, the multicast packets are sent over to the peer link for each receiving VLAN. You can use the **no ip igmp snooping mrouter vpc-peer link** command to avoid sending multicast traffic over a peer link for each receiver VLAN when there are no orphan ports.

This example shows how to avoid sending the multicast traffic in this scenario:

```
switch-Layer 3-1(config)# no ip igmp snooping mrouter vpc-peer link
Warning: IGMP Snooping mrouter vpc-peer link should be globally disabled on peer VPC
switch as well.
switch-Layer 3-1(config)#
```

With the above CLI configured, the multicast packet is only sent to peer link for VLANs that have orphan ports.

This example shows how to display the list of all orphan ports:

```
switch-Layer 3-1# show vpc orphan-ports
Note:
--------::Going through port database. Please be patient.::--------

VLAN           Orphan Ports
-------        -------------------------
1              Eth1/15
switch-Layer 3-1#
```

Note As of Cisco NX-OS Release 5.0(3)N1(1b), the **no ip igmp snooping mrouter vpc-peer link** command cannot be applied with FEX dual-homed topologies due to a software limitation. The command is used only for interfaces on a Cisco Nexus 5500 Platform switch. This software limitation will be removed in a future software release.

One post-routed multicast packet is sent to a vPC peer link using a reserved VLAN. To configure the reserved VLAN, use the follow commands:

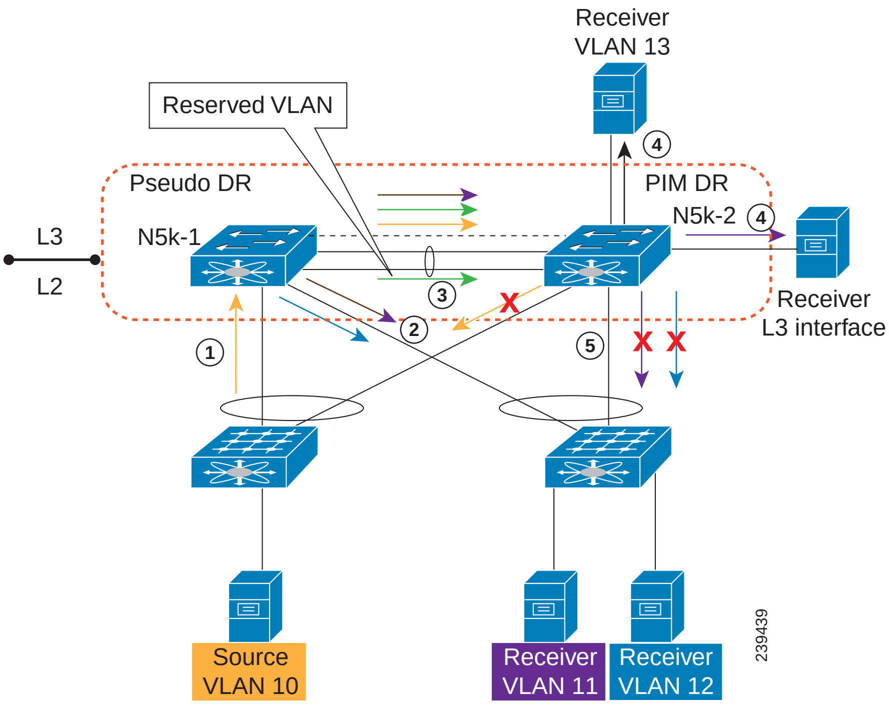
```
switch-Layer 3-1(config)# vpc bind-vrf vrf name vlan VLAN ID
switch-Layer 3-1(config)# vpc bind-vrf default vlan 3000
```

One reserved VLAN is required for each VRF. Without these commands, the receivers in non-vPC VLAN and the receivers connected to Layer 3 interfaces may not receive multicast traffic. The non-vPC VLANs are the VLANs that are not trunked over a peer link.

Multicast traffic that is received over a peer link (with a VLAN ID other than the reserved VLAN ID) is not routed. The multicast traffic is treated as Layer 2 frames that are sent to orphan ports only and not to vPC member ports. The multicast traffic that is received over a peer link with a reserved VLAN ID is routed to a non-vPC VLAN (shown as VLAN 13 in Figure 2-10) and receivers behind the Layer 3 interface. The receivers behind the Layer 3 interface can be hosts directly connected to the Cisco Nexus 5500 Platform switch using Layer 3 interfaces or a router joins the source tree.

Figure 2-10 shows the multicast forwarding rules in a vPC dual-DR topology. In this topology, the source in VLAN 10 and receivers in VLAN 11 and VLAN 12 are the vPC hosts (although in this example they are hosts behind a dual-homed FEX topology where the same rule applies to hosts directly to a Cisco Nexus 5500 Platform switch in a vPC topology). VLAN 13 is a non-vPC VLAN and resides only on N5k-2.

Figure 2-10 Multicast Data Forwarding



The forwarding process is as follows:

1. IGMP joins from the hosts are synchronized between the two vPC peer switches. N5k-2 is elected as the PIM DR for VLAN 10. Multicast traffic is sent over to N5k-1.
2. The routing engine of N5k-1 performs an mroute lookup and replicates packets to VLAN 11 and VLAN 12. The data packets for VLAN 11 and VLAN 12 are sent to the FEX which in turn sends packets to the two receivers;
3. By default, the replicated packets are sent to the vPC peer link for the source VLAN as well as each receiver VLAN (VLAN 10, VLAN 11, and VLAN 12) in this example. When you use the **no ip igmp snooping mrouter vpc-peer-link** command, the multicast packets are not sent to the peer link for VLAN 10, VLAN 11, and VLAN 12 because there are no orphan ports. One copy of the packets is sent to the peer link with the reserved VLAN 3000 which was configured using the **vpc bind-vrf default vlan 3000** command.

Note In Cisco NX-OS Release 5.0(3)N1(1b), the **no ip igmp snooping mrouter vpc-peer-link** command cannot be applied with a FEX dual-homed topology.

4. For the multicast traffic received from the peer link, if the VLAN ID is the reserved VLAN ID 3000, the N5k-2 route engine performs a Layer 3 lookup and replicates packets to VLAN 13 (a non-vPC VLAN) and receivers behind Layer 3 interfaces.
5. For the multicast packets received over the peer link, VLAN 10, VLAN 11, and VLAN 12 are dropped by N5k-2 to prevent duplicated packets being sent to the vPC hosts. If any orphan ports are in VLAN 10, VLAN 11, and VLAN 12, the packets are bridged to the orphan ports.

Software Upgrade and Downgrade Impact

In Cisco NX-OS Release 5.0(3)N1(1b), the Cisco Nexus 5500 Platform switch does not support ISSUs when Layer 3 modules are installed and Layer 3 features are enabled. Use the **install all** command and the **show install all impact** command to determine the impact of the software upgrade and to indicate whether the software upgrade with Layer 3 features enabled will be disruptive and would require a switch and FEX reload.

show install all impact kickstart

This example shows the output of the **show install all** command:

```
Layer 3-N5548-2# show install all impact kickstart
n5000-uk9-kickstart.5.0.3.N1.0.271.bin.upg system n5000-uk9.5.0.3.N1.0.271.bin.upg

Verifying image bootflash:/n5000-uk9-kickstart.5.0.3.N1.0.271.bin.upg for boot variable
"kickstart".
[####################] 100% -- SUCCESS

Verifying image bootflash:/n5000-uk9.5.0.3.N1.0.271.bin.upg for boot variable "system".
[####################] 100% -- SUCCESS

Verifying image type.
[###########         ]  50%
[####################] 100% -- SUCCESS

Extracting "system" version from image bootflash:/n5000-uk9.5.0.3.N1.0.271.bin.upg.
[####################] 100% -- SUCCESS
```

```
Extracting "kickstart" version from image
bootflash:/n5000-uk9-kickstart.5.0.3.N1.0.271.bin.upg.
[####################] 100% -- SUCCESS

Extracting "bios" version from image bootflash:/n5000-uk9.5.0.3.N1.0.271.bin.upg.
[####################] 100% -- SUCCESS

Extracting "fexth" version from image bootflash:/n5000-uk9.5.0.3.N1.0.271.bin.upg.
[####################] 100% -- SUCCESS

Performing module support checks.
[####################] 100% -- SUCCESS

Notifying services about system upgrade.
[####################] 100% -- SUCCESS



Compatibility check is done:
Module  bootable          Impact  Install-type  Reason
------  --------  --------------  ------------  ------
     1       yes      disruptive         reset  Non-disruptive install not supported if
Layer 3 was enabled
   100       yes      disruptive         reset  Non-disruptive install not supported if
Layer 3 was enabled



Images will be upgraded according to following table:
Module       Image         Running-Version            New-Version  Upg-Required
------  ----------  ----------------------  ----------------------  ------------
     1      system            5.0(3)N1(1b)           5.0(3u)N1(1u)           yes
     1   kickstart            5.0(3)N1(1b)           5.0(3u)N1(1u)           yes
     1        bios      v3.4.0(01/13/2011)      v3.4.0(01/13/2011)           no
   100       fexth            5.0(3)N1(1b)           5.0(3u)N1(1u)           yes
     1   power-seq                    v3.0                    v3.0           no
     2   power-seq                    v1.0                    v1.0           no
     1          uC               v1.0.0.14               v1.0.0.14           no

Layer 3-N5548-2#
```

You can perform a nondisruptive ISSU from an earlier release to NX-OS Release 5.0(3)N1(1b) when upgrading without Layer 3 features enabled.

show spanning-tree issu-impact

To verify that the current STP topology is consistent with ISSU requirements, use the **show spanning-tree issu-impact** command to display the STP configuration and whether or not there are potential STP issues.

This example shows how to display information about the STP impact when performing an ISSU:

```
nexus5010# show spanning-tree issu-impact
For ISSU to Proceed, Check the Following Criteria :
1. No Topology change must be active in any STP instance
2. Bridge assurance(BA) should not be active on any port (except MCT)
3. There should not be any Non Edge Designated Forwarding port (except MCT)
4. ISSU criteria must be met on the VPC Peer Switch as well

Following are the statistics on this switch
```

```
No Active Topology change Found!
Criteria 1 PASSED !!

No Ports with BA Enabled Found!
Criteria 2 PASSED!!

No Non-Edge Designated Forwarding Ports Found!
Criteria 3 PASSED !!

ISSU Can Proceed! Check Peer Switch.
```

For information on upgrade procedures, see the *Cisco Nexus 5000 Series NX-OS Upgrade and Downgrade Guide*.

INDEX