Design and Configuration Guide: Best Practices for Virtual Port Channels (vPC) on Cisco Nexus 7000 Series Switches

Revised: November 2015

Contents

Introduction

This guide provides best practices for using virtual Port Channels (vPCs) on Cisco Nexus® 7000 Series Switches.

Use this document in conjunction with the complete Cisco Nexus 7000 Series documentation, which you will find at: http://www.cisco.com/en/US/products/ps9402/tsd_products_support_series_home.html.

vPC user guide is located at the following link (CCO):
http://www.cisco.com/en/US/docs/switches/datacenter/sw/6_x/nx-os/interfaces/configuration/guide/if_vPC.html.

(vPC user guide is contained within NX-OS interface configuration guide).

The best practices in this document follow a consistent pattern that makes the information in each section easy to find. Best practices for vPCs are organized in the following ways:

- vPC description and Terminology
- vPC deployment scenario
- Best Practices for Building a vPC Domain
- Best Practices for vPC Components Configuration
- Best practices for vPC in mixed chassis mode (M1/F1 ports in same system or VDC)
- Best practices for attaching a device to vPC domain
- Best practices for Data Center Interconnect and Encryption
- Best Practices for Spanning Tree Protocol Interoperability
- Best practices for Layer 3 and vPC
- Best practices for HSRP/VRRP and vPC
- Best practices for Network Services and vPC
- Best practices for Multicast and vPC
- Best practices for FEX and vPC
- Best practices for VDC and vPC

This document also covers ISSU operations related to vPC and give details about latest vPC enhancements (object-tracking, peer-gateway, peer-switch, reload restore, delay restore, graceful type-1 check, auto-recovery, orphan ports suspend, host vPC).

vPC scalability numbers are published at the following link (CCO):
http://www.cisco.com/en/US/docs/switches/datacenter/sw/verified_scalability/b_Cisco_Nexus_7000_Series_NX-OS_Verified_Scalability_Guide.html#reference_32EB4DB289634F6FA8885FDFD8E71F5F.

Take into consideration these scale numbers to design properly a network based on vPC technology.

Note: This document does not cover the following topic:

- vPC+ (vPC used in the context of FabricPath)

vPC Description and Terminology

Benefits of vPC

vPC is a virtualization technology that presents both Cisco Nexus 7000 Series paired devices as a unique Layer 2 logical node to access layer devices or endpoints. vPC belongs to Multichassis EtherChannel [MCEC] family of technology.

A virtual port channel (vPC) allows links that are physically connected to two different Cisco Nexus 7000 Series devices to appear as a single port channel to a third device. The third device can be a switch, server, or any other networking device that supports link aggregation technology.

vPC provides the following technical benefits:

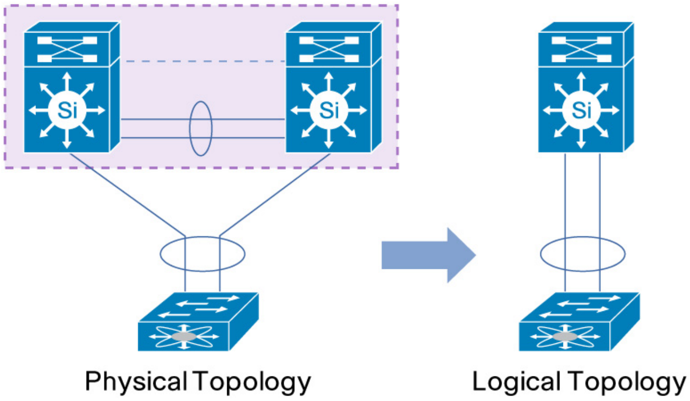
- Eliminates Spanning Tree Protocol (STP) blocked ports
- Uses all available uplink bandwidth
- Allows dual-homed servers to operate in active-active mode
- Provides fast convergence upon link or device failure
- Offers dual active/active default gateways for servers

vPC also leverages native split horizon/loop management provided by port-channeling technology: a packet entering a port-channel cannot immediately exit that same port-channel.

By using vPC, users get the immediate operational and architectural advantages:

- Simplifies network design
- Build highly resilient and robust Layer 2 network
- Enables seamless virtual machine mobility and server high-availability clusters
- Scales available Layer 2 bandwidth, increasing bisectional bandwith
- Grows the size of the Layer 2 network

Figure 1. Creating a Single Logical Node through vPC (virtual Port Channel) Technology



vPC leverages both hardware and software redundancy aspects:

- vPC uses all port-channel member links available so that in case an individual link fails, hashing algorithm will redirect all flows to the remaining links.
- vPC domain is composed of two peer devices. Each peer device processes half of the traffic coming from the access layer. In case a peer device fails, the other peer device will absorb all the traffic with minimal convergence time impact.
- Each peer device in the vPC domain runs its own control plane, and both devices work independently. Any potential control plane issues stay local to the peer device and does not propagate or impact the other peer device.

From a Spanning-Tree standpoint, vPC eliminates STP blocked ports and uses all available uplink bandwidth. Spanning-Tree is used as a fail safe mechanism and does not dictate L2 path for vPC-attached devices.

Withing a vPC domain, user can connect access devices in multiple ways: vPC-attached connections leveraging active/active behavior with port-channel, active/standby connectivity using spanning-tree, single attachment without spanning-tree running on the access device.

All these connectivity configurations are fully supported and will be detailed in the following document.

NX-OS Version Requirement for vPC

vPC technology is supported since NX-OS 4.1.3. (i.e since the inception of NEXUS 7000 platform).

NX-OS appropriate version depends on line cards configuration (M1, F1 or F2), chassis type (7010, 7018 or 7009) and Fabric Module generation (FM generation 1 [46Gbps per module] or generation 2 [110Gbps per module]).

Please refer to the following URL to check the recommended NX-OS version: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/recommended_releases/recommended_nx-os_releases.html.

[Minimum Recommended Cisco NX-OS Releases for Cisco Nexus 7000 Series Switches].

NX-OS release notes for each code release can be found at this location: http://www.cisco.com/en/US/products/ps9402/prod_release_notes_list.html.

NX-OS License Requirement for vPC

vPC feature is included in the base NX-OS software license.

Hot Standby Router Protocol (HSRP), Virtual Router Redundancy Protocol (VRRP), Link Aggregation Control Protocol (LACP) are also included in this base license.

Layer 3 features like Open Shortest Path First (OSPF) protocol or Intermediate-System-to-Intermediate System (IS-IS) protocol require LAN_ENTERPRISE_SERVICES_PKG license.

Virtual device contexts (VDCs) requires LAN_ADVANCED_SERVICES_PKG license.

Components of vPC

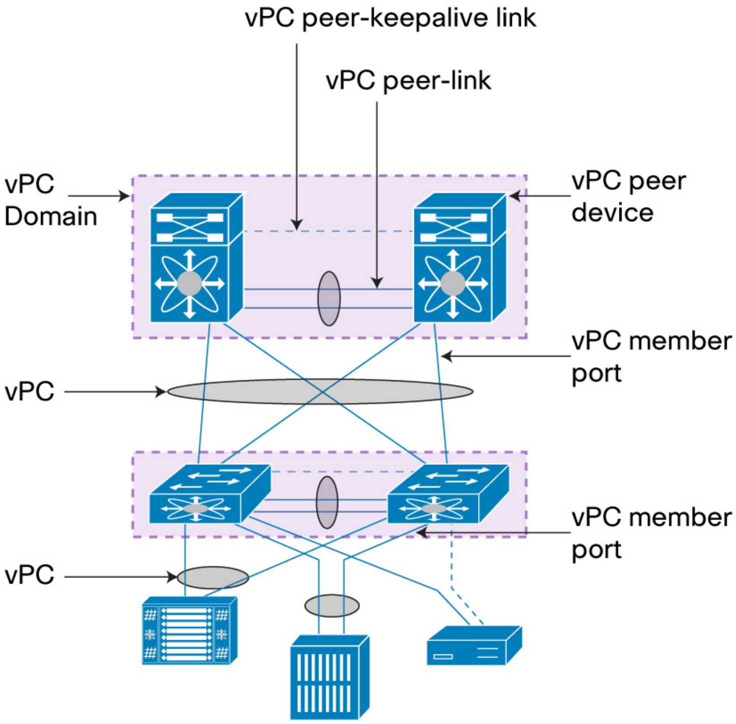
Table 1 lists important terms you need to know to understand vPC technology. These terms are used throughout this guide.

Table 1. vPC Terms

Term	Meaning
vPC	The combined port-channel between the vPC peers and the downstream device. A vPC is a L2 port type: switchport mode trunk or switchport mode access
vPC peer device	A vPC switch (one of a Cisco Nexus 7000 Series pair).
vPC domain	Domain containing the 2 peer devices. Only 2 peer devices max can be part of same vPC domain.
vPC member port	One of a set of ports (that is, port-channels) that form a vPC (or port-channel member of a vPC).
vPC peer-link	Link used to synchronize the state between vPC peer devices. It must be a 10-Gigabit Ethernet link. vPC peer-link is a L2 trunk carrying vPC VLAN.
vPC peer-keepalive link	The keepalive link between vPC peer devices; this link is used to monitor the liveness of the peer device.
vPC VLAN	VLAN carried over the vPC peer-link and used to communicate via vPC with a third device. As soon as a VLAN is defined on vPC peer-link, it becomes a vPC VLAN
non-vPC VLAN	A VLAN that is **not** part of any vPC and not present on vPC peer-link.
Orphan port	A port that belong to a single attached device. vPC VLAN is typically used on this port.
Cisco Fabric Services (CFS) protocol	Underlying protocol running on top of vPC peer-link providing reliable synchronization and consistency check mechanisms between the 2 peer devices.

Figure 2 shows the different components of vPC and how they are related.

Figure 2. vPC Components



Best practices to build vPC peer-link and vPC peer-keepalive link will be described in the section “Best Practices for Building a vPC Domain”.

Recommendations to connect an orphan port to vPC domain will be described in the section “ Best practices for attaching a device to vPC domain”.

vPC Data-Plane Loop Avoidance

vPC performs loop avoidance at data-plane layer instead of control plane layer for Spanning Tree Protocol.

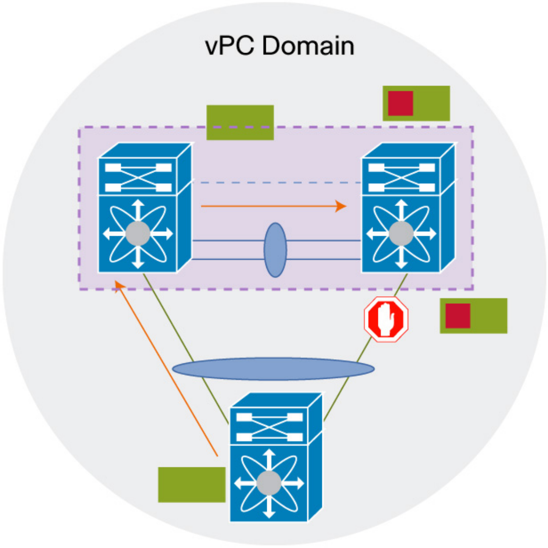
All logics are implemented directly in hardware on vPC peer-link ports, avoiding any dependancy to CPU utilization.

vPC peer devices always forward traffic locally when possible. vPC peer-link does not typically forward data packets and it is usually considered as a control plane extension in a steady state network (vPC peer-link used to synchronize information between the 2 peer devices as mac address, vPC member state information, IGMP).

vPC loop avoidance rule states that traffic coming from vPC member port, then crossing vPC peer-link is NOT allowed to egress any vPC member port; however it can egress any other type of port (L3 port, orphan port, …).

The only exception to this rule occurs when vPC member port goes down. vPC peer devices exchange member-port states and reprogram in hardware the vPC loop avoidance logic for that particular vPC. The peer-link is then used as backup path for optimal resiliency. Traffic need not ingress a vPC member port for this rule to be applicable. The vPC loop avoidance rulle exception is depicted in the figure below:

Figure 3. vPC Loop Avoidance Rule Exception



vPC Deployment Scenarios

vPC is typically used at the access or aggregation layer of the data center. At access layer, it is used for active/active connectivity from network endpoint (server, switch, NAS storage device.) to vPC domain. At aggregation layer, it is used for both active/active connectivity from network endpoint to vPC domain and active/active default gateway for L2/L3 boundary.

However, because vPC provides capabilites to build a loop free topology, it is also commonly used to interconnect two separate data centers together at layer 2, allowing extension of VLAN across the 2 sites.

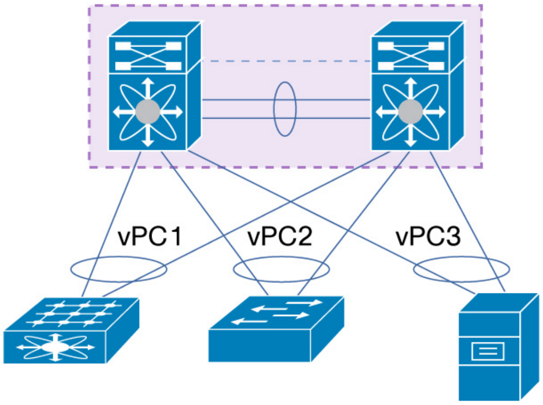
The 2 common deployment scenarios using vPC technology are listed as below:

- Inside Data Center:
  - Single-sided vPC (access layer or aggregation layer)
  - Double-sided vPC, also called multilayer vPC (access layer using vPC interconnected to aggregation layer using vPC)
- Across Data Center i.e vPC for Data Center Interconnect (DCI):
  - Multilayer vPC for Aggregation and DCI
  - Dual Layer 2 /Layer 3 Pod Interconnect

Single-Sided vPC

Figure 4 shows a single-sided vPC topology. In single-sided vPC, access devices are directly dual-attached to pair of Cisco Nexus 7000 Series Switches forming the vPC domain.

Figure 4. Single-Sided vPC Topology



The access device can be any endpoint equipement (L2 switch, rack-mount server, blade server, firewall, load balancer, network attached storage [NAS] device). Only prerequisite for the access device is to support port-channeling (or link aggregation) technology:

- LACP mode active
- LACP mode passive
- Static bundling (mode ON)

Strong Recommendation:
Use LACP protocol when connecting access device to vPC domain.

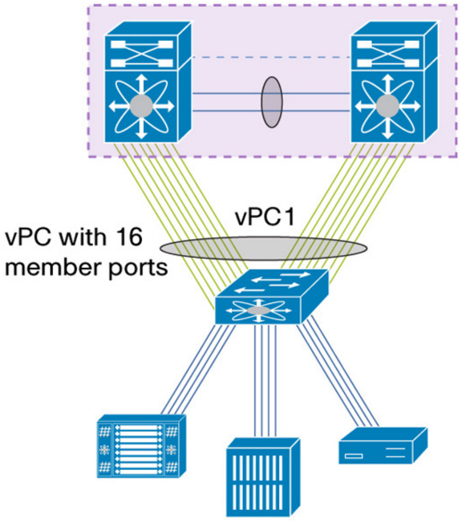
Depending on type of line card used for vPC member ports, maximum number of port-channel member ports can vary from 16 to 32:

- vPC with Cisco Nexus M1 Series module line-card: 16 active member ports (8 on peer device 1 and 8 on peer device 2)

- vPC with Cisco Nexus F1/F2 Series module line card: 32 active member ports (16 on peer device 1 and 16 on peer device 2)

Beginning with Cisco NX-OS Software Release 4.1(3)N1(1a), the Cisco Nexus 5000 Series is capable of supporting 16 active member ports per port-channel. As Figure 5 shows, connecting the Cisco Nexus 5000 Series to a vPC domain gives a compelling topology where vPC can be sized up to 160 Gbps (16 x 10-Gbps ports).

Figure 5. 16-Way Port-Channel in Single-Sided vPC Topology



Note: Although not displayed in Figure 4 or 5, orphan ports or active/standby attached devices (that is, using Spanning Tree Protocol) are fully supported in a vPC topology.

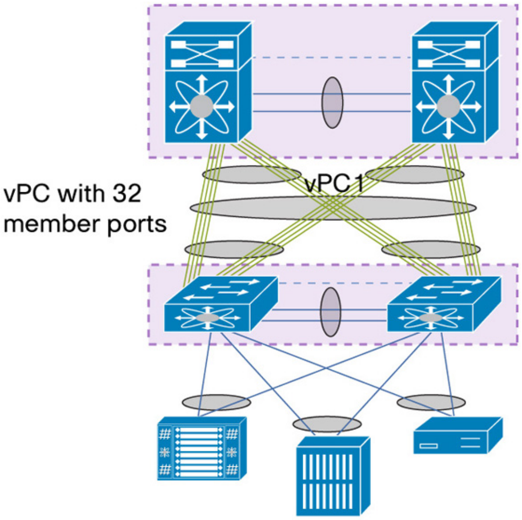
Double-Sided vPC

Figure 6 shows a double-sided vPC topology. This topology superposes two layers of vPC domain and the bundle between vPC domain 1 and vPC domain 2 is by itself a vPC.

vPC domain at the bottom is used for active/active connectivity from enpoint devices to network access layer.

vPC domain at the top is used for active/active FHRP in the L2/L3 boundary aggregation layer.

Figure 6. Double-Sided vPC Topology



Benefits of double-sided vPC over single-sided vPC topology are listed below:

- Enables a larger Layer 2 domain.
- Provides a higher resilient architecture. In double-sided vPC, two access switches are connected to two aggregation switches whereas in single-sided vPC, one access switch is connected to two aggregation switches.
- Provides more bandwidth from the access to aggregation layer. Using a Cisco Nexus F1 or F2 Series modules line card for vPC and Cisco Nexus 5000 Series Switches with Release 4.1(3)N1(1a) or later, a vPC with 32 active member ports (that is, 320 Gbps) can be instantiated.

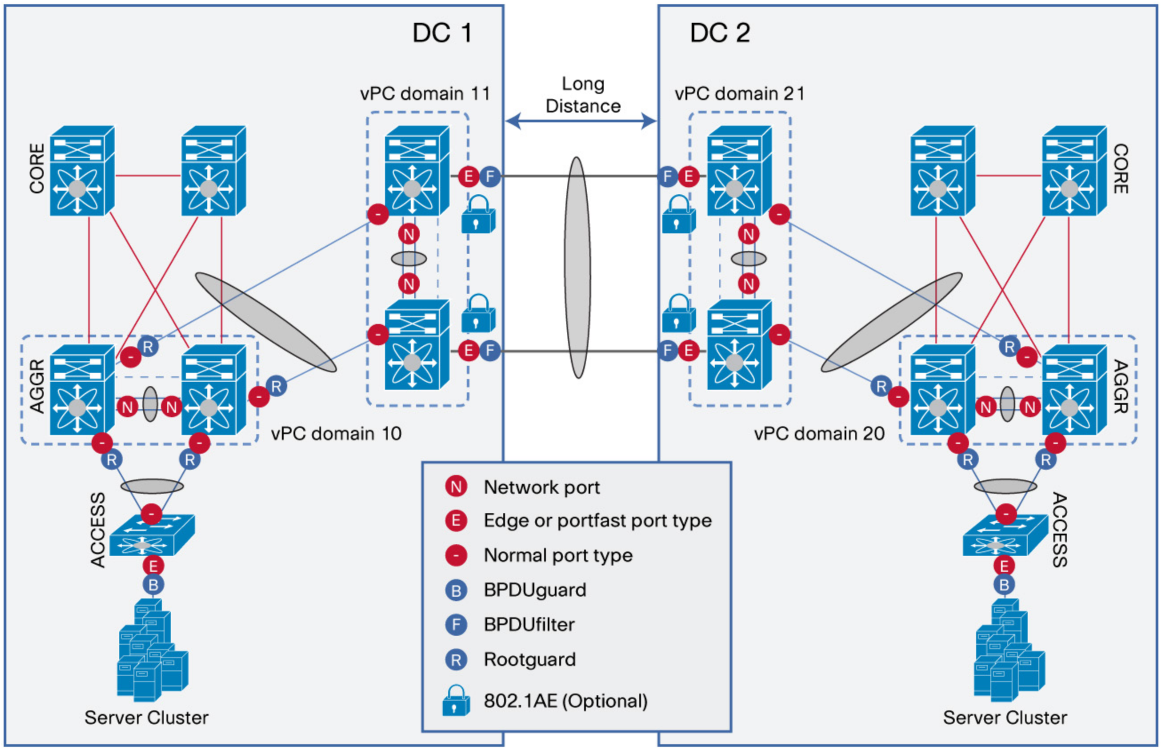
Configuration sample for double-sided vPC is provided in the “attaching to vPC” section.

Multilayer vPC for Aggregation and DCI

vPC provides capabilities to build a loop-free topology, and as such it makes the technology a good fit for Data Center Interconnect (DCI) deployments. In this scenario, a dedicated layer of vPC domain (adjacent to aggregation layer which also runs vPC) is used to interconnect the 2 data centers together.

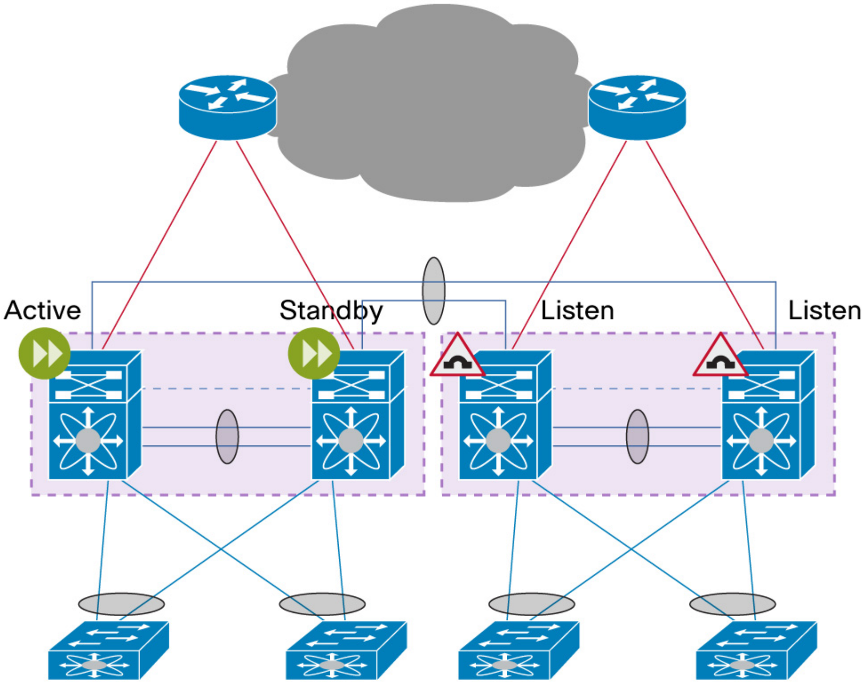
The design is called multilayer vPC for aggregation and DCI, as shown in Figure 7.

Figure 7. vPC for Data Center Interconnect: Multilayer vPC for Aggregation and DCI



Another design is to interconnect directly between vPC aggregation layer, without using any dedicated vPC layer for DCI. This design is referred to as dual Layer 2 /Layer 3 pod interconnect and is shown in Figure 8.

Figure 8. vPC for Data Center Interconnect: Dual Layer 2 /Layer 3 Pod Interconnect



vPC as DCI technology is intended to interconnect two data centers in maximum. If you need to interconnect more than 2 data centers, recommendation is to use Overlay Transport Virtualization (OTV) solution.

“Best practices for Data Center Interconnect and Encryption” section will describe vPC used as DCI technology.

Strong Recommendation:
Use vPC to interconnect a maximum of 2 data centers. Use OTV when more than 2 data centers need to be interconnected.

Best Practices for Building a vPC Domain

Building a vPC Domain

A vPC domain defines the grouping of switches participating in the vPC. As of today, only two Cisco NEXUS 7000 Series Switches can form a vPC domain.

From a configuration standpoint, vPC domain provides context to define global vPC system parameters.

User enters into vPC domain sub-commands to configure vPC options and features like peer-gateway, peer-swtich and so on.

The process of building a vPC domain involves multiple steps that should be completed in the following order:

1. Globally configure a vPC domain identifier on both vPC devices. The domain ID must be the same on both peer devices.
2. Configure vPC peer-keepalive link on both peer devices and ensure that the vPC peer-keepalive link is operational. If not, vPC domain cannot successfully be formed.

3. Configure or reuse an ISL (Inter Switch Link) L2 trunk port-channel between the vPC peer devices. Configure the port-channel as a vPC peer-link on both peer devices and ensure that the port-channel is operational.
4. Configure or reuse port-channels from the access devices to Cisco Nexus 7000 Series forming vPC domain. Then configure a unique logical vPC and join the port-channels across different vPC peer devices.

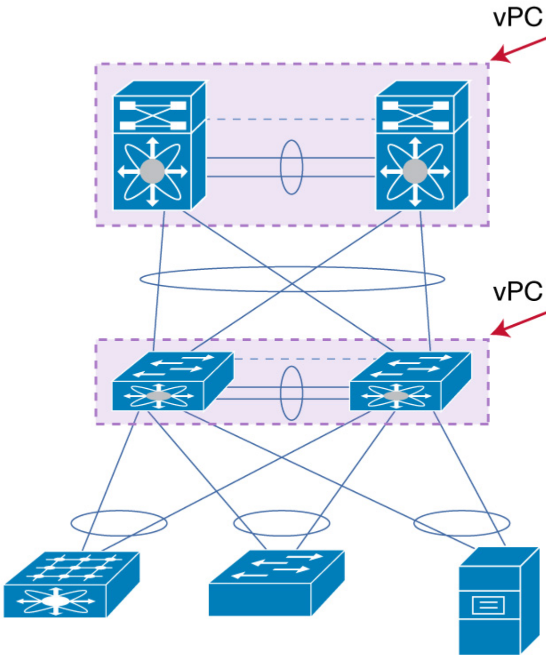
vPC Domain Identifier

vPC domain identifier is defined by the command **vpc domain <domain-id>**.

It must be identical across the 2 peer devices.

There are some situations where vPC domain ID must be configured with caution. Typical case deals with double-sided vPC topology (shown in Figure 9) where Cisco Nexus 5000 vPC layer is connected to Cisco Nexus 7000 vPC layer using a vPC. In this scenario, vPC domain identifiers **must** be different on both layers because this information is used as part of the LACP protocol. Using the same vPC domain identifiers will generate continous flaps on vPC interconnecting the NEXUS 5000 to NEXUS 7000.

Figure 9. Using a Different vPC Domain Identifier in Double-Sided vPC Topology



In case of vPC use for DCI purposes, vPC domain identifiers also **must** be different across the 2 data centers (same reason as previously, vPC domain identifier is used as part of the LACP protocol).

If user absolutely wants to use the same domain-id on both vPC domains, then knob **system-mac** (under vPC domain configuration context) must be used to force different vPC system-mac values.

Required Recommendation:
Always use different domain ID in double-sided vPC topology and in vPC for DCI topology.

vPC System-Mac and vPC Local System-Mac

Once configured, both peer devices use the vPC domain ID to automatically assign a unique vPC system MAC address, as defined below:

```
vPC system-mac = 00:23:04:ee:be:<vpc domain-id in hexadecimal>
```

For instance, vPC domain 10 will result in vPC system-mac of 00:23:04:ee:be:0a.

vPC system mac is then identical on both peer devices. This is the foundation for L2 virtualization technique with vPC: when vPC systems need to present itself as a unique logical device, it will use this unique and shared information across the 2 peer devices.

Note: It is possible to configure manually vPC system-mac value with the command **system-mac** inside vPC domain configuration context.

vPC local system mac is owned by each peer devices so it is unique per device. vPC local system mac is derived from system or VDC mac address (show vdc command to view it). vPC local system mac is used whenever vPC systems do not need to present itself as a unique logical device. This is for instance the case with orphan ports.

The show commands to visualize vPC domain ID, vPC system mac and vPC local system mac are:
show vpc and **show vpc role**.

Sample output of these commands are shown below:

```
APULIA-2-VPC_AGG2# 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   : success
Type-2 consistency status          : success
Type-2 consistency reason          : success
vPC role                           : secondary
Number of vPCs configured          : 6
Track object                       : 1
Peer Gateway                       : Enabled
Dual-active excluded VLANs         : -

vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ --------------------------------------------------
1    Po10   up     1-20,23-24,40,50,100,200,300,400-401,501,600,1000-1100
```

```
vPC status
----------------------------------------------------------------------
id   Port   Status Consistency Reason                     Active vlans
--   ----   ------ ----------- ------                     ------------
1    Po1    up     success     success                    1000-1100
<snip>
```

```
APULIA-2-VPC_AGG2# sh vpc role

vPC Role status
----------------------------------------------------
vPC role                                    : secondary
Dual Active Detection Status                : 0
vPC system-mac                              : 00:23:04:ee:be:0a
vPC system-priority                         : 32667
vPC local system-mac                        : 00:22:55:79:aa:c2
vPC local role-priority                     : 65534
```

check that vPC local system mac is derived from system or VDC mac address:

```
APULIA-2-VPC_AGG2# sh vdc

vdc_id  vdc_name              state          mac                        lc
------  --------              ---           ----------                ------
2       VPC_AGG2              active       00:22:55:79:aa:c2    m1 f1 m1xl
```

vPC system-mac and vPC local system-mac are both used in the LACP protocol as the LACP system ID.

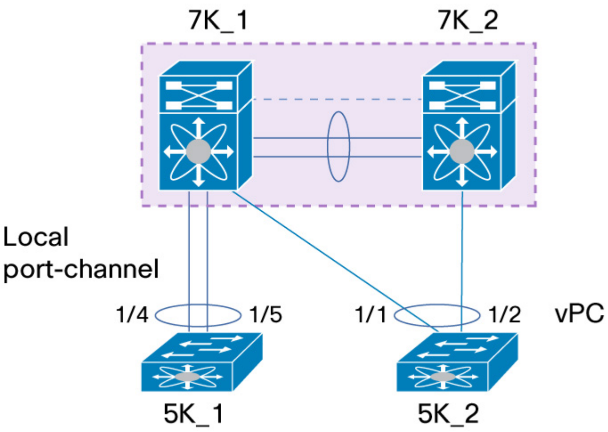
However, vPC system-mac is used only with vPC attached access devices while vPC local system-mac is used with single attached devices (orphan port or active/standby with or without STP)

Figure 10 illustrates how vPC system-mac and vPC local system-mac are used.

In this figure, the Cisco Nexus 5000 Series device 1 (5K_1) is forming a local port-channel with the Cisco Nexus 7000 Series device 1 (7K_1). As a result, 7K_1 will use its vPC local system-mac to exchange LACP information with 5K_1.

On the contrary, the Cisco Nexus 5000 Series device 2 (5K_2) is forming a vPC with the Cisco Nexus 7000 Series device 1 (7K_1) and Cisco Nexus 7000 Series device 2 (7K_2). As a result, both 7K_1 and 7K_2 will use their common vPC system-mac to exchange LACP information with 5K_2.

Figure 10. Use of vPC System-Mac and vPC Local System-Mac



show lacp neighbor on 5K_1 displays LACP system ID used by 7K_1 (which is the vPC local system-mac):

```
5K_1# sh lacp neighbor interface port-channel 1
Flags:  S - Device is sending Slow LACPDUs F - Device is sending Fast LACPDUs
        A - Device is in Active mode       P - Device is in Passive mode
port-channel1 neighbors
Partner's information
                 Partner                   Partner
   Partner
Port             System ID                 Port Number         Age         Flags
Eth1/4      32667, 0-22-55-79-ab-42        0x4206              18999       SA

  LACP Partner                       Partner                   Partner
  Port Priority                      Oper Key                  Port State
  32768                              0x8001                    0x3d

Partner's information
                 Partner                          Partner
   Partner
Port             System ID            Port Number         Age         Flags
Eth1/5      32667,0-22-55-79-ab-42    0x4208              18999       SA

LACP Partner                  Partner                    Partner
Port Priority                 Oper Key                   Port State
32768                         0x8001                     0x3d
```

show lacp neighbor on 5K_2 displays LACP system ID used by 7K_1 and 7K_2 (which the common vPC system-mac):

```
5K_2# sh lacp neighbor interface port-channel 1
Flags:  S - Device is sending Slow LACPDUs F - Device is sending Fast LACPDUs
        A - Device is in Active mode       P - Device is in Passive mode
port-channel1 neighbors
Partner's information
         Partner              Partner                    Partner
Port             System ID                 Port Number         Age         Flags
Eth1/1      32667,0-23-4-ee-be-a           0x4206              18999       SA

    LACP Partner                     Partner                     Partner
    Port Priority                    Oper Key                    Port State
    32768                            0x8001                      0x3d

Partner's information
   Partner                    Partner                    Partner
Port             System ID                 Port Number         Age         Flags
Eth1/2      32667,0-23-4-ee-be-a           0x4208              18999       SA

 LACP Partner                 Partner                    Partner
 Port Priority                Oper Key                   Port State
 32768                        0x8001                      0x3d
```

vPC Role

There are two defined vPC roles: primary and secondary.
vPC role defines which of the two vPC peer devices processes Bridge Protocol Data Units (BPDUs) and responds to Address Resolution Protocol (ARP).

Use **role priority <value>** command (under vPC domain configuration context) to force vPC role to primary for a dedicated peer device.

<value> ranges from 1 to 65535 and the lowest value will dictate the primary peer device.

In case of tie (same role priority value defined on both peer devices), lowest system mac will dictate the primary peer device.

To know which of the 2 peer devices is primary or secondary, use **show vpc role** command:

```
APULIA-1-VPC_AGG1# sh vpc role

vPC Role status
----------------------------------------------------
vPC role                                   : primary
Dual Active Detection Status               : 0
vPC system-mac                             : 00:23:04:ee:be:0a
vPC system-priority                        : 32667
vPC local system-mac                       : 00:22:55:79:ab:42
vPC local role-priority                    : 2
```

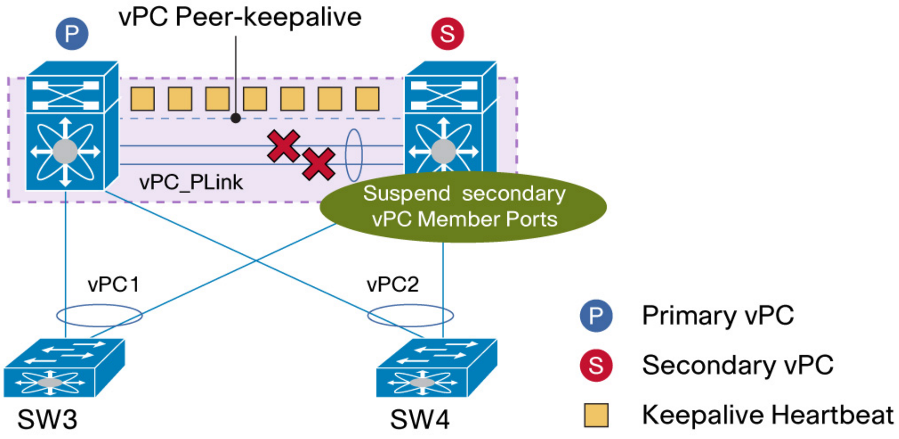
```
APULIA-2-VPC_AGG2# sh vpc role

vPC Role status
----------------------------------------------------
vPC role                                   : secondary
Dual Active Detection Status               : 0
vPC system-mac                             : 00:23:04:ee:be:0a
vPC system-priority                        : 32667
vPC local system-mac                       : 00:22:55:79:aa:c2
vPC local role-priority                    : 65534
```

General Recommendation:

For ease of operations, define vPC primary peer device on the left Nexus 7000 series device. Define vPC secondary peer device on the right Nexus 7000 series device.

vPC role (primary or secondary) matters for the behavior in a peer-link failure. As Figure 11 illustrates, when a peer-link failure occurs, only the secondary peer device shuts its vPC member ports to down state and in addition shuts all its vPC VLAN interface (or SVIs - Switch Virtual Interface) [SVI associated to vPC VLAN].

Figure 11. vPC Peer-Link Failure - Behavior On Secondary Peer Device



vPC also adds the concept of operationa role: operational primary and operational secondary.

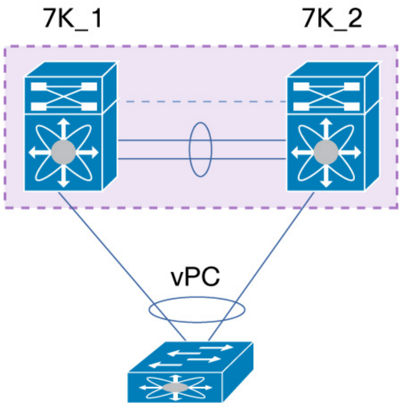
vPC primary or secondary role is dictated by the CLI configuration with role priority command.

vPC operational primary or operational secondary role is driven by the real-time behavior of peer devices.

Once vPC domain is powered up, vPC role is always equal to vPC operational role. However, as the network lives and encounters some operational events, the operational role may afterwards differ from the original vPC role.

The following sequence shows an example of divergence between vPC role and vPC operational role:

Figure 12. vPC Role and vPC Operational Role



Event 1: vPC devices powered up	7K_1: vPC role = primary	7K_2: vPC = secondary
Event 2: 7K_1 reloads and recovers	7K_1: vPC role = primary, operational secondary	7K_2: vPC role = secondary, operational primary
Event 3: 7K_2 reloads and recovers	7K_1: vPC role = primary	7K_2: vPC role = secondary

vPC role is nonpreemptive so vPC operational role is the most relevant of the 2 information.

Note: To preempt manually the operational primary role for a vPC peer device, administrator must perform following procedure: Log in to the peer device you want to change to operational primary role and configure role priority with a lower value than the other peer device. Then bounce the vPC peer-link (shut, then no shut) to force the change.

Be careful that this operation is disruptive as operational secondary peer device will shut its vPC member ports once peer-link is down.

CLI alias facility can be used to automate the sequence of commands to change vPC operational role:
cli alias name vpcpreempt conf t; vpc domain 1; role priority 1; int po 1; shut; no sh

Cisco Fabric Services (CFS) Protocol

Cisco Fabric Services (CFS) protocol provides reliable synchronization and consistency check mechanisms between the 2 peer devices and runs on top of vPC peer-link. The protocol was first implemented on MDS products (network storge devices) and then ported to NEXUS 7000.

Cisco Fabric Services (CFS) protocol performs the following functions:

- Configuration validation and comparison (consistency check)
- Synchronization of MAC addresses for vPC member ports
- vPC member port status advertisement

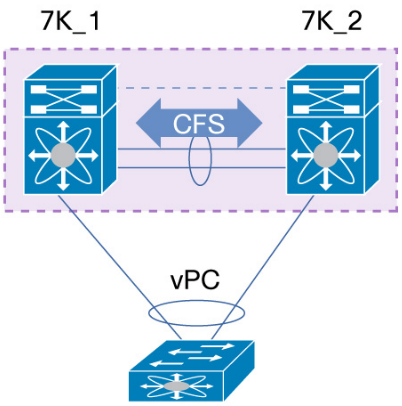
- Spanning Tree Protocol management
- Synchronization of HSRP and IGMP snooping

Cisco Fabric Services is enabled by default when vPC feature is turned on.

There is no specific Cisco Fabric Services configuration to implement. Figure 13 shows the Cisco Fabric Services message path in a vPC domain.

Cisco Fabric Services messages are encapsulated in standard Ethernet frames that are delivered between peers exclusively on the peer-link. Cisco Fabric Services messages are tagged with CoS = 4 for reliable communication.

Figure 13. CFS Message Path in a vPC Domain



To check the Cisco Fabric Services application for vPC and Cisco Fabric Services status, use **show cfs application** and **show cfs status** commands**,** as follows:

```
7K1# sh cfs application
----------------------------------------------
Application    Enabled   Scope
----------------------------------------------
arp              Yes       Physical-eth
pim              Yes       Physical-eth
stp              Yes       Physical-eth
vpc              Yes       Physical-eth
igmp             Yes       Physical-eth
l2fm             Yes       Physical-eth
```

```
7K1# sh cfs status
Distribution: Enabled
Distribution over IP: Disabled
IPv4 multicast address: 239.255.70.83
IPv6 multicast address: ff15::efff:4653
```

```
   Distribution over Ethernet: Enabled
```

Checking vPC Configuration Consistency When You Build a vPC Domain

This section contains recommendations to help ensure that there are no incompatible parameters when building a vPC domain.

Both switches in the vPC domain maintain distinct control planes. Cisco Fabric Services protocol will take care of state synchronization between both peers (including the MAC address table, Internet Group Management Protocol (IGMP) state, vPC states and so on.)

System configuration must be kept in sync. Currently this is a manual process (configuration is done separately on each device) with an automated consistency check to help ensure correct network behavior.

There are two types of consistency checks:

- Type 1 - Puts peer device or interface into a suspended state to prevent invalid packet forwarding behavior. With vPC Graceful Consistency check, suspension occurs only on the secondary peer device.
- Type 2 - Peer device or Interface still forward traffic. However they are subject to undesired packet forwarding behavior.

Type 1 and Type 2 consistency check apply both for global configuration and for vPC interface configuration.

Configuration Parameters That Must Be Identical (Type-1 Consistency Check)

After you enable vPC feature and configure vPC peer-link on both peer devices, Cisco Fabric Services messages provide a copy of the configuration on the local vPC peer device to the remote vPC peer device. The system then determines whether any of the crucial configuration parameters differ on the two devices.

Many global configuration parameters must be identical as well as vPC interfaces parameters in the same vPC domain.

The devices automatically check for compatibility. The per-interface parameters must be consistent per interface, and the global parameters must be consistent globally.

When Type 1 inconsistency check is detected, radical actions are taken:

For global configuration type 1 inconsistency check, all vPC member ports are set to down state.

For vPC interface configuration type 1 inconsistency check, the misconfigured vPC is set to down state.

Since NX-OS version 5.2, graceful consistency check has been introduced to soften vPC system reaction in occurrence to type 1 consistency check:

For global configuration type 1 inconsistency check, only vPC member ports on secondary peer device are set to down state.

For vPC interface configuration type 1 inconsistency check, only vPC member ports on secondary peer device are set to down state.

Enter the **show vpc consistency-parameters** command to display the global configuration values and vPC interfaces parameters. The displayed output are only those configurations that would limit the vPC from coming up.

Table 2 lists global configuration parameters that are taken into account for type-1 consistency check.

Table 2. Global configuration Type-1 Consistency Check

Parameter Name	Value
Spanning Tree Protocol (STP) mode	RPVST (Rapid Per LAN Spanning Tree) or MST (Multiple Spanning Tree)
STP Enable/disable state per VLAN	Yes or No
STP region configuration for Multiple Spanning Tree (MST)	Region name, region revision, region instance to VLAN mapping
STP global settings	Bridge Assurance settings Port type settings Loop Guard settings BPDU filter settings MST Simulate PVST enabled or disabled

show vpc consistency-parameters global command displays global type 1 consistency parameters.

An example of output is shown below:

```
7K1# sh vpc consistency-parameters global

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

Name                          Type   Local Value                Peer Value
-------------                 ----   ----------------------     ------------
STP Mode                      1      Rapid-PVST                 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
Interface-vlan admin up       2      200,3966-3967              200,3966-3967
Interface-vlan routing        2      1,40,200,3966-3967         1,40,200,3966-3967
capability

Allowed VLANs                 -      1-20,23-24,40,50,100,2     1-20,23-
                                                                24,40,50,100,2
                                     00,300,501,600,1000-11
                                     00,300,501,600,1000-11
                                     00,2015,3966-3967          00,2015,3966-3967
```

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

Table 3 lists the per vPC interface parameters that are taken into account for type-1 consistency check.

Table 3. Per vPC Interface Type-1 Consistency Check

Parameter	Value
Port-channel LACP mode	ON, ACTIVE, PASSIVE
Link speed per port-channel	Speed in mbps
Duplex mode per port-channel	Half duplex or full duplex
Switchport mode per port-channel	Trunk or access Native VLAN
STP interface settings	Port type setting Loop Guard Root Guard
MST Simulate PVST	Enable or disable
MTU per port-channel	Maximum transmission Unit (MTU) value

show vpc consistency-parameters interface port-channel <id> command displays per vPC interface type 1 consistency parameters.

An example of output is shown below:

```
7K1# sh vpc consistency-parameters interface port-channel 80

    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-a,8050,  0-23-4-ee-be-a, 8050,
                              0, 0), (8000,         0, 0), (8000,
                              0-22-90-c2-8-3e, 1, 0, 0-22-90-c2-8-3e, 1, 0,
                              0)]                 0)]
mode                   1      active              active
Speed                  1      1000 Mb/s           1000 Mb/s
Duplex                 1      full                full
Port Mode              1      trunk                trunk
Native Vlan            1      1                    1
MTU                    1      1500                1500
Allowed VLANs          -      100,200,300         100,200,300
Local suspended VLANs  -      -                   -
```

Configuration Parameters That Should Be Identical (Type-2 Consistency Check)

When Type 2 inconsistency check is detected, moderate action or no action are taken:

For global configuration type 2 inconsistency check, all vPC member ports remain in up state and vPC systems trigger to protective actions.

For vPC interface configuration type 2 inconsistency check, the misconfigured vPC remains in up state. However, depending on the discrepancy type, vPC systems will trigger protective actions. The most typical one deals with allowed VLAN in vPC interface trunking configuration. In that case, vPC systems will disable from the vPC interface VLAN that do not match on both sides.

Table 4 lists type 2 consistency check parameters.

Some global configuration consistency check parameters appear in **sh vpc consistency-parameters global** (mentioned above). However, most of vPC interface parameters (for type 2 consistency check) do not appear in **sh vpc consistency-parameters interface port-channel** command.

If any of the parameters listed in Table 3 is not configured identically on both vPC peer devices, the inconsistent configuration can cause undesirable behavior in the traffic flow.

Table 4. Type-2 Consistency Parameters

Parameter	Description
MAC aging timers	MAC aging timer for a particular VLAN should be the same on both vPC peer devices.
Static MAC entries	Static MAC entries in a particular VLAN should be applied on both vPC peer devices.
VLAN interface (switch virtual interface [SVI])	Each peer device must have a VLAN interface configured for the same VLAN on both ends, and this VLAN interface must be in the same operational state.
Access Control List (ACL) configurations and parameters	ACL configurations should be identical on both vPC peer devices.
Quality of Service (QoS) configuration and parameters	QoS configuration should be identical on both vPC peer devices.
Spanning Tree Protocol interface settings	Bridge Protocol Data Unit (BPDU) filter Link type (auto, point-to-point, shared) Cost Port-priority STP interface settings should be identical on both vPC peer devices.
VLAN database	You must create all VLANs on both the primary and secondary vPC peer devices, or the VLAN will be suspended. Those VLANs configured on only one peer device do not pass traffic using the vPC or vPC peer-link.
Port security	Network Access Control (NAC) Dynamic ARP Inspection (DAI) IP source guard (IPSG) Port security settings should be identical on both vPC peer devices.
Cisco TrustSec	Cisco TrustSec configuration should be identical on both vPC peer devices.
Dynamic Host Configuration Protocol (DHCP) snooping	DHCP snooping configuration should be identical on both vPC peer devices.
Internet Group Management Protocol (IGMP) snooping	IGMP snooping configuration should be identical on both vPC peer devices.
Hot Standby Router Protocol (HSRP)	HSRP configuration should be identical on both vPC peer devices.

Parameter	Description
Protocol Independent Multicast (PIM)	PIM configuration should be identical on both vPC peer devices.
Gateway Load-Balancing Protocol (GLBP)	GLBP configuration should be identical on both vPC peer devices.
All routing protocol configurations	Routing configuration should be consistent on both vPC peer devices.

General Recommendation:

To help ensure that all the configuration parameters are compatible, we recommend that you display the configurations for each vPC peer device once you configure the vPC.

Building a vPC Domain: Guidelines and Restrictions

To build a vPC domain, use the following configuration guidelines:

- You must enable feature vPC (conf t; feature vpc) before you can start configuring a vPC domain.
- You must configure peer-keepalive link before peer-link in order for vPC system to come up.
- You must configure both vPC peer devices; the configuration is not sent from one device to the other.
- To configure double-sided vPC topology, you must assign a unique vPC domain ID for each respective vPC layer.
- To use vPC in a DCI topology, you must assign a unique vPC domain ID for each respective data center.
- Check that the necessary configuration parameters are consistent on both sides of the vPC peer-link.
- We recommend that you activate the LACP feature and configure vPC member ports with LACP mode set to ACTIVE.
- All ports for a given vPC peer must be in the same VDC.
- Only Layer 2 port channels (switchport mode trunk or switchport mode access) can be configured on vPC member ports.
- PIM SM (Sparse Mode) is fully interoperable with vPC. The software does not support PIM BiDIR or PIM SSM (Source Specific Multicast) with vPC.
- The software does not support DAI (Dynamic ARP Inspection) or IPSG (IP Source Guard) in a vPC environment.
- DHCP relay and DHCP snooping are supported with vPC.
- The software does not support Cisco Fabric Services regions with vPC.
- Port security is not supported on vPC member ports.
- Configure a separate Layer 3 link for routing from the vPC peer device (backup routing path), rather than using vPC peer-link and SVI for this purpose.
- We recommend that you create an additional Layer 2 trunk port-channel as an interswitch link to transport non-vPC VLAN traffic.

Note: When using vPC, it is a best practice to use default timers for HSRP (Hot Standby Router Protocol), VRRP (Virtual Router Redundancy Protocol) and PIM (Protocol Independent Multicast) configurations.

There is no gain in regards to network convergence times when using aggressive timers in vPC configurations.

Best Practices for vPC Components Configuration

The following sections provide recommendations to configure the different vPC components: vPC VLAN, vPC Peer-keepalive link, vPC peer-link, vPC member ports.

Specific consideration for vPC peer-link using only 1 10G module in the chassis is also discussed in this chapter.

Recommendation for vPC VLAN Configuration

vPC VLAN is a VLAN that is allowed on vPC member port and vPC peer-link.

The first step to make a switch working is to create the VLAN database. Enter the following command in global configuration mode:

```
N7k(config) # vlan <vlan-id range>
```

When configuring large number of VLAN in vPC environment, it is recommended to configure the VLANs using range command, instead of individually configuring one VLAN at a time.

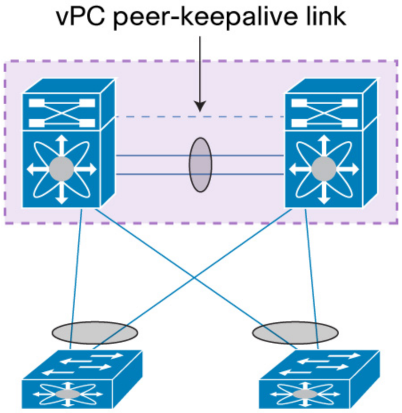
If you need to name the different VLAN, create first all the VLAN using range command. Exit from global configuration mode to make effective VLAN creation. Then name each VLAN as needed.

Strong Recommendation:
When configuring large number of VLAN in vPC environment, it is recommended to configure the VLANs using range command, instead of individually configuring one VLAN at a time.

Recommendations for vPC Peer-Keepalive Link Configuration

vPC peer-keepalive link is a Layer 3 link that joins one vPC peer device to the other vPC peer device, as illustrated in Figure 14.

Figure 14. vPC Peer-Keepalive Link



The vPC peer-keepalive link carries periodic heartbeat between vPC peer devices. It is used at the boot up of the vPC systems to guarantee both peer devices are up before forming vPC domain and also when vPC peer-link fails to down state; in the latter case, vPC peer-keepalive link is leveraged to detect split brain scenario (both vPC peer devices are active-active) [when vPC peer-link is down, there is no more real time synchronization between the 2 peer devices so vPC systems must react to this active-active situation; this is done by shutting down vPC member ports on secondary peer device].

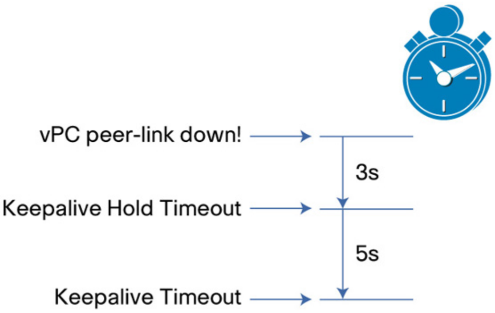
In term of data structure, a vPC peer-keepalive message is a User Datagram Protocol (UDP) message on port 3200 that is 96 bytes long, with a 32-byte payload. Keepalive messages can be captured and displayed using the onboard Wireshark Toolkit.

Table 5 lists default values for vPC peer-keepalive link timers and Figure 15 illustrates the vPC timers concept.

Table 5. Default Values for vPC Peer-Keepalive Links

Timer	Default value
Keepalive interval	1 seconds
Keepalive hold timeout (on vPC peer-link loss)	3 seconds
Keepalive timeout	5 seconds

Figure 15. vPC Timers Concept



Keepalive Hold Timeout

This timer gets started once the vPC peer-link goes to down state. During this time period, the secondary vPC peer device will ignore any peer-keepalive hello messages (or the lack of). This is to assure that network convergence can happen before any action is taken.

Keepalive Timeout

During this time period, the secondary vPC peer device will look for vPC peer-keepalive hello messages from the primary vPC peer device. If a single hello is received, the secondary vPC peer concludes that there must be a dual active scenario and therefore will disable all its vPC member ports (that is, all port-channels that carry the keyword **vpc**).

Command line configuration to modify vPC timers is (under vPC domain configuration context):

N7k(config-vpc-domain)# peer-keepalive destination ipaddress [**source** ipaddress | **hold-timeout** secs | **interval** msecs {**timeout** secs}]

The **show vpc peer-keepalive** command displays all information about peer-keepalive link, as follows:

```
7K1# sh vpc peer-keepalive
```

```
vPC keep-alive status                : peer is alive
--Peer is alive for                  : (22) seconds, (255) msec
--Send status                        : Success
--Last send at                       : 2011.06.07 15:24:28 339 ms
--Sent on interface                  : Eth1/24
--Receive status                     : Success
--Last receive at                    : 2011.06.07 15:24:27 597 ms
--Received on interface                : Eth1/24
--Last update from peer              : (0) seconds, (857) msec

vPC Keep-alive parameters
--Destination                        : 192.168.100.2
--Keepalive interval                 : 1000 msec
--Keepalive timeout                  : 5 seconds
--Keepalive hold timeout             : 3 seconds
--Keepalive vrf                      : peerkeepalive
--Keepalive udp port                 : 3200
--Keepalive tos                      : 192
```

Strong Recommendations:

When building a vPC peer-keepalive link, use the following in descending order of preference:

1. Dedicated link(s) (1-Gigabit Ethernet port is enough) configured as L3. Port-channel with 2 X 1G port is even better.
2. Mgmt0 interface (along with management traffic)
3. As a last resort, route the peer-keepalive link over the Layer 3 infrastructure

WARNING: Do not configure vPC peer-keepalive link on top of vPC peer-link; peer-keepalive messages must not be carried over vPC peer-link to avoid fate sharing in case peer-link goes down.

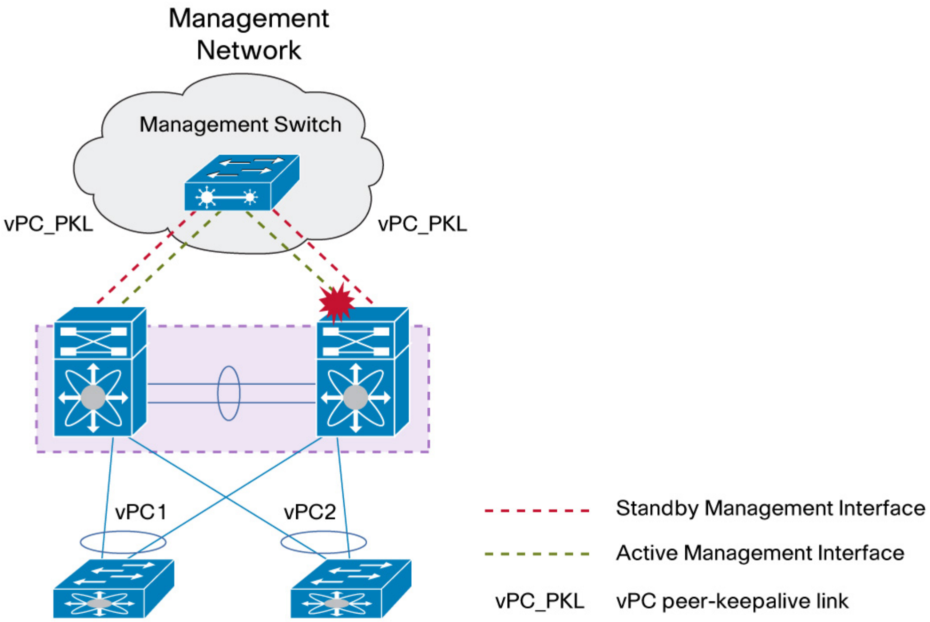
Note: If you are using a pure Cisco Nexus F1 Series system or VDC (that is, only F1 line cards used in the chassis or only F1 ports in the VDC), the peer-keepalive link can be formed with mgmt0 interface or 10-Gigabit Ethernet front panel port. In the latter case, use the **management** command under the SVI to enable it for inband management (otherwise, the SVI is brought down because no M1 modules exist in the system or VDC).

vPC Peer-Keepalive Link Using mgmt0 Cisco Nexus 7000 Series Pairs with Dual Supervisors Each

When using dual supervisors and mgmt0 interfaces to carry the vPC peer-keepalive link, do not connect mgmt0 ports in back-to-back mode across the two switches (i.e mgmt0 on supervisor 1 of peer device 1 connected directly to mgmt0 on supervisor 1 of peer device 2 and so on). Reason is that active supervisor owns control of mgmt0 port and in case of supervisor switchover, keep-alive connectivity may be broken (active supervisor on peer device 1 sending keep-alive to standby supervisor on peer device 2).

Best practice is to insert a L2 switch between the different supervisors to avoid this kind of situation as depicted in figure 16.

Figure 16. vPC Peer-Keepalive Link Using mgmt0 and Dual Supervisors



Strong Recommendation:
When using mgmt0 port for vPC peer-keepalive link in a dual supervisor configuration, always use an intermediate L2 switch to interconnect the different supervisors together.

vPC Peer-Keepalive Link and VRF

By default, vPC peer-keepalive is placed in VRF management.

If needed, vPC peer-keepalive can be placed in another VRF using the following command (under vPC domain configuration context):

N7k(config-vpc-domain)# peer-keepalive destination <destination IP> source <source IP> **vrf <VRF name>**

General Recommendation:
Create a dedicated VRF for vPC peer-keepalive link (for instance VRF PKL-VRF)

Recommendations for vPC Peer-Link Configuration

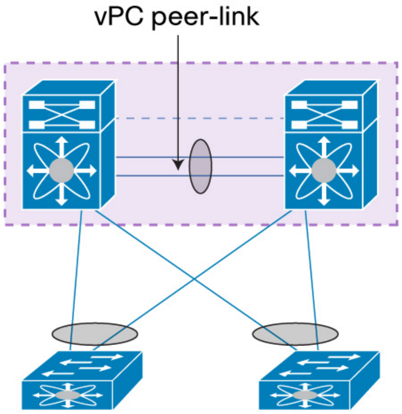
The vPC peer-link is a standard 802.1Q trunk that can perform the following actions:

- Carry vPC and non-vPC VLANs.
- Carry Cisco Fabric Services messages that are tagged with CoS=4 for reliable communication.
- Carry flooded traffic from the other vPC peer device.
- Carry STP BPDUs, HSRP hello messages, and IGMP updates.

It is at the vPC peer-link level that we implement vPC loop avoidance mechanism (data plane layer). This is done in hardware, without any software sollicitation.

Figure 17 shows vPC peer-link component in a vPC domain.

Figure 17. vPC Peer-Link



Strong Recommendations:
When building a vPC peer-link, follow these guidelines:

- Ensure that member ports are 10-Gigabit Ethernet interfaces.
- Use a minimum of two 10-Gigabit Ethernet ports. vPC peer-link member ports can be scaled up to line card capacity in regards to port-channel (M1 line card supports up to 8 members ports while F1 and F2 support up to 16 member ports).
- Use at least 2 different line cards to increase high availability of peer-link.
- Use dedicated 10-Gigabit Ethernet ports with M1 32 10G line card. Do not use shared mode ports.
- Split vPC and non-vPC VLANs on different interswitch port channels (use vPC peer-link to carry vPC VLAN and the other interswitch port-channel to carry non-vPC VLAN).
- Do not insert any device between vPC peers. A peer-link is a point-to-point link.

vPC peer-link is supported on all shipping 10G line card. It is not supported on any 1G line card nor on any FEX ports (including the 2232 model which has 10G front panel ports).

Table 6 list all line cards that are able to support vPC peer-link

Table 6. Supported line cards for vPC peer-links.

Line Card Part Number	Line Card Description
N7K-M132XP-12 **N7K-M132XP-12L**	32 10-Gigabit Ethernet port M1- Series
N7K-M108X2-12L	8 10-Gigabit Ethernet port M1- Series
N7K-F132XP-15	32 1/10-Gigabit Ethernet port F1 Series
N7K-F248XP-25	48 1/10-Gigabit Ethernet port F2 Series

vPC Peer-link can be formed with M1 ports on both sides (M1 to M1), or F1 ports on both sides (F1 to F1), or F2 ports on both sides (F2 to F2). It is mandatory that both sides of vPC peer-link are strictly identical.

As a consequence, it is not possible to mix F1 port on 1 side and M1 port on the other side (for instance).

The M132XP and M108X2 (that is, a 32 10-Gigabit Ethernet line card and 8 10-Gigabit Ethernet line card, respectively) can form a port-channel together. However, be aware of the requirement that ports on M132XP can form a port-channel with ports on M108X2 only if the port is configured in dedicated mode.

This is consistent with vPC peer-link recommendations: vPC peer-link member ports should have full 10-Gigabit Ethernet capability - in other words, no oversubscription.

Supported and unsupported configurations are depicted in figure 18 and 19.

Figure 18. Supported Configurations for vPC Peer-Links (Both Sides Must Be of Same Port Type)

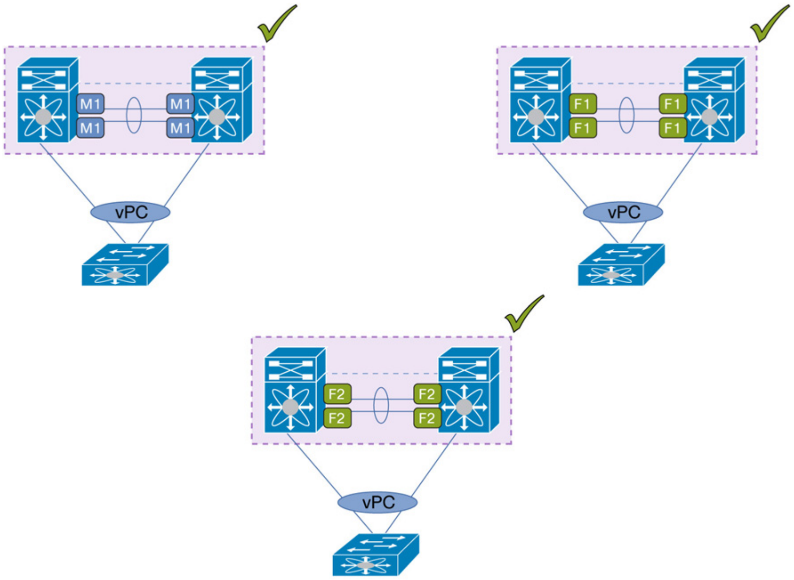
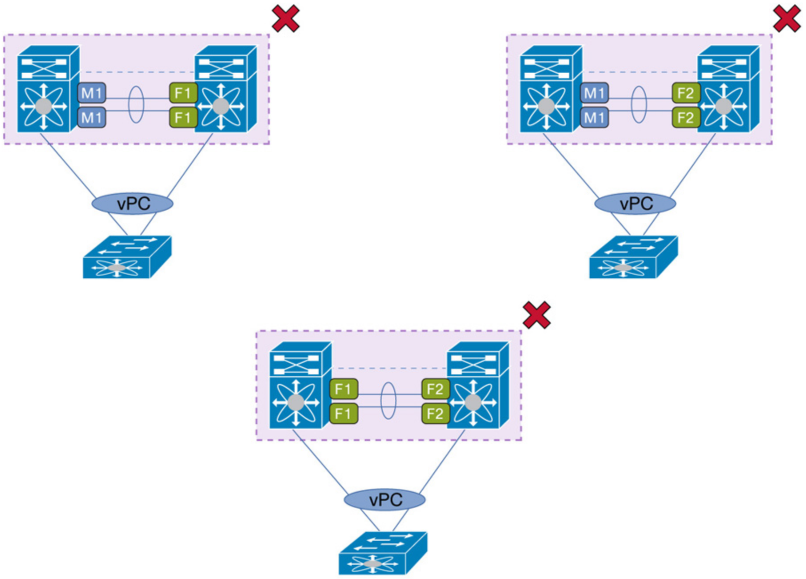


Figure 19. Unsupported Configurations for vPC Peer-Links (Both Sides Must Be of Same Port Type)



It is not possible to mix different port type inside the same vPC peer-link side (for example bundling M1 port and F1 port on same side of the vPC peer-link) because each port type have different hardware characteristics (in terms of forwarding, queuing, and security).

This rule does not only apply to vPC peer-link; this is a generic statement for port-channel.

Note: In case of mixed chassis mode (that is, with an M1 and F1 ports in the same system or same VDC) and vPC peer-link on F1 ports, be careful if you use the vPC peer-gateway feature and backup routing path capability.

Starting with Cisco NX-OS Release 5.1.3, a knob is available to exclude specific VLANs from the peer-gateway. These VLANs are typically used for backup routing paths. The command is:

```
    N7k(config-vpc-domain)# peer-gateway exclude-vlan <VLAN list>
```

The knob is not useful in the other configurations of vPC systems and vPC peer-link.

More details are provided in the peer-gateway section of this document.

For unicast traffic, vPC peer devices always use local forwarding preference using vPC member port.

vPC peer-links is usually not loaded with unicast traffic unless vPC member port fail on 1 side of vPC domain.

For multicast traffic, a copy of the stream is replicated on vPC peer-link (except when vPC peer-link is built with F2 ports as the line card do not support dual DR [Designated Router] in multicast environment).

This type of traffic needs to be taken into account when dimensioning the vPC peer-link.

To view statistics for vPC peer-link, use the command **show vpc statistics peer-link**.

One important aspect related to vPC peer-link is VLAN pruning. For a vPC VLAN to become operational, it must be defined on vPC peer-link using the command:

```
N7k(config)# interface port-channel 1          (int Po1: vPC peer-link)
   N7k(configif)# switchport trunk allowed vlan <VLAN-id list>
```

Of course, vPC VLAN has previously been allowed (i.e pruned) on vPC member port.

Required Recommendation:
Always perform VLAN pruning on vPC peer-link with allowed list of vPC VLAN. vPC VLAN must have been pruned on vPC member port previously.

vPC Systems Behavior When a vPC Peer-Link Goes Down

When vPC peer-link fails down and vPC peer-keepalive link is still up, the vPC secondary peer device performs the following operations:

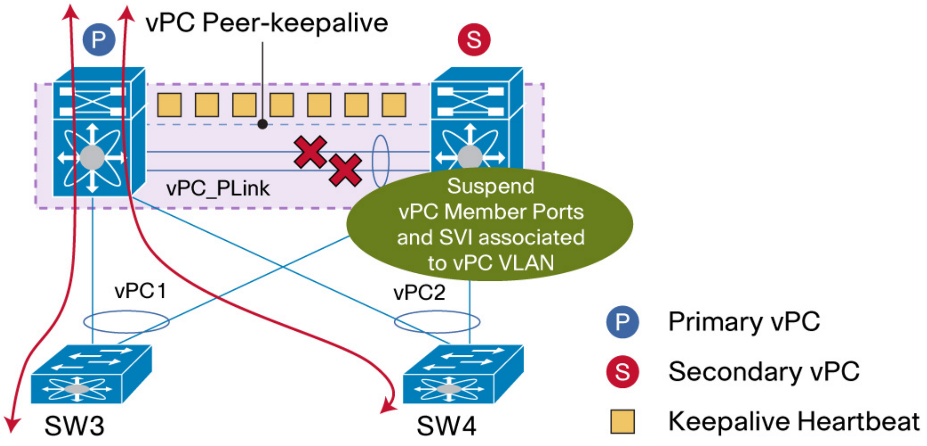
- Suspends its vPC member ports
- Shuts down the SVI associated to the vPC VLAN

This protective behavior from vPC allows to redirect all southbound and northbound traffic to primary peer device.

Note that when vPC peer-link is down, both vPC peer devices cannot synchronize each other anymore so designed protection mechanism leads to isolate one of the peer device (in occurrence the secondary peer device) from the data path.

Figure 20 illustrates what happens when a vPC peer-link fails down

Figure 20. vPC Peer-Link Down - Behavior on Secondary Peer Device



If orphan ports are connected to vPC secondary peer device, they become isolated once peer-link is down. In a VXLAN + vPC implementation, when a vPC peer-link shuts down, all Layer 2 or Layer 3 orphan receivers behind the non-forwarder (shut down vPC peer-link) will not receive any traffic.

To maintain Layer 3 connectivity to these orphan ports, a command is available to prevent the SVI (associated to vPC VLAN) from being shut down: **dual-active exclude interface-vlan**.

Use this command to keep desired SVI in UP state when vPC peer-link goes down:

```
N7k(config-vpc-domain)# dual-active exclude interface-vlan <VLAN list>
```

VLAN listed in the knob must be associated to vPC VLAN. Using a non-vPC VLAN has no effect since SVI associated to these VLAN are not shut down when vPC peer-link goes down.

Recommendations for vPC Peer-Link Configuration with Systems Containing Only One M1 10-Gbps Module

Some Cisco Nexus 7000 Series Switch configurations have only one M1 10-Gbps module and several 1-Gbps modules. Problems may occur if these Switches are defined as Layer 2/Layer 3 boundary (that is, same 10-Gbps line card used for both Layer 3 uplinks connectivity and for vPC peer-link).

For this type of configuration, it is necessary to use vPC Object Tracking feature (available since NX-OS 4.2)

vPC Object Tracking

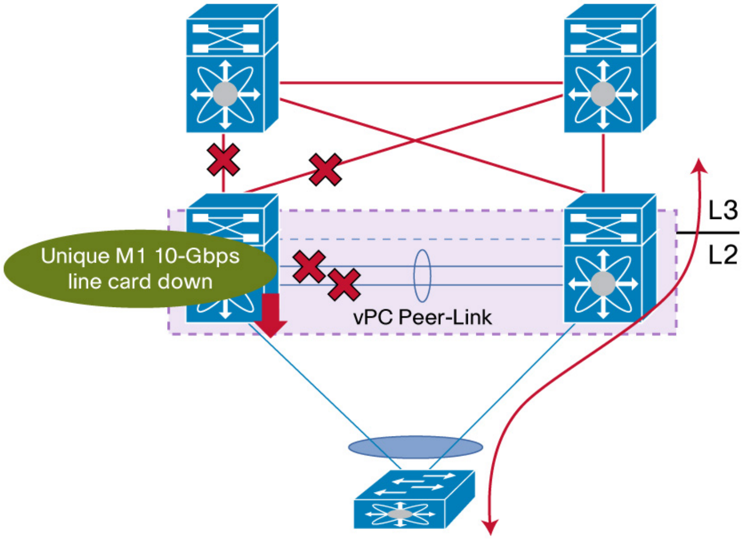
vPC object tracking is used to track failure of all modules on a Cisco Nexus 7000/7700 Series switch on wich peer-link and uplinks are hosted. It is also used when the L3 core links and vPC peer-link interfaces are localised on the same module and it fails.

Without the vPC object tracking feature enabled, if the module/modules fails on the vPC primary device that hosts the peer-link and ulinks, it will lead to a complete traffic blackhole even though the vPC secondary device is up and running.

This happens because the module hosting the peer-link fails and the peer-link goes down and vPC secondary device suspends all vPC (vPC loop prevention). The vPC’s on the vPC primary will be still up but as the module/modules hosting the uplinks have failed, therefore the uplinks will be down and traffic (south to north) will be dropped.

Figure 21 illustrates the effect of the vPC Object Tracking feature.

Figure 21. vPC Object Tracking Feature - Behavior When vPC Peer-Link Fails Down



The vPC Object Tracking feature suspends the vPCs on the impaired device so that traffic can be diverted over the remaining vPC peer.

To use vPC object tracking, track both Peer-link interfaces and L3 core interfaces as a list of Boolean objects. Note that the Boolean AND operation is not supported with vPC object tracking.

The vPC object tracking configuration must be applied on both vPC peer devices.

Following configuration sample shows the different commands needed to activate vPC object tracking:

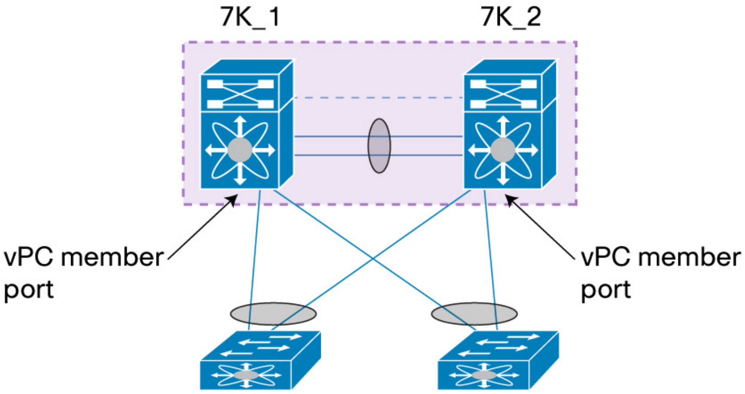
```
! Track the vpc peer link
track 1 interface port-channel11 line-protocol
! Track the uplinks to the core
track 2 interface Ethernet1/1 line-protocol
track 3 interface Ethernet1/2 line-protocol

! Combine all tracked objects into one.
! “OR” means if ALL objects are down, this object will go down
! ==> we have lost all connectivity to the L3 core and the peer link
track 10 list boolean OR
  object 1
  object 2
  object 3

! If object 10 goes down on the primary vPC peer,
! system will switch over to other vPC peer and disable all local vPCs
vpc domain 1
  track 10
```

Strong Recommendation:
For vPC peer device with a single line card, use vPC object tracking feature.

Recommendations for vPC Member Port Configuration

By definition, a vPC member port is a port-channel member of a vPC, as illustrated in Figure 22.

Figure 22. vPC Member Port



Port-channel defined as vPC member port always contains the keywords **vpc <vpc id>.**

Port-channel member port contain from 1 member port up to the hardware limit of the line card (M1 series line card support up to 8 member ports while F1 and F2 series line card support up to 16 member ports).

A Layer 2 port-channel only is supported with vPC (no Layer 3). The port-channel can be configured in access or trunk switchport mode. Any VLAN allowed in the vPC member port is by definition called vPC VLAN. The vPC VLAN must then be allowed on vPC peer-link.

Note: Whenever a vPC VLAN is defined on vPC member port, it MUST be defined also on vPC peer-link. Not defining a vPC VLAN on vPC peer-link will will make the VLAN not operational.

Required Recommendation:

Any vPC VLAN allowed on vPC member port MUST be allowed on vPC peer-link.

Below is a sample configuration of vPC member port:

```
7K1:
interface port-channel201
  switchport mode trunk
  switchport trunk native vlan 100
  switchport trunk allowed vlan 100-105
  vpc 201
```

```
7K2:
interface port-channel201
  switchport mode trunk
  switchport trunk native vlan 100
  switchport trunk allowed vlan 100-105
  vpc 201
```

Use the following recommendations to build properly vPC member port:

Strong Recommendations:

- The configuration of vPC member port must match on both vPC peer devices.
- If there is an inconsistency, a VLAN or the entire port channel may suspend (depending on type-1 or type-2 consistency check for the vPC member port). For instance, a MTU mismatch will suspend the vPC member port.
- Use same vPC ID as port-channel ID for ease of configuration, monitoring, and troubleshooting.
- With the M1 Series line card: There can be up to eight active ports bundled in the same vPC member port (allowing as a result a 16-way port channel to be built for the whole vPC).
- With F1 and F2 Series line card: There can be up to 16 active ports bundled in the same vPC member port (allowing as a result a 32-way port channel to be built for the whole vPC).
- Do not mix different port types (M1 port, F1 port or F2 port) in the same vPC member port. This is not allowed by the software.
- Both sides of the vPC member ports (i.e vPC member port on 7K1 and vPC member port on 7K2) must be of same port type (M1 port or F1 port or F2 port).

A vPC can be formed with same port type on both sides (i.e on both vPC peer devices):

If vPC member port uses M1 ports type on peer device 1, then vPC member port on peer device 2 must use M1 ports type as well. Same statement applies for F1 and F2 ports type.

It is not possible to mix different port types on same vPC member port (also true for port-channel) because these ports have different hardware characteristics (in terms of forwarding, queuing, and security).

Supported and unsupported vPC member port configurations are depicted in Figures 23 and 24:

Figure 23. Supported Configurations for vPC Member Ports

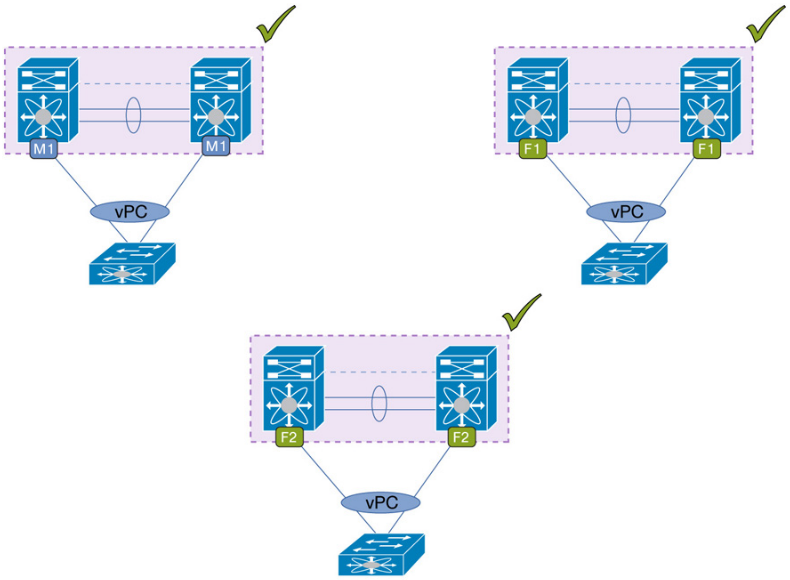
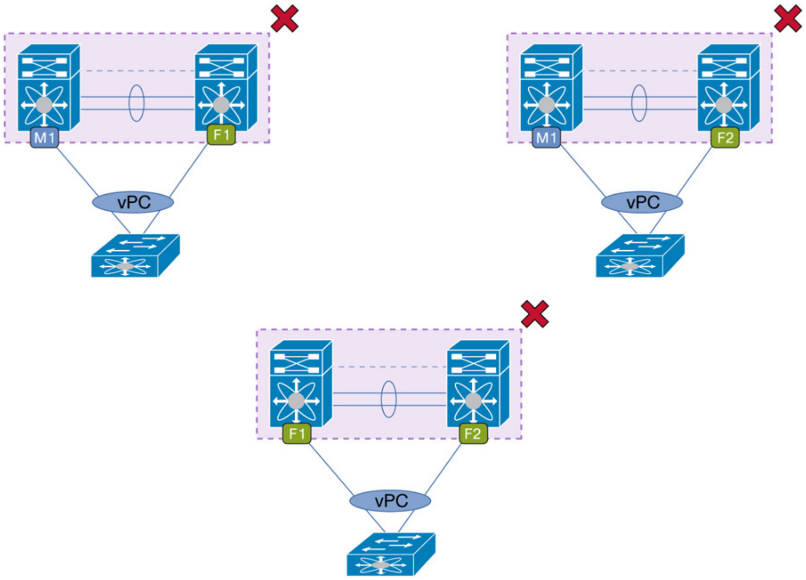


Figure 24. Unsupported Configurations for vPC Member Ports



Note: M132XP and M108X2 (32 10-Gigabit Ethernet line cards and 8 10-Gigabit Ethernet line cards, respectively) can form a port-channel together and then can co-exist in the same vPC member port. However, be aware of the requirement that ports on M132XP can form a port-channel or vPC member port with ports on M108X2 only if M132XP ports are configured in dedicated mode.

Best Practices for vPC in Mixed Chassis Mode (M1/F1 Ports in Same System or VDC)

Mixed chassis mode is a system where both M1 ports and F1 ports are used simultaneously.

M1 Series line cards provide scalable Layer 2 and Layer 3 capabilities. F1 Series line cards provide high-density cost-effective Layer 2 10-Gigabit Ethernet connectivity. Interoperability between M1 and F1 ports are provided by L3 internal proxy routing where M1 ports are used for L3 proxy when traffic entering a F1 port need to be routed (L3 traffic for inter VLAN routing or traffic going outside of data center). M1 line card typically host the interface VLAN (i.e SVI - Switch Virtual Interface) on behalf of F1 line card.

Layer 3 Internal Proxy Routing

L3 internal proxy routing present the following characteristics:

- For Unicast traffic - Proxy Layer 3 routing for F1 modules can be spread among any or all M1 interfaces in the VDC
- For Multicast traffic - Proxy Layer 3 replication for F1 modules can be spread among all M1 replication engines.

L3 internal proxy routing is enabled by default when the system or VDC is configured in mixed chassis mode. The Cisco Nexus 7000 Series automatically makes all M1 modules in VDC available for proxy Layer 3 forwarding.

All M1 front-panel ports or port groups in the system or VDC become part this L3 internal proxy routing.

The user can modify the proxy routing configuration using the command: **hardware proxy layer-3 forwarding**. The knob allows to include or exclude specific front-panel ports or port group from participating to L3 internal proxy routing.

Note: Starting with Cisco NX-OS Release 5.1(2), the maximum number of proxy forwarders that can be used to proxy Layer 3 traffic (traffic that ingresses from an F1 Series module) has increased from 16 to 128. The output of the **show hardware proxy layer-3 detail** command displays up to 128 Layer 3 forwarders.

```
7K1# sh hardware proxy layer-3 detail

Global Information:
        F1 Modules:     Count: 1           Slot: 8
        M1 Modules:    Count: 4            Slot: 1-3,9

        Replication Rebalance Mode:               Manual
        Number of proxy layer-3 forwarders:       16
        Number of proxy layer-3 replicators:    12

Forwarder Interfaces                     Status     Reason
------------------------------------------------------------------------------
Eth1/1-12                                  up         SUCCESS
Eth1/13-24                                 up         SUCCESS
Eth1/25-26                                 up         SUCCESS
```

```
Eth1/37-48                                 up         SUCCESS
<snip>

Replicator Interfaces                    #Interface-Vlan   Interface-Vlan
------------------------------------------------------------------------------
Eth1/1-24                                  7                 1,40,200,400-401,
                                                               1000,3000
Eth1/25-48                                 6                 3030,3039,3049,3142,
                                                               3966-3967
```

sh hardware proxy layer-3 counters brief gives some statistics data about the usage of L3 forwarders.

```
7K1# sh hardware proxy layer-3 counters brief

Summary:
--------
Proxy packets sent by all F-series module:
-----------------------------------------------------------------------
Router Interfaces             Tx-Pkts             Tx-Rate (pkts/sec approx.)
-----------------------------------------------------------------------
Eth1/1-12                         34453                  1443
Eth1/13-24                       323242                234
Eth1/25-26                       345435                213
Eth1/37-4                         345341                0234
<snip>
==========================================================
Total                                2341434             3453
==========================================================
```

vPC in Mixed Chassis Mode

Mixing M1 ports and F1 ports in an aggregation switch (called mixed chassis mode) provides several benefits:

- Bridged traffic remains in F1 ports
- Routed traffic coming from F1 ports are proxied to M1 ports, providing F1 port the capability to support any type of traffic.

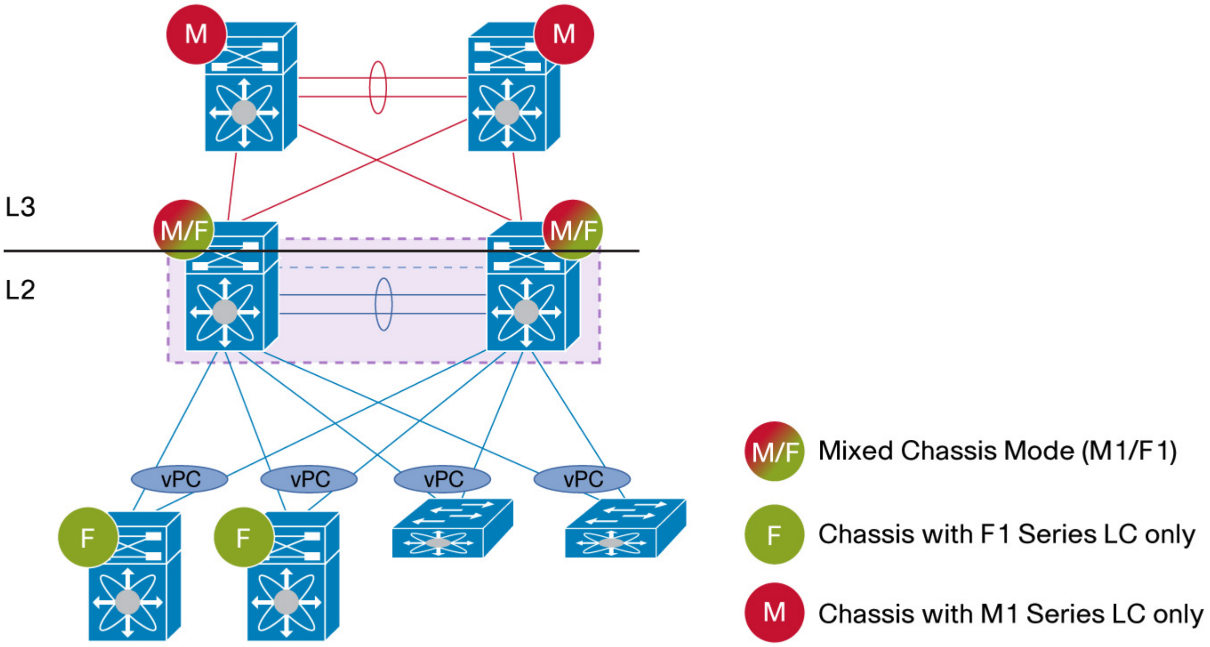
vPC is fully supported in a mixed chassis mode configuration, for both types of vPC peer-link (i.e vPC peer-link formed with F1 ports or with M1 ports)

Note: Configuring a VDC or system in mixed chassis mode (M1/F1) is a prerequisite when vPC domain must assume the role of L2/L3 boundary (i.e hosting interface VLAN and HSRP or VRRP feature).

In a situation where vPC domain is needed only as L2 network, then F1 ports are largely self sufficient; there is no need for M1 port as L3 function is not required.

Figure 25 shows how a vPC system in Mixed Chassis Mode configuration typically fits in the data center.

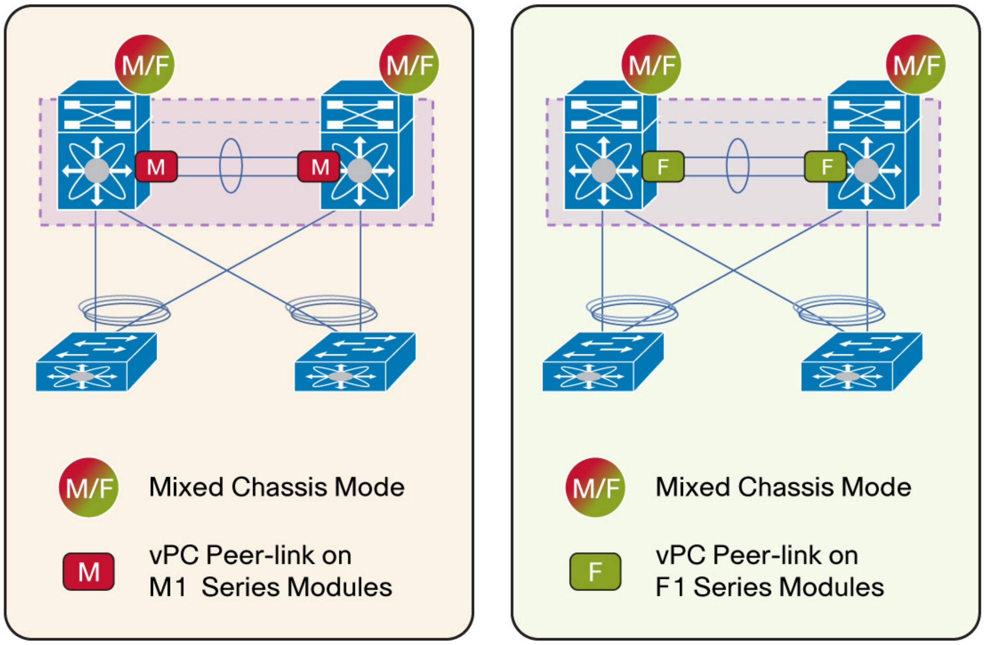
Figure 25. vPC System in Mixed Chassis Mode Configuration


M
M
L3
M/F
M/F
L2
vPC
vPC
vPC
vPC
F
F
M/F Mixed Chassis Mode (M1/F1)
F Chassis with F1 Series LC only
M Chassis with M1 Series LC only

2 different topologies can be implemented with vPC in mixed chassis mode, depending on the nature of vPC peer-link: peer-link on M1 ports and peer-link on F1 ports.

Figure 26 shows the 2 possible configurations.

Figure 26. vPC in Mixed Chassis Mode - Peer-Link on F1 Ports or M1 Ports



vPC system in mixed chassis mode with peer-link on F1 ports present the following characteristics:

- Total number of MAC addresses supported is 16K (capacity of one forwarding engine [i.e switch on chip] on F1 series line card)
- M1 ports are used only for L3 uplinks
- F1 ports used for vPC member ports (can use M1 ports as well if needed)
- Need to use **peer-gateway exclude-vlan <VLAN list>** to exclude VLAN that belong to backup routing path (this command only applied to vPC system in mixed chassis mode with vPC peer-link on F1).

Attention: When you deploy a vPC system in a mixed chassis mode with peer-link on F1 ports, it is recommended that you disable the vPC peer gateway. If the peer gateway is configured, all the packets forwarded over the vPC peer link will be punted to the CPU, resulting in high CPU usage. Factors that affect CPU usage in this scenario are the layer 2 vPC peer gateway rate limiter (which is defaulted to 5000 pps) settings and the default control plane policing (CoPP) settings.

vPC system in mixed chassis mode with peer-link on M1 ports present the following characteristics:

- Total number of MAC addresses supported is 128K (capacity of forwarding engine on M1 series line card)
- M1 port are used for L3 uplinks and vPC peer-link
- F1 ports used for vPC member ports (can use M1 ports as well if needed)
- No need to use **peer-gateway exclude-vlan <VLAN list>** knob

Be careful when you deploy a vPC system in mixed chassis mode with peer-link on F1 ports and you need to use peer-gateway function in addition to backup routing path.

By default, backup routing path configured over vPC peer-link are processed in software with CPU intervention when peer-gateway knob is enabled.

To force the traffic carried over backup routing path to be processed in hardware without any performace penalty, use **peer-gateway exclude-vlan <VLAN list>** knob.

This command (available beginning NX-OS Release 5.1.3) allows to disassociate desired VLAN from the peer-gateway mechanim.

Peer-gateway exclude-vlan is described later in this document (in the peer-gateway section).

Strong Recommendation:
For vPC system in mixed chassis mode (peer-link on F1 ports or on M1 ports), recommendation is to use at least 2 M1 series line card in the same system or VDC. Having 2 M1 series line cards provides higher resiliency for L3 internal proxy routing and for L3 features (L3 uplinks, interface VLAN or SVI and HSRP/VRRP feature).

vPC Mixed Chassis Mode with Peer-Link on F1 and Only One M1 Line Card

vPC system in mixed chassis mode with peer-link on F1 ports and only one M1 line card is not a recommended configuration (use 2 M1 line card to comply with the recommendation).

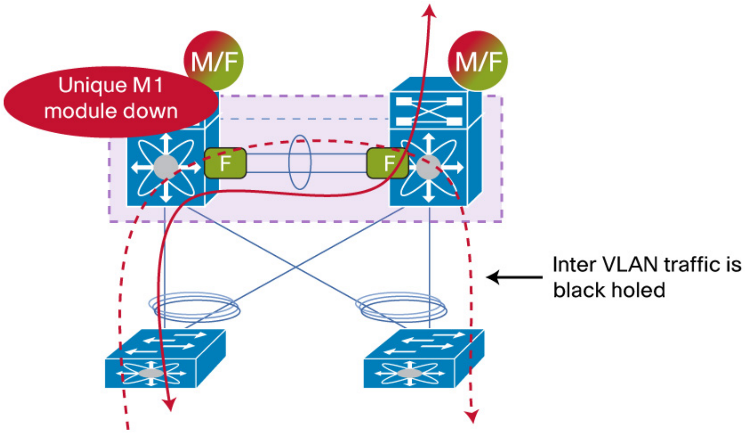
However, vPC works fine in this type of configuration. The M1 Series module is used for Layer 3 internal proxy routing and Layer 3 uplink connectivity. F1 modules are used for Layer 2 domain bridging.

Be cautious about vPC system behavior when the unique M1 module fails:

- Inter VLAN traffic are black holed because of vPC loop avoidance rule (alive M1 module process all packets with destination MAC equals HSRP or VRRP vMAC)
- L3 traffic (southbound or northbound traffic) are flowing seamlessly without any issue (all routed traffic being diverted to alive M1 module)

This behavior is represented in figure 27.

Figure 27. vPC Mixed Chassis Mode with Peer-Link on F1 and Only One M1 Line Card - Traffic Flows When vPC Peer-Link Is Down



Strong Recommendation:

For mixed chassis mode with a vPC peer-link on F1 ports, use at least two M1 line cards.

Having 2 M1 series line cards provides higher resiliency for L3 internal proxy routing and for L3 features (L3 uplinks, interface VLAN or SVI and HSRP/VRRP feature).

Best Practices for Attaching a Device to vPC Domain

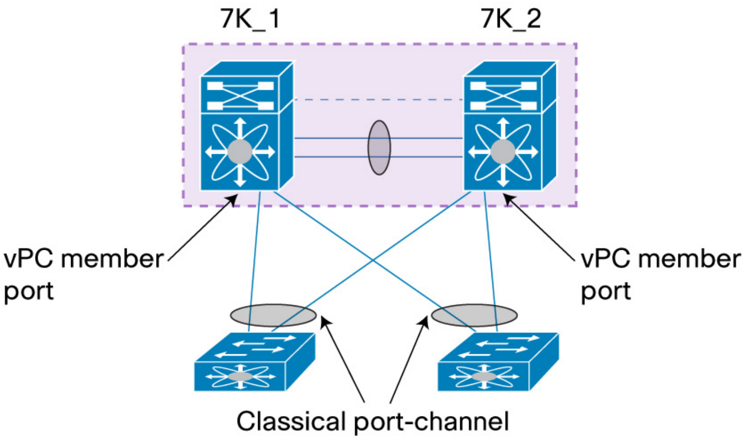
This section describes best practices for attaching an access device or an endpoint device to the vPC domain.

How to Attach Devices to a vPC Domain

Attaching a device to a vPC domain involves creating a Layer 2 port-channel from the access device to the 2 vPC peer devices. From the access device standpoint, this is a classical port-channel. From each vPC peer device standpoint, this is a vPC member port (i.e port-channel with keyword vPC).

Figure 28 depicts how classical port-channel and vPC member ports are differentiated in a vPC topology

Figure 28. Classical Port-Channel and vPC Member Port



Note: The Layer 3 port-channel is not supported with vPC technology.

Attaching to a vPC domain provides load-balancing capability across the different port-channel member ports.

Port-channel is considered as a single logical entity from spanning tree protocol (STP). As a consequence, adding or removing port-channel member ports do not create any topology change.

The access device can be of any type. It can be a switch, a server, a firewall, a load-balancer, a NAS (and so on). The requirements for access devices in order to properly connect to vPC domain are the following:

- Support of standard 802.3ad capability (LACP protocol)
- Support of static port-channels (channel-group mode ON)

The use of the Link Aggregation Control Protocol (LACP) is strongly recommended when available for improved failover convergence time and for misconfiguration protection. If not possible, use the manual bundling mechanism (channel-group mode ON).

Note: The Cisco Nexus 7000 Series does not support Port Aggregation Protocol (PAgP).

NEXUS 7000 supports different options for port-channel load-balancing hashing algorithm. This is configured globally in the default VDC and the command is **port-channel load-balance**.

Fields that can be used for the hashing algorithm are the following:

• Ip	IP
• ip-l4port	IP and L4 port
• ip-l4port-vlan	IP, L4 port and VLAN
• ip-vlan	IP and VLAN
• l4port	L4 port
• mac	MAC

They can be selected as source fields only or destination fields only or source and destination fields.

General Recommendations:

- Use LACP when available for graceful failover and misconfiguration protection.
- LACP mode active-active (on both sides of the port-channel) is the recommended configuration. Otherwise use LACP mode active-passive. Port-channel in mode active-active initiates more quickly than port-channel in mode active-passive.
- If access device does not support LACP, use manual bundling mechanism (channel-group mode ON).
- If the downstream access switch is a Cisco Nexus device, enable the LACP graceful-convergence option (this option is ON by default).
- If the downstream access switch is a not a Cisco Nexus device, disable the LACP graceful-convergence option.
- Use source-destination IP, L4 port and VLAN as fields for port-channel load-balancing hashing algorithm. This improves fair usage of all member ports forming the port-channel.

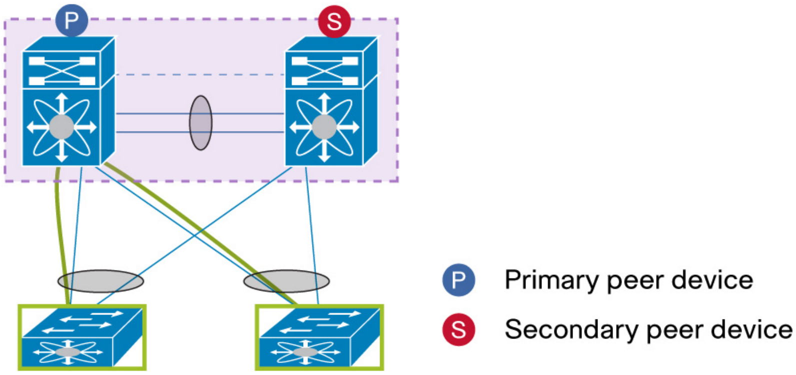
The following sections list all the possible ways to connect an access device or endpoint to vPC domain.

Access Device Dual-Attached to vPC Domain

Access device attached to vPC domain using half of the port-channel member ports to connect to peer device 1 and the other half member ports to peer device 2 is the natural way to use vPC technology (as represented on figure 29).

This is the recommendation number 1 when dealing with connecting to vPC domain.

Figure 29. Access Devices Dual-Attached to a vPC Domain



Dual-attaching access device to vPC domain using port-channel provide the following benefits:

- It ensures minimal disruption in case of a peer-link failover and provides consistent behavior in vPC dual-active scenarios.
- It ensures full redundant active-active paths through a vPC.

“Dual-attaching access device” is also referred as “vPC-attaching access device” in the document.

Strong Recommendation:
When possible, always dual-attach access device to vPC domain using port-channel

Single-Sided vPC with 16-Way Port-Channel

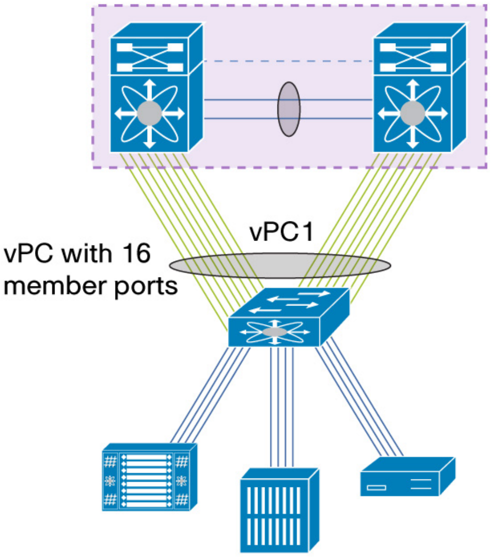
Single-sided vPC with 16-way port-channel is a specific implementation of access device dual-attached to vPC domain. This type of design leverages the maximum port-channel capability of the access device.

Cisco Nexus 5000 with 16-port port-channel support was introduced in Cisco NX-OS Software Release 4.1(3)N1(1a). This is the foundation for Single-sided vPC with 16-way port-channel topology (Figure 30 illustrates this configuration).

The single-sided vPC with 16-way port-channel topology has the following characteristics:

- Access device supports port-channel with 16 active member ports.
- Each vPC peer device has vPC member ports composed with eight active links, and the pair (i.e vPC) has 16 active load-balanced links to the downstream device
- vPC member ports can be selected from Cisco Nexus M1 or F1 or F2 Series line cards
- for north to south traffic, vPC peer device always perform **local** load-balancing out of its vPC member port unless the unique path to southbound device is only through vPC peer-link.

Figure 30. Single-Sided vPC with 16-Way Port-Channel



Double-Sided vPC with 32-Way Port-Channel

Double-sided vPC with 32-way port-channel is the second specific implementation of access device dual-attached to vPC domain. This type of design leverages the maximum port-channel capability of the access devices forming a vPC domain.

Double-sided vPC is a configuration where 2 access layer switches forming vPC domain are connected to 2 aggregation layer switches forming another vPC domain through a big fat vPC (up to 32 member ports).

Figure 31 illustrates double-sided vPC with 32-way port-channel topology.

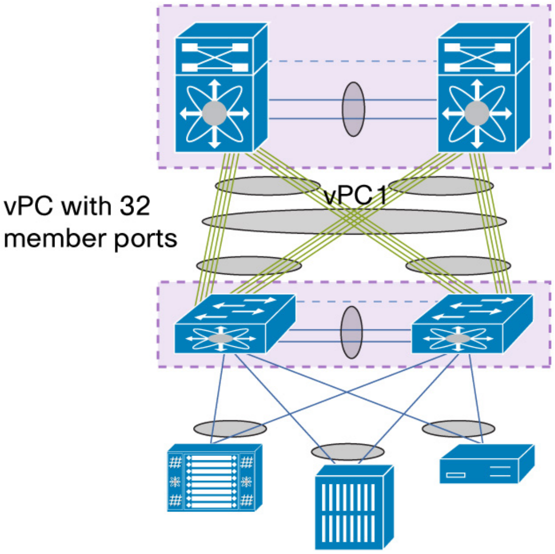
Upper vPC domain is usually used as aggregation layer (L2/L3 boundary).

Downer vPC domain is usually used as access layer (L2 only).

The double-sided vPC with 32-way port-channel topology has the following characteristics:

- Each vPC peer at aggregation layer has vPC member port composed with 16 active links, and the pair (i.e vPC) has 32 active load-balanced links
- Only The F1 and F2 Series line cards support 16 way active port-channel load balancing. So vPC peer device needs to be populated with either of these modules (M1 series line card support only 8 members ports max within a port-channel).
- For north to south traffic and the opposite direction, vPC peer device (at access or aggregation layers) always perform **local** load-balancing out of its vPC member port unless the unique path to southbound/northbound device is only through vPC peer-link.

Figure 31. Double-Sided vPC with 32-Way Port-Channel

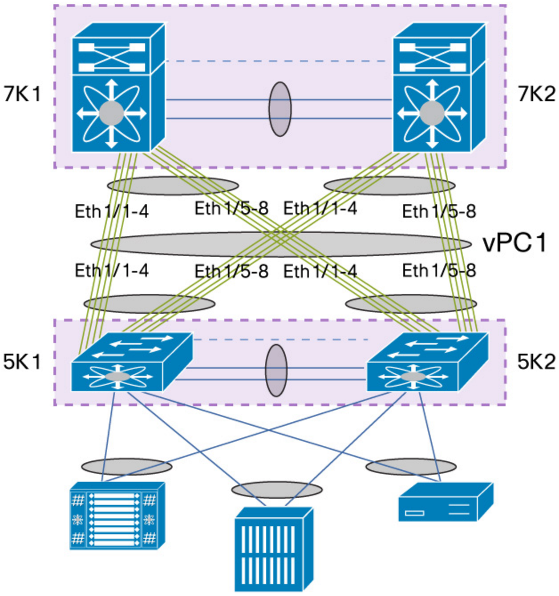


Double-Sided vPC Configuration Sample:

Here's how to configure double-sided vPC for both NEXUS 7000 devices at aggregation layer and both NEXUS 5000 devices at access layer.

Ports connectivity for the sample configuration is depicted in figure 32.

Figure 32. Double-Sided vPC Sample Configuration



7K1 configuration:

```
interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  vpc 1

interface Ethernet1/1-4
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown

interface Ethernet1/5-8
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown


! vPC peer-link
interface port-channel10
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  spanning-tree port type network
  vpc peer-link
```

7K2 configuration:

```
interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  vpc 1

interface Ethernet1/1-4
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown

interface Ethernet1/5-8
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown


! vPC peer-link
interface port-channel10
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  spanning-tree port type network
  vpc peer-link
```

5K1 configuration:

```
interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  vpc 1

interface Ethernet1/1-4
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown

interface Ethernet1/5-8
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown


! vPC peer-link
interface port-channel10
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  spanning-tree port type network
  vpc peer-link
```

```
5K2 configuration:

interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  vpc 1

interface Ethernet1/1-4
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown

interface Ethernet1/5-8
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  channel-group 1 mode active
  no shutdown


! vPC peer-link
interface port-channel10
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 1000-1100
  spanning-tree port type network
  vpc peer-link
```

Particularity of double-sided vPC in regards to configuration is the following:

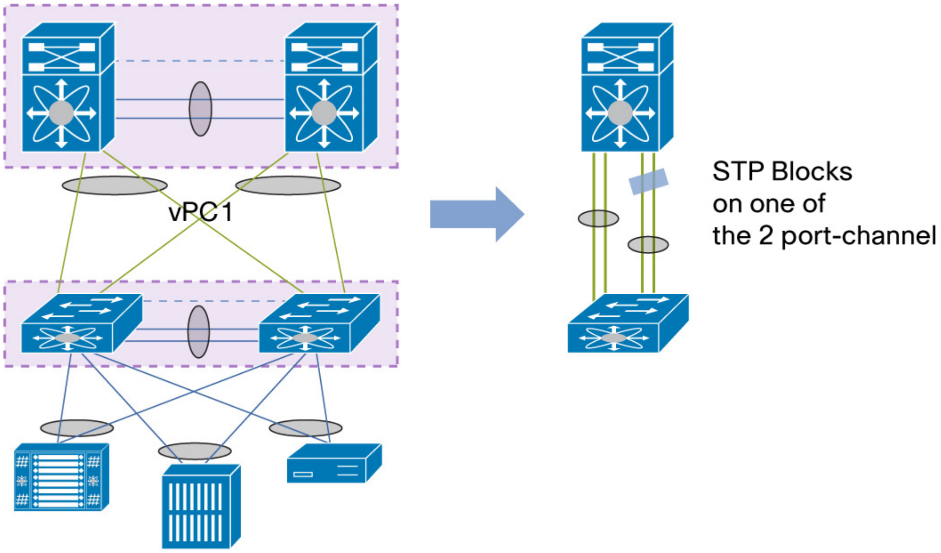
All ports from access layer vPC peer device to aggregation layer vPC domain belong to same port-channel.

All ports from aggregation layer vPC peer device to access layer vPC domain belong to same port-channel.

For ease of configuration and operations, best practice is to use same port-channel id and same vPC id on both vPC domains for the interconnect link (i.e vPC in the middle of the 2 vPC domains).

One common mistake with double-sided vPC topology is to forget to create the port-channel on access layer vPC domain as depicted in figure 33.

Figure 33. Incorrect Configuration of Double-Sided vPC Topology



Creating port-channel on aggregation layer vPC domain but not on access layer vPC domain results in topology where 2 separate port-channels interconnect the 2 layers. In this case, spanning tree protocol detects the network loop and then blocks one of the two port-channels.

Required Recommendation:
In a double-sided vPC topology, all interconnect links between the 2 vPC domains MUST belong to the same vPC. All links form a unique vPC (on both sides of the 2 vPC domains). VPC id can be different across the 2 vPC domains. However, vPC id must be the same across the 2 peer devices of the same domain.

Access Device Single-Attached to vPC Domain

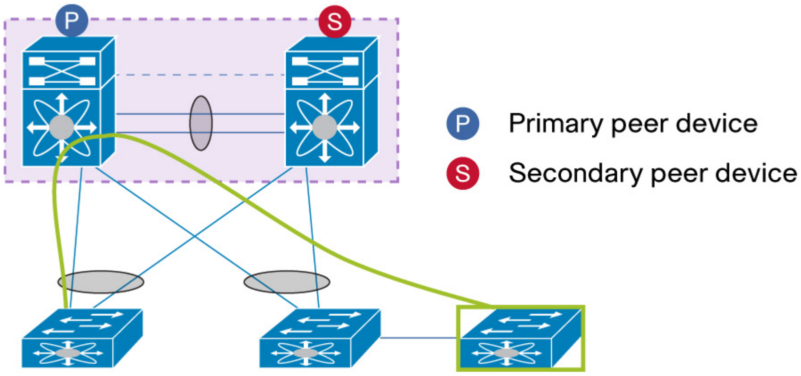
When you cannot dual-attach devices to a vPC domain, there are three main alternatives. We'll discuss these alternatives in order of preference.

The best alternative if dual-attaching is not an option is to connect the device through a vPC-attached access switch (Figure 34). The advantage of this approach is to ensure minimum disruption in case of a peer-link failover and it also provides consistent behavior in vPC dual-active scenarios.

There are two disadvantages of connecting the access device through a vPC-attached access switch:

- Need to use an additional access switch (can be an external device or an instantiation of VDCs)
- Additional administrative burden to configure and manage the physical device or virtual device in case of VDC.

Figure 34. Access Device Attached Through a vPC-Attached Access Switch



If you can’t use an intermediate vPC-attached switch or VDC, the next best alternative is to connect the device to vPC peer devices using a non-vPC VLAN (by definition, a non-vPC VLAN is a VLAN that is not part of any vPC and then not present on vPC peer-link) and create a dedicated inter-switch port-channel to carry these non-vPC VLAN (figure 35).

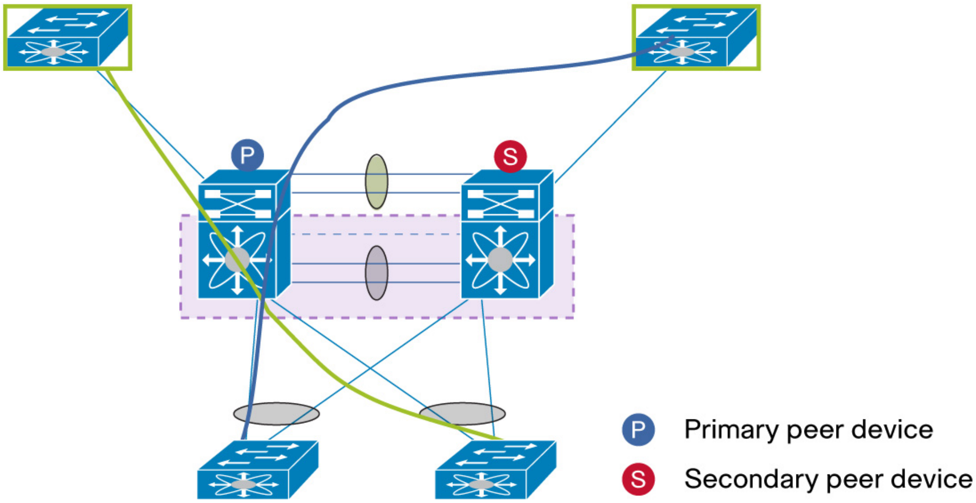
Access device can be connected to primary peer device or secondary peer device. It does not matter because the dedicated inter-switch port-channel guarantees a backup path in case vPC peer-link fails down.

The advantage of this approach is that in case of a peer-link failover, traffic is diverted on a secondary path, avoiding isolation for any device attached to secondary peer device.

The disadvantage is that you need to configure and manage additional port channels between the Cisco Nexus 7000 Series devices.

Communications from non-vPC VLAN to vPC VLAN must be done through inter-VLAN routing as the 2 VLAN belong to different bridging domain.

Figure 35. Access Device Single-Attached to vPC Domain Using Non-vPC VLAN

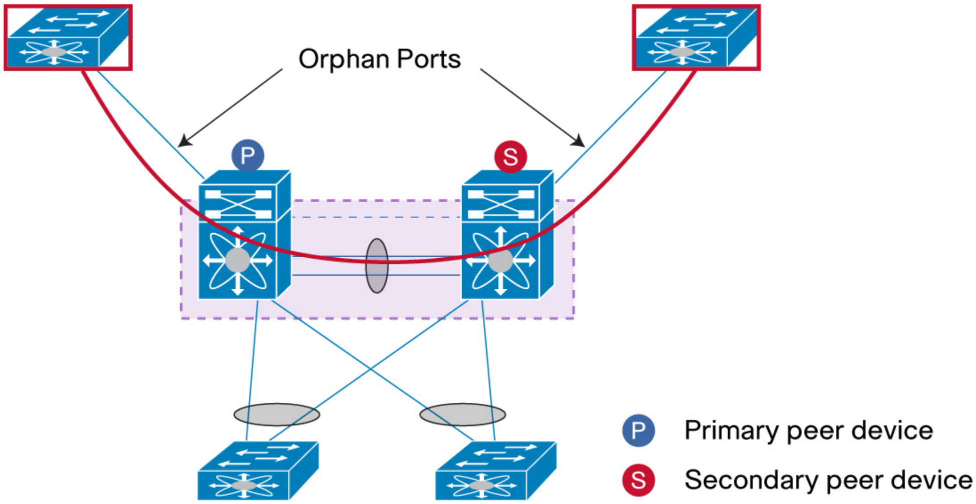


Finally, if you don't want to use non-vPC VLAN to connect the access device to vPC domain (and don't want to create additional inter-switch link between the 2 peer devices), then you can connect the access device to primary peer device using vPC VLAN and leveraging vPC peer-link (Figure 36).

The advantage of this method is the relative ease of deployment (no new VLAN to provision, no new additional inter-switch links to add).

Because of lack of role preemption with vPC, this method results in the vPC domain being bound to vPC roles. As a result, access devices connected to operational secondary peer device become fully isolated in case of peer-link failure.

Figure 36. Access Device Single-Attached to vPC Domain Using vPC VLAN



This particular topology introduces the concept of Orphan Port.

Cisco Nexus 7000 Series ports that are connected to these single-attached devices using vPC VLAN are called orphan ports.

An orphan port has the following characteristics:

- A port on vPC peer device (primary or secondary) that is connected to a single attached device.
- A port on vPC peer device (primary or secondary) that carries vPC VLAN. If the port carries a non-vPC VLAN, it is no more defined as Orphan Port.

Use **show vpc orphan-ports** command to display all Orphan Ports on vPC peer device.

A sample output of this command is shown below:

```
7K1# sh vpc orphan-ports
Note:
--------::Going through port database. Please be patient.::--------

VLAN           Orphan Ports
-------        -------------------------
11             Eth3/23
23             Eth3/21
50             Eth3/14
600            Eth1/41, Eth1/48
```

In this example, VLAN 11, 23, 50, 600 are vPC VLAN (so they are also defined on vPC peer-link).

General Recommendation:
When connecting a single-attached access device to vPC domain using vPC VLAN, always connect it to vPC primary peer device. Reason is when vPC peer-link fails down, any single attached device connected to secondary peer device (and using vPC VLAN) will become completely isolated wih the rest of the network.

In conclusion of this section (single attached devices to vPC domain):

General Recommendations (by descending order of priority):

- Connect access device to an intermediate switch which is dual-attached to vPC domain
- Connect single-attached device to vPC domain using non-vPC VLAN. Create an inter-switch link between the 2 peer devices to transport non-vPC VLAN.
- Connect single-attached device to vPC domain using vPC VLAN and leveraging vPC peer-link.

Access Device STP-Attached to vPC Domain

In case access device supports Spanning Tree Protocol (STP) and needs to be connected to vPC domain, just follow the same recommendations as previously described in the section “access device single-attached to vPC domain”.

Likewise the previous situation, there are 2 options to connect the STP-attached acces device to vPC domain:

Option 1:
Connect the access device through two independent links using Spanning Tree Protocol. Use non-vPC VLAN only on the Spanning Tree Protocol switch. Run the same STP mode as the vPC domain (RVPST or MST), and enable port type edge (i.e port fast) or port type edge trunk (in case access port is connected to hypervisor server) on the host-facing ports. Insert an additional inter-switch link between the 2 vPC peer devices to carry non-vPC VLAN.

An advantage of this best practice is that it helps ensure minimal disruption in case of a peer-link failover and provides consistent behavior with vPC dual-active scenarios. It also helps to ensure full redundant active-active paths on vPC VLANs.

A disadvantage is that it requires an additional Spanning Tree Protocol port-channel between the 2 vPC peer devices.

There is also an operational burden in provisioning and configuring separate STP and vPC VLAN instances. There are only active-standby paths on a Spanning Tree Protocol VLAN.

Option 2:
Connect the access device through two independent links using Spanning Tree Protocol. Use vPC VLANs on this switch. Run the same STP mode as the vPC domain (RVPST or MST), and enable port type edge (i.e port fast) or port type edge trunk (in case access port is connected to hypervisor server) on the host-facing ports.

There is no need to insert an additional inter-switch link between the 2 vPC peer devices as vPC peer-link is re used to carry vPC VLAN from access device.

This best practice simplifies VLAN provisioning (no need to create new VLAN) and does not require allocation of an additional 10 Gigabit Ethernet port-channel. The disadvantage is that access device may be isolated to the rest of network if STP forwarding link is connected to secondary peer device and peer-link fails down. That’s why for option 2, best practices is to connect the STP port in forwarding state to vPC primary peer device.

General Recommendations (by descending order of priority):

- Connect STP-attached device to vPC domain using non-vPC VLAN. Create an inter-switch link between the 2 peer devices to transport non-vPC VLAN.
- Connect STP-attached device to vPC domain using vPC VLAN and leveraging vPC peer-link. Always connected the STP port in forwarding state to vPC primary peer device.

Best Practices for Data Center Interconnect and Encryption

Data Center Interconnect (DCI) is becoming increasingly popular as most of the companies build 2 distincts data centers (at least) to increase High Availabilty and Business Operations. Purpose of DCI is to extend specific VLAN across the different data centers, offering L2 adjacency for servers and NAS devices while separated by multiple miles distance.

Several technologies can be used for DCI like OTV (Overlay Transport Virtualization), VPLS (Virtual Private LAN Services) and vPC.

vPC can be used to interconnect two data centers max. If more than 2 data centers needs to be interconnected to offer VLAN extension across the different sites, recommendation is then to use OTV.

vPC presents the benefit of isolating spanning-tree between the two sites (no BPDU across the DCI vPC) so any outage in 1 data center is not propagated to the other one.

vPC is easy to configure and it provides robust and resilient interconnect solution: vPC protects from any DCI link failure as well as vPC peer device failure.

vPC as DCI technology can be deployed in 2 ways:

- Multilayer vPC for aggregation and DCI
- Dual L2/L3 POD interconnect

In the first configuration, a dedicated vPC domain for DCI is inserted between the aggregation layer and the other data center.

In the second configuration, the same aggragation vPC domain is re used to interconnect the 2 data centers together.

M1 series line card support 802.1ae MACsec encryption directly in hardware. Using M1 ports as member ports for DCI vPC allows to leverage this security feature across the 2 data centers: flows are encrypted within the interconnect vPC making any network intrusion between the 2 facilities useless.

To use 802.1ae MACsec encryption, LAN advanced services license is required. Enable feature cts before being able to configure the M1 port with associated security commands (int eth1/1; cts manual; no propagate-sgt; sap pmk <key>).

F1 ans F2 series line cards do not support hardware 802.1ae MACsec encryption.

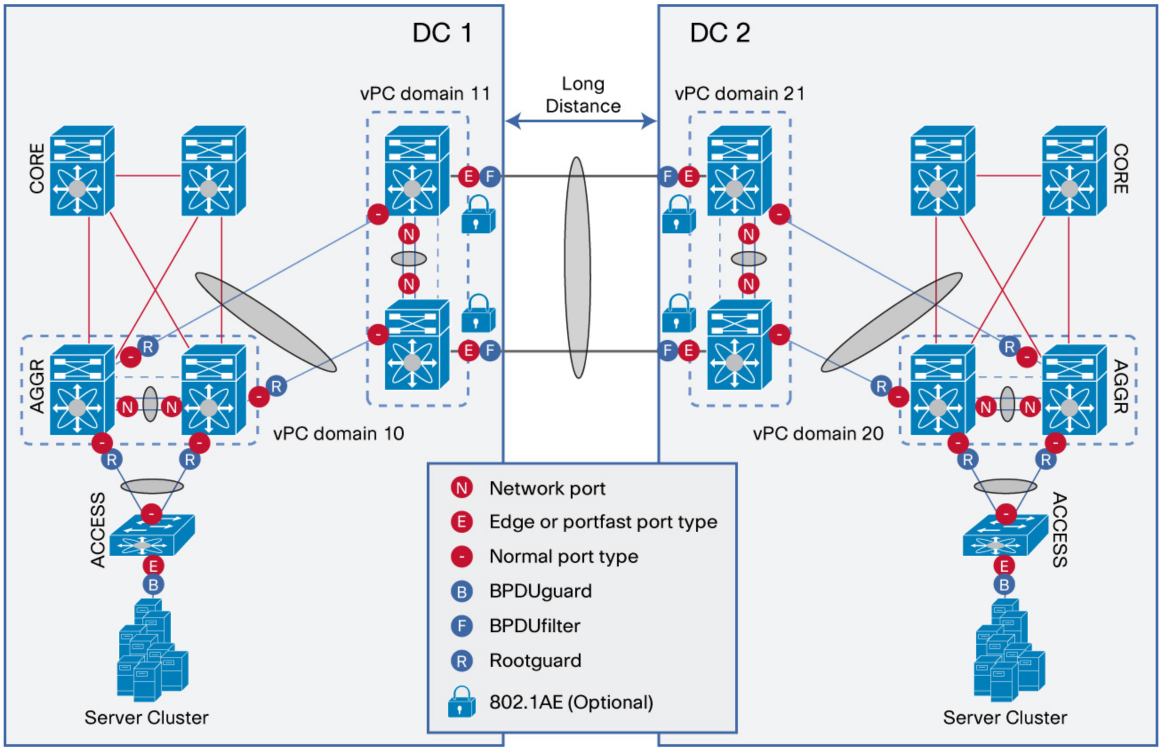
Considerations for HSRP operations with vPC used for DCI are described lated in the section “Best practices for HSRP and vPC”

Multilayer vPC for Aggregation and DCI

Multilayer vPC for aggregation and DCI is a solution where dedicated vPC domains are used for DCI purposes.

One vPC domain is used for server connectivity (called “aggregation vPC domain”) while the other vPC domain is exclusively used for DCI (called “DCI vPC domain”). This topology, shown in Figure 37, clarifies the function of each vPC domain and permits fast troubleshooting and easy monitoring.

Figure 37. Multilayer vPC for Aggregation and DCI



To build successfully a Multilayer vPC for aggregation and DCI solution, use the following guidelines/best practices.

Required Recommendations:

- Use different vPC domain-id for each vPC domain (DC1: vPC domain for aggregation, vPC for DCI. DC2: vPC domain for aggregation, vPC for DCI)
- For each data center, interconnect the aggregation vPC domain to the DCI vPC domain using a vPC (double-sided topology)
- Interconnect the 2 data centers using a vPC (vPC between DCI vPC domain in site 1 and site 2)
- Enable BPDU filter on the vPC used for DCI (under the port-channel configuration, activate the following command: **spanning-tree bpdufilter enable**) to avoid BPDU propagation
- Configure the vPC used for DCI as spanning-tree port type edge (i.e port fast) to fasten port state forwarding mode when port is operationnaly up
- Remember by default vPC peer-link runs in spanning-tree port type network i.e bridge assurance is activated on the link
- Configure root guard on aggregation vPC domain (more exactly on vPC between this vPC domain and DCI vPC domain). STP root must remain on aggregation vPC domain on each side of the data center
- No loop must exist outside the vPC domains

- Do not use Layer 3 peering between data centers (in other words, there is no Layer 3 over vPC)
- Do not use bridge assurance for interconnect vPC (DCI vPC)
- Use M1 ports for DCI vPC if flows between the 2 data centers need to be encrypted using 802.1ae MACsec protocol

Dual Layer 2 /Layer 3 pod Interconnect

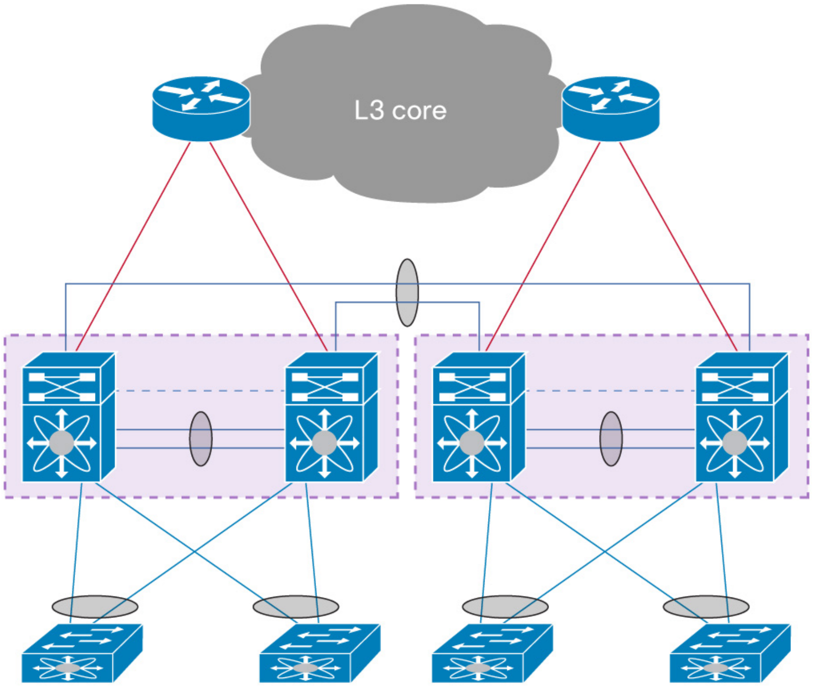
DCI vPC domain can be merged directly in the vPC aggregation domain if cost and consolidation are important considerations.

A dual Layer 2 /Layer 3 pod interconnect provides 2 separate functions within the same comfiguration:

It provides classical aggregation layer functionalities (L2/L3 bundary) and DCI interconnect across the 2 data centers. Bridged traffic will use the dedicated vPC DCI link, while routed traffic will use the interface VLAN (i.e SVI) and dedicated Layer 3 links to the upstream routed core.

Figure 38 shows Dual Layer 2 /Layer 3 Pod Interconnect topology.

Figure 38. Dual Layer 2 /Layer 3 Pod Interconnect Topology



To build successfully a dual Layer 2 /Layer 3 pod interconnect topology, use the following guidelines/best practices.

Required Recommendations:

- Use different vPC domain-id for each vPC domain
- Interconnect the 2 data centers using a vPC

- Enable BPDU filter on the vPC used for DCI (under the port-channel configuration, activate the following command: **spanning-tree bpdufilter enable**) to avoid BPDU propagation
- Configure the vPC used for DCI as spanning-tree port type edge (i.e port fast) to fasten port state forwarding mode when port is operationnaly up
- Remember by default vPC peer-link runs in spanning-tree port type network i.e bridge assurance is activated on the link
- Configure root guard on vPC for DCI. STP root must remain local on each side of the data center.
- No loop must exist outside the vPC domains
- Do not use Layer 3 peering between data centers (in other words, there is no Layer 3 over vPC)
- Do not use bridge assurance for interconnect vPC (DCI vPC)
- Use M1 ports for DCI vPC if flows between the 2 data centers need to be encrypted using 802.1ae MACsec protocol

Best Practices for Spanning Tree Protocol Interoperability

This section describes best practices for Spanning Tree Protocol interoperability with vPC.

About Spanning Tree Protocol Interoperability with vPC

vPC technology allows to build a loop free topology by levering port-channel from access device to vPC domain. A port-channel is seen as a logical link from STP standpoint so vPC domain with vPC-attached access device globally form a star topology at L2 (there is no STP blocked ports in this type of topology). For that case, STP is used as a fail-safe mechanism to protect any network loop caused by human error (like plugging a loopback cable across the 2 vPC peer device).

As seen previously, all kinds of connectivity to vPC domain are fully supported: single-attached access device as well as STP-attached access device in addition to traditional vPC-attached connectivity.

For the first 2 attachment modes, STP must be actively involved to dictate the L2 path, again avoiding any loop in the network.

So depending on the type of connectivity between access device and vPC domain, STP is more or less engaged to build the loopfree L2 path.

Logical ports considerations still apply for STP in the context of vPC. Logical ports count, however calculation is done now on the vPC member port (i.e port-channel) and no more on individual links. STP sees the port-channel as a unique logical link.

RPVST (Rapid Per Vlan Spanning Tree) supports 16,000 logical ports and MST (Multiple Spanning Tree) supports up to 90,000 logical ports.

Role of Spanning Tree Protocol within vPC Domain

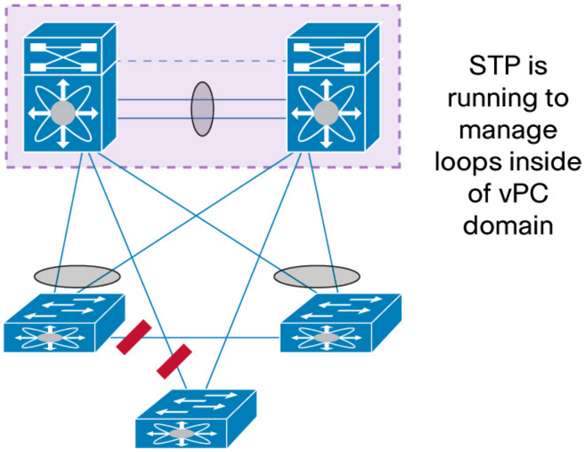
In the context of vPC technology, Spanning Tree Protocol provides the following functionalities:

- Protect the L2 network by detection and breaking any loops
- Dictates Layer 2 path for non-vPC attached devices (i.e single-attached device or STP-attached device)

- Loop management on the addition or removal of a vPC (avoid L2 loops in the particular network configuration events related to vPC)

In Figure 39, Spanning Tree Protocol is running to manage loops inside of the vPC domain, or before initial vPC configuration. Spanning Tree Protocol is also actively running in case a device is single-attached to vPC domain.

Figure 39. Role of STP within vPC Domain



Recommended Spanning Tree Protocol Configuration with vPC

We recommend that you configure both ends of vPC with identical Spanning Tree Protocol configurations for the following global parameters and interface settings:

Global parameters:

- STP mode (RPVST or MST)
- STP region configuration for MST
- Enable/disable state per VLAN
- Bridge Assurance setting
- STP Port type setting (Enable or Disable edge port type by default on all access ports)
- Loop Guard settings (Enable or Disable loop guard by default on all ports)
- BPDU Guard settings (Enable or Disable BPDU guard by default on all edge ports)
- BPDU filter settings ((Enable or Disable BPDU filter by default on all edge ports)

Interface settings:

- STP Port type setting (edge, network or normal)
- Loop Guard (enabled or disabled)
- Root Guard (enabled or disabled)

Note: If any of these parameters are misconfigured, the Cisco NX-OS Software suspends all interfaces in the vPC (this is a type-1 consistency check error). Starting NX-OS 5.2 and introduction of vPC graceful consistency check feature, only secondary peer device suspends its interfaces in the vPC. Primary peer device maintains its vPC member ports to up and operational state.

Syslog messages are sent when vPC member ports are suspended; **show vpc brief** command indicates the status of vPC state.

To help avoid unpredictable behavior in the traffic flow, ensure that the following Spanning Tree Protocol interface configurations are identical on both sides of the vPC (this is a type-2 consistency check error):

- BPDU Filter
- BPDU Guard
- STP Cost (STP port path cost)
- STP Link type (auto, point-to-point, shared)
- STP Priority (STP port priority)
- VLANs

Display the configuration on both sides of the vPC to ensure that the settings are identical by using the command **sh run int port-channel <id> membership.**

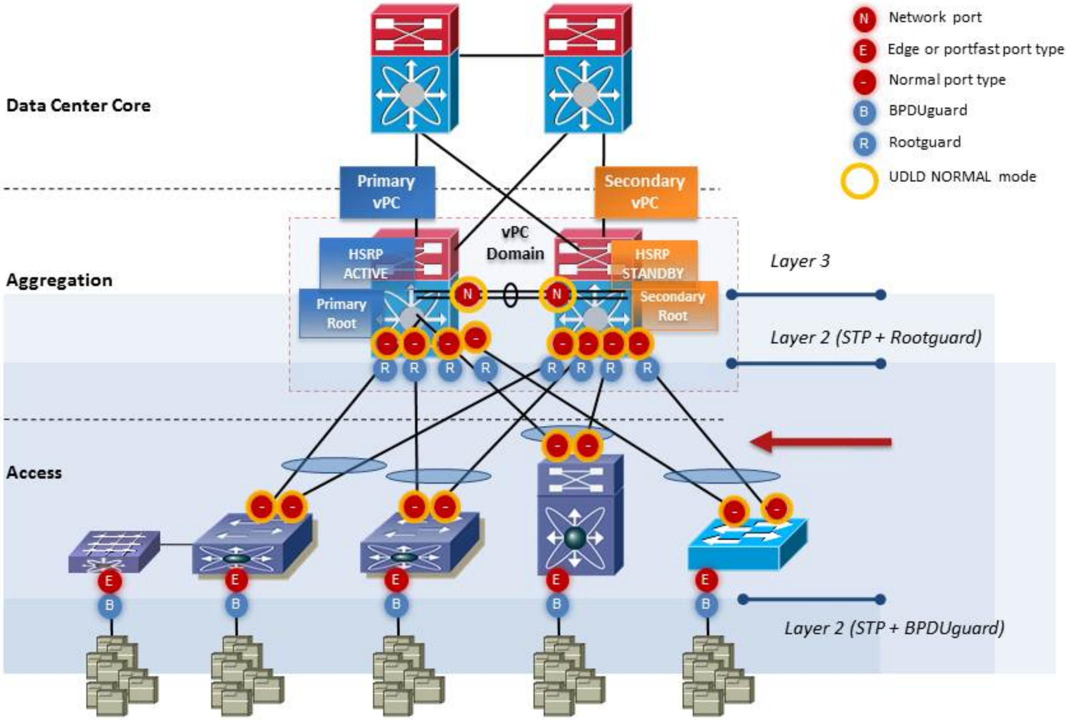
Strong Recommendations:

- Spanning Tree Protocol must remain enabled for all VLAN (even if all access devices are vPC-attached to vPC domain). Do not disable spanning-tree protocol!
- Use MST with vPC if you need to build a large L2 domain. Plan ahead to avoid future configuration changes that can trigger vPC type-1 consistency failure (Sample global type-1 parameters include MST region configuration, STP mode, STP global configuration)
- Implement consistent STP mode in the same L2 domain: Ensure that all switches in your Layer 2 domain are running with Rapid-PVST+ or MST to avoid slow Spanning Tree Protocol convergence (30 seconds or more)
- Perform VLAN pruning on vPC member ports to reduce internal resources consumption

STP Interoperability with vPC - Blueprint Diagram

The following blueprint diagram (Figure 40) displays the Spanning Tree Protocol port configuration recommended with vPC.

Figure 40. Spanning Tree Protocol and vPC: Port Configuration Recommendation



Strong Recommendations:

- Keep Spanning Tree Protocol root function on the aggregation layer of the network (aggregation vPC domain)
- For each vPC peer device, configure root guard on ports connected to access devices
- Bridge Assurance is enabled by default when configuring vPC peer-link. Do not disable it on vPC peer-link
- It is not necessary to enable Bridge Assurance on vPC (Bridge Assurance is enabled when the vPC member port is defined as spanning-tree port type network)
- Configure vPC member port as spanning-tree port type normal (so not using Bridge Assurance on the link)
- Configure port fast (edge port type) on the host-facing interfaces to avoid slow Spanning Tree Protocol convergence (30 seconds or more) when port transitions to up state.

- Configure BPDU guard on host-facing interfaces to block any BPDU sent from the host (access switch port receiving the BPDU will be put in errdisable mode).

vPC and Spanning Tree Protocol Bridge Protocol Data Units

vPC maintains dual active control planes and Spanning Tree Protocol still runs on both switches.

To interoperate with vPC technology, STP implementation has been adapted to operate in a dual-pair device configuration.

For vPC ports only the vPC primary switch runs the STP topology for those vPC ports. In other words, Spanning Tree Protocol for vPCs is controlled by the vPC primary peer device, and only this device generates then sends out Bridge Protocol Data Units (BPDUs) on Spanning Tree Protocol designated ports. This happens irrespectively of where the designated Spanning Tree Protocol root is located.

STP on the secondary vPC switch must be enabled but it doesn't dictate vPC member port state.

vPC secondary peer device proxies any received Spanning Tree Protocol BPDU messages from access switches toward the primary vPC peer device (Figure 41).

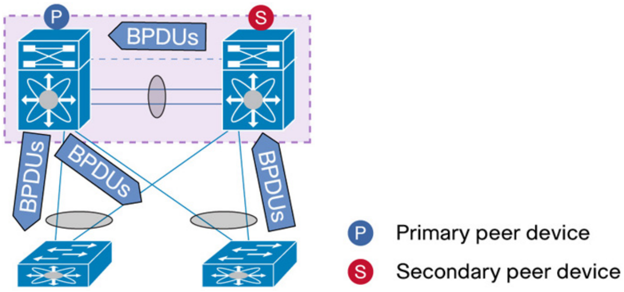
Both vPC member ports on both peer devices always share the same STP port state (FWD state in a steady network).

By default, STP implementation (in the context of vPC) allocates each vPC peer device with its own bridgeID value. As vPC domain is usually the STP root for all VLAN in the domain, rootID value is equal to bridgeID of primary peer device or secondary peer device.

Strong Recommendations:

- Always define the vPC domain as STP root for all VLAN in that domain (configure aggregation vPC peer devices as STP root primary and STP root secondary)
- Enforce this rule by implementing STP root guards on vPC peer devices ports connected to another L2 switch.

Figure 41. vPC and Spanning Tree Protocol BPDUs



For access switches connected to vPC domain, use the following guidelines/best practices:

Strong Recommendations:

- Enable STP port type "edge" and port type "edge trunk" on host ports

- Enable STP BPDU-guard globally
- Disable STP channel-misconfig guard if supported by access switches
- Do not enable Loopguard on vPC (disabled by default)
- Do not enable Bridge Assurance on vPC

For ease of operations and quick diagnostics, the following recommendation applies to STP in the context of vPC:

General Recommendations:

- Configure the Spanning Tree Protocol root for all VLAN on vPC primary device (spanning-tree vlan 100-102 root primary)
- Configure the Spanning Tree Protocol secondary root for all VLAN on vPC secondary device (spanning-tree vlan 100-102 root secondary)

vPC Peer-link is a regular port for STP.

However, vPC imposes the rule that the peer-link should never be blocking because this link carries important traffic such as the CFSoE. As a consequence, the peer-link is always forwarding for any VLAN that is a member of that link.

When user configures the port-channel as vPC peer-link (adding keyword “vpc peer-link”), the system automatically turns on Bridge Assurance on the link. Bridge Assurance is a STP extension that protects L2 network from any unidirectional link event caused by physical cable failure or adjacent switch control plane failure.

Strong Recommendation:

- Let Bridge Assurance running on vPC peer-link (default mode) and do not disable it.
- Do not enable Bridge Assurance on vPC member ports.

vPC Peer-Switch

Since NX-OS 4.2(6), 5.0(2a) and further, an enhancement, called vPC peer-switch, was brought to STP in the context of vPC.

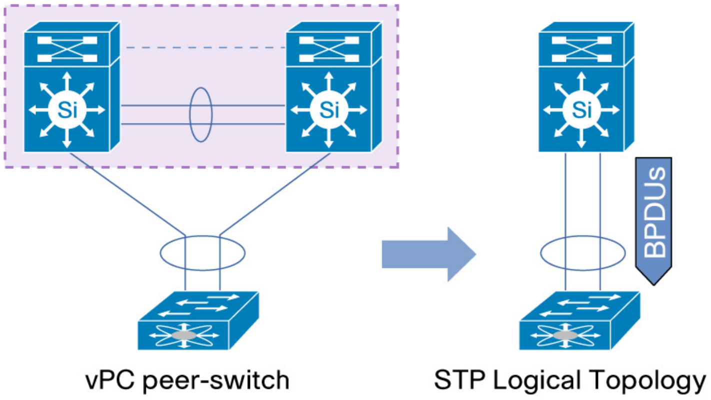
The vPC Peer-Switch feature (Figure 42) allows a pair of vPC peer devices to appear as a single Spanning Tree Protocol root in the Layer 2 topology (they have the same bridge ID). vPC peer-switch must be configured on both vPC peer devices to become operational. The command is the following:

N7K(config-vpc-domain)# **peer-switch**

This feature simplifies Spanning Tree Protocol configuration by configuring vPC VLAN on both peer devices with the same Spanning Tree Protocol priority. A vPC Peer-Switch eliminates the need to map the Spanning Tree Protocol root to the vPC primary peer device.

Main advantage of vPC peer-swtich is the improvement in term of convergence time during vPC primary peer device failure/recovery. Without vPC peer-switch feature, vPC primary peer device failure and recovery usually create around 3 seconds of traffic disruption (for south to north traffic). With vPC peer-switch, traffic disruption is lowered to sub-second value because peer device down an up events do not generate any Rapid Spanning Tree Protocol Sync behavior (from a STP standpoint, there is no change in L2 topology).

Figure 42. vPC Peer-Switch



Most common mistake with vPC peer-switch deals with spanning tree configuration.

When vPC peer-switch is activated, it is mandatory that both peer devices have the exact same spanning tree configuration and more precisely the same Spanning Tree Protocol priority for all vPC VLAN.

This requirement comes from the origin of vPC peer-switch: both peer devices present themselved as a unique STP root device using the same bridge ID.

Typical STP configuration with vPC peer-switch looks are displayed below:

```
7K1 (vPC peer device 1):
spanning-tree vlan 10-101 priority 8192

vpc domain 1
 peer-switch
```

```
7K2 (vPC peer device 2):
spanning-tree vlan 10-101 priority 8192

vpc domain 1
 peer-switch
```

Required Recommendations:

- When vPC peer-swtich is activated, both vPC peer devices MUST have the same spanning tree configuration (same Spanning Tree Protocol priority for all vPC VLAN)

General Recommendation:

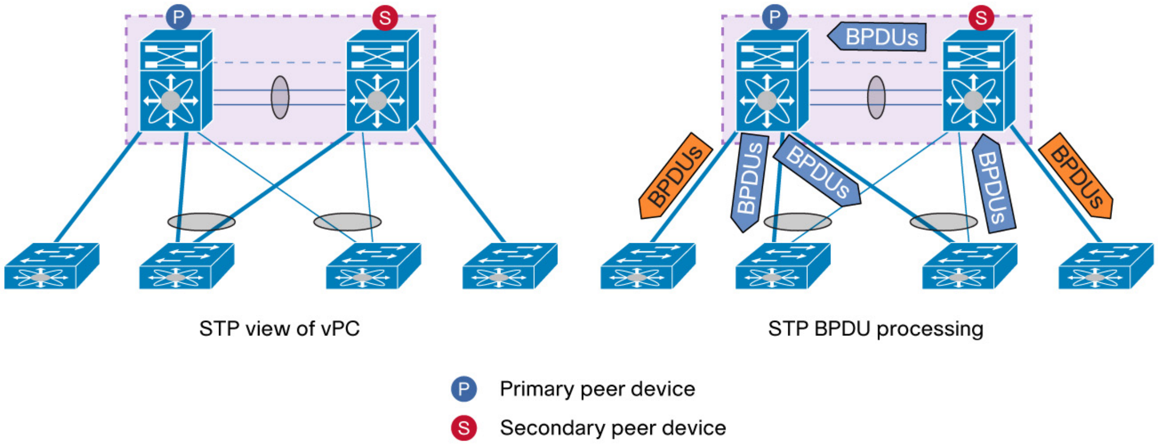
- Activate vPC peer-switch in a vPC environment vPC peer-switch can be used in a vPC domain with different type of access device attachment (This is also called hybrid-topology): access device can be vPC-attached to vPC domain and they can be STP-attached as well.

Single-attached access device are naturally supported in a vPC domain configured with peer-switch.

Let’s have a look at Spanning Tree Protocol BPDU with and without vPC peer-switch functionality.

Figure 43 depicts STP BPDU processing in a vPC domain without peer-switch while Figure 44 depicts STP BPDU processing in the same topology with peer-switch activated.

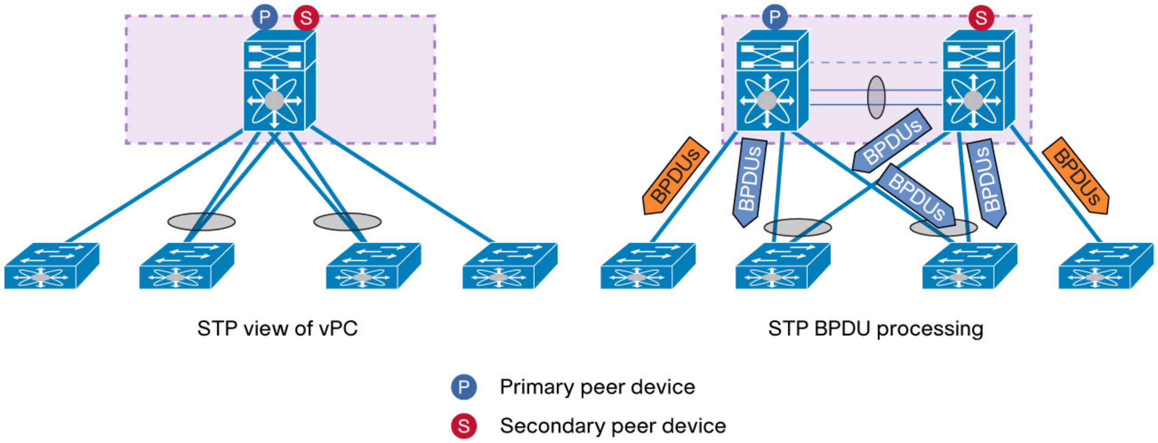
Figure 43. STP BPDU Processing without vPC Peer-Switch



STP BPDU processing behavior without vPC peer-switch:
BPDU is processed only by primary peer device for vPC attached switched.

For directly single-attached switches, the respective connected NEXUS 7000 switch will process locally the BPDU.

Figure 44. STP BPDU Processing with vPC Peer-Switch



STP BPDU processing behavior with vPC peer-switch:

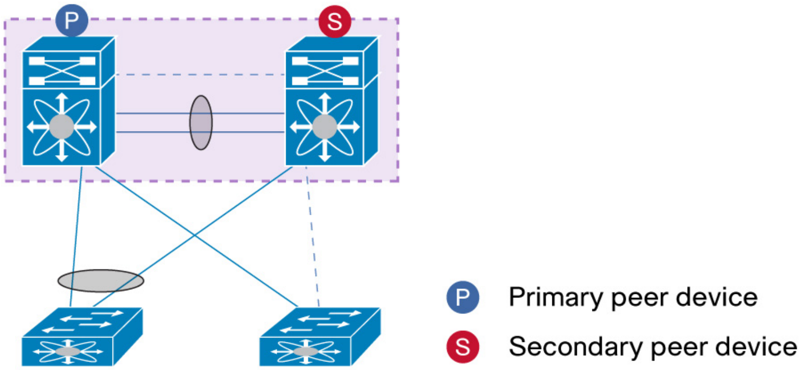
With vPC peer-swtich activated, STP BPDUs are directly processed by the logical Spanning Tree Protocol root formed by the 2 peer devices. Note that a vPC-attached access device will receive 2 BPDU: 1 per vPC peer device. The content of the BPDU is exactly the same. BPDU proxing over vPC peer-link is no more needed once vPC peer-switch is activated.

For directly single-attached switches, the respective connected NEXUS 7000 switch will process locally the BPDU.

vPC peer-switch supports a hybrid topology. Hybrid topology means both vPC-attached access device and STP-attached access device co-exist in the vPC domain.

Figure 45 represents a hybrid topology in the context of vPC with peer-switch activated on both peer devices.

Figure 45. Hybrid Topology with vPC Peer-Switch (vPC and STP-Attached Access Device)



Spanning-tree pseudo-information knob has been introduced to enable VLAN load-balancing for STP-attached access device and to avoid spanning tree topology change when a peer device recovers (after a failure or reload).

Spanning-tree pseudo-information configuration containts 2 sub-commands: **designated priority** and **root priority**.

Designated priority defines the STP priority for the VLAN on the bridge (i.e peer deivce) and is used to effectively load balance the different VLAN across the 2 peer devices.

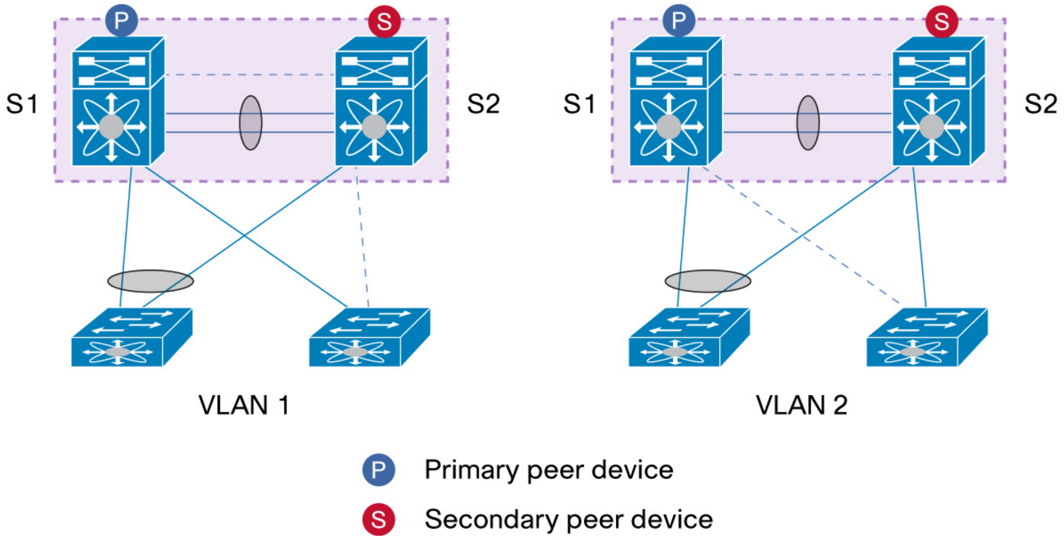
Root priority is used for a specific case when one of the two peer devices fails and recovers: In hybrid topology, STP topology change when a vPC peer device (say S1) recovers because of the difference between the regular STP links (non-vPC) and vPC link bring-up. Regular STP link can be up prior to vPC, and hence the vPC peer-switch formation. Since vPC peer-switch is not formed, the peer device S1 will use the local system MAC for STP bridge ID an if that local MAC address is better than the vPC system MAC then it will trigger STP topology change since the STP bridge priority is the same on both vPC peer devices.

In order to void the STP topology change when S1 recovers, the STP bridge priority for vPC peer-switch should be better than the local bridge ID priority.

Below is a sample configuration showing how to use **Spanning-tree pseudo-information.**

Figure 46 is used as the reference topology for the example. S1 is the STP root for VLAN 1 and S2 is the STP root for VLAN 2.

Figure 46. Hybrid Topology Reference for Peer-Switch with Spanning-Tree Pseudo-Information Configuration Sample



```
S1 configuration:

S1(config)# spanning-tree pseudo-information
S1(config-pseudo)# vlan 1 designated priority 4096
S1(config-pseudo)# vlan 2 designated priority 8192
S1(config-pseudo)# vlan 1 root priority 4096
S1(config-pseudo)# vlan 2 root priority 4096

S1(config)# vpc domain 1
S1(config-vpc-domain)# peer-switch
```

```
S2 configuration:

S2(config)# spanning-tree pseudo-information
S2(config-pseudo)# vlan 1 designated priority 8192
S2(config-pseudo)# vlan 2 designated priority 4096
S2(config-pseudo)# vlan 1 root priority 4096
S2(config-pseudo)# vlan 2 root priority 4096

S2(config)# vpc domain 1
S2(config-vpc-domain)# peer-switch
```

Strong Recommendation:

When using vPC peer-switch in a hybrid environment (i.e dual-attached access device and STP-attached access device co-exist in the same vPC domain), use spanning-tree pseudo-information to load-balance VLAN across the

2 peer devices and to avoid spanning tree topology change when a peer device recovers from a failure or from a reload event.

Bridge Assurance and vPC

Bridge Assurance is a STP extension that protects L2 network from any unidirectional link event caused by physical cable failure or adjacent switch control plane failure.

Bridge Assurance causes the switch to send BPDUs on all operational ports that carry a STP port type setting of "network", including alternate and backup ports for each hello time period. If a neighbor port stops receiving BPDUs, the port is moved into the blocking state. If the blocked port begins receiving BPDUs again, it is removed from bridge assurance blocking state, and goes through normal Rapid-PVST transition.

This bidirectional hello mechanism helps prevent looping conditions caused by unidirectional links or a malfunctioning switch (control plane down but still operational/forwarding on data plane side).

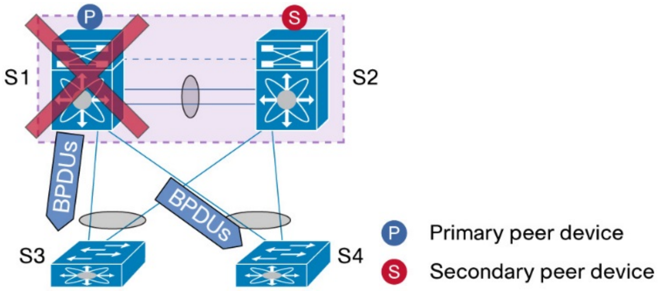
Bridge Assurance BPDU are processed directly by SUPERVISOR CPU (Bridge Assurance BPDU is STP BPDU). It is not hardware off-loaded down to line card (like BFD for instance).

Bridge Assurance BPDU are sent periodically. Default hello time is 2 seconds.

Withing the context of vPC, Bridge Assurance is not recommended on vPC.

Assume the topology shown in Figure 47 with Bridge Assurance enabled on each vPC member ports.

Figure 47. Using Bridge Assurance (BA) on vPC



S1 is the primary peer device and is configured to be STP root for all vPC VLAN.

S1, S2, S3 and S4 are configured with Bridge Assurance enabled on vPC member ports and port-channels respectively.

In a steady state, primary peer device processes the STP BPDU. If the primary peer device fails over, the secondary peer device needs to start sending BPDUs. As the primary peer device was also the Spanning Tree Protocol root, the secondary also has to take over the STP role as root. If this process lasts too long, the uplink ports on access devices (S3 and S4) may go into Bridge Assurance inconsistent (BA_Inconsistent) state. This can occur in specific conditions of intense CPU utilization.

Note: Bridge Assurance is enabled by default on vPC peer-link (at the creation of the link). Bridge assurance on the peer-link is fine so there is no need to disable it.

Bridge Assurance makes sense in a pure Spanning Tree Protocol environment with STP blocked ports to help ensure that those ports will not transition to a forwarding state by error (by sending bidirectional BPDUs). But in a vPC environment, there are no STP blocked ports with vPC (vPC member ports are always in forwarding state), so making the Bridge Assurance feature less useful.

Strong Recommendations:

- Do not enable Bridge Assurance (BA) on vPC member port.
- Bridge Assurance is enabled automatically on vPC peer-link at creation of the link. Bridge assurance on the peer-link is fine so there is no need to disable it.
- Even if peer-switch is used in the vPC domain, recommendation is still to disable Bridge Assurance on the vPC.

NX-OS and IOS Internal VLAN Range Allocation

Cisco NX-OS Software and Cisco IOS Software have different internal VLAN range allocation.

The range number is shown below.

Cisco NX-OS Software internal VLAN range allocation:

```
7K1# sh vlan internal usage

VLAN        DESCRIPTION
---------   -------------------------------------------------------
3968-4031        Multicast
4032             Online diagnostics vlan1
4033             Online diagnostics vlan2
4034             Online diagnostics vlan3
4035             Online diagnostics vlan4
4036-4041        Reserved
4042             Satellite
4043-4047        Reserved
4094             Reserved
```

Cisco IOS Software internal VLAN range allocation:

```
6500-1#sh vlan internal usage

VLAN Usage
---- --------------------
1006      online diag vlan0
1007      online diag vlan1
1008      online diag vlan2
1009      online diag vlan3
1010      online diag vlan4
1011      online diag vlan5
```

```
  1012       PM vlan process (trunk tagging)
  1013       Port-channel103
  1014       Control Plane Protection
  1015       Layer 3 multicast partial shortcuts for VPN 0
  1016       Egress internal vlan
  1017       Multicast VPN 0 QOS vlan
  1018       IPv6 Multicast Egress multicast
```

When connecting a Cisco Nexus device to a Cisco Catalyst® device, be cautious with the VLAN used for that purpose in order to avoid any reserved VLANs from the NX-OS range or IOS range.

Note: Starting with Cisco NX-OS Release 5.2, it is possible to change the reserved VLAN range with the command **system vlan {start-vlan} reserve**.
Please refer to the following URL for more information (Changing the range of reserved VLAN - L2 switching configuration guide): http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/layer2/configuration/guide/b_Cisco_Nexus_7000_Series_NX-OS_Layer_2_Switching_Configuration_Guide_chapter_0100.html#task_67E5266F50104AF38E5149C1CC56B1A7

Best Practices for Layer 3 and vPC

This section describes best practices for using and configuring Layer 3 with vPC.

About Layer 3 and vPC

When a Layer 3 device (a router or a firewall configured in routed mode for instances) is connected to a vPC domain through a vPC, it gets the following views:

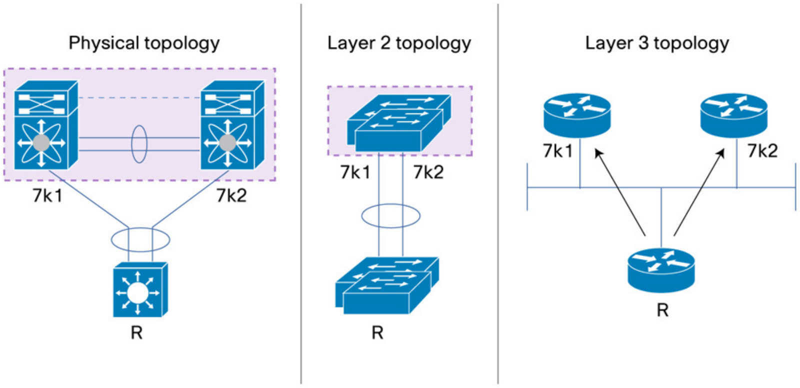
- At Layer 2, the L3 device sees an unique Layer 2 switch formed by the vPC peer devices.
- At Layer 3, the L3 device sees two distinct Layer 3 devices (one for each vPC peer device).

vPC is a Layer 2 virtualization technology. That is why at Layer 2, both vPC peer devices present themselves as a unique logical device to in the rest of the network.

However, at Layer 3 no virtualization mechanism is implemented with vPC technology. This is the reason why each vPC peer device is seen as a distinct L3 device by the rest of the network.

Figure 49 represent the 2 distinct views with vPC technology depending on Layer 2 or Layer 3 vision of the network.

Figure 48. Different Views for vPC Peer Devices Depending on Layer 2 or Layer 3 Vision of the Network



Layer 3 and vPC: Guidelines and Restrictions

Attaching a L3 device (router or firewall configured in routed mode for instance) to vPC domain using a vPC is not a supported design because of vPC loop avoidance rule.

To connect a L3 device to vPC domain, simply use L3 links from L3 device to each vPC peer device.

L3 device will be able to initiate L3 routing protocol adjacencies with both vPC peer devices.

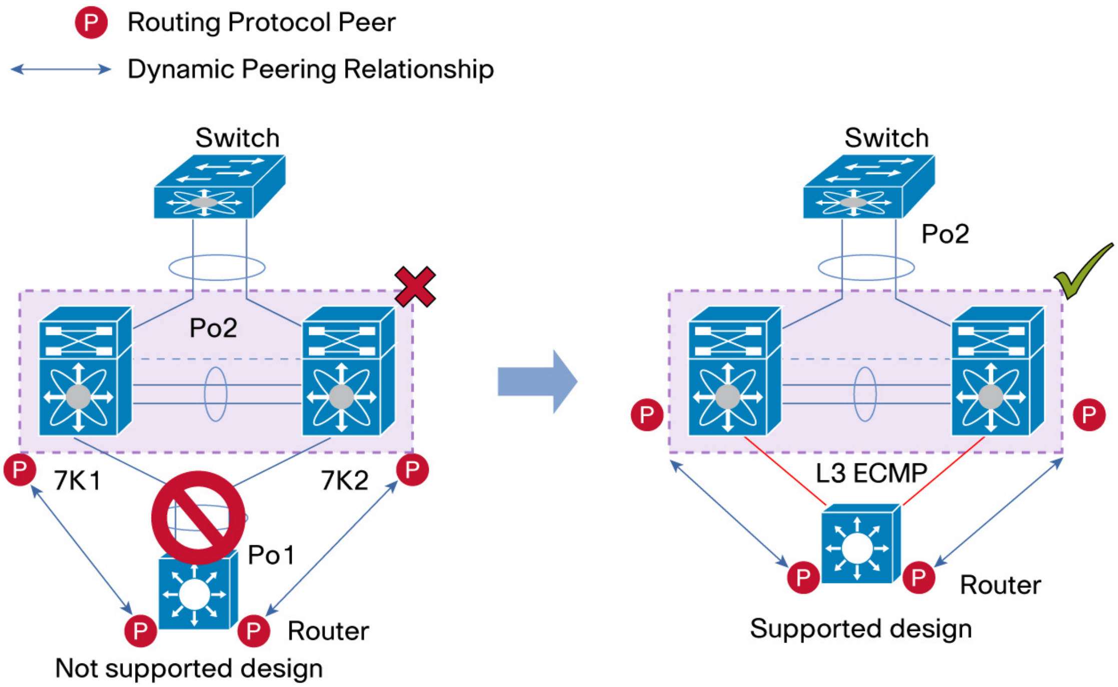
One or multiple L3 links can be used to connect to L3 device to each vPC peer device. NEXUS 7000 series support L3 Equal Cost Multipathing (ECMP) with up to 16 hardware load-sharing paths per prefix. Traffic from vPC peer device to L3 device can be load-balanced across all the L3 links interconnecting the 2 devices together.

Using Layer 3 ECMP on the L3 device can effectively use all Layer 3 links from this device to vPC domain. Traffic from L3 device to vPC domain (i.e vPC peer device 1 and vPC peer device 2) can be load-balanced across all the L3 links interconnecting the 2 entities together.

There is then no penalty to use L3 links to connect L3 device to vPC domain compared to a vPC connection (in the sense that multiple links can also be leveraged with L3 connectivity).

The supported connection model for L3 device to vPC domain is illustrated in figure 50.

Figure 49. Using Separate Layer 3 Links to Connect L3 Device to a vPC Domain



We strongly recommend that you follow these guidelines when connecting a Layer 3 device to vPC domain.

Strong Recommendations:

- Use separate Layer 3 links to connect L3 device (like router or firewall in routed mode for instance) to a vPC domain (Figure 50).
- Do not use a Layer 2 vPC to attach L3 device to a vPC domain unless L3 device can statically route to the HSRP address configured on vPC peer devices.
- Use individual Layer 3 links for routed traffic and a separate Layer 2 port-channel for bridged traffic if both routed and bridged traffic are required.
- Enable Layer 3 connectivity between vPC peer device by configuring a VLAN network interface for the same VLAN from both devices or by using a dedicated L3 link between the 2 peer devices (for L3 backup routing path purposes).

Layer 3 and vPC Interactions: Supported Designs

Network topologies shown from figures 51 through 59 represent supported designs for Layer 3 and vPC.

Figure 51 defines the symbols used in these diagrams.

Figure 50. Signification of Symbols Used in the Following Diagrams (Figures 51 Through 59)

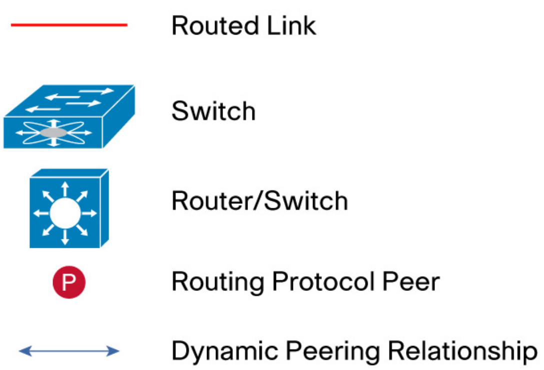
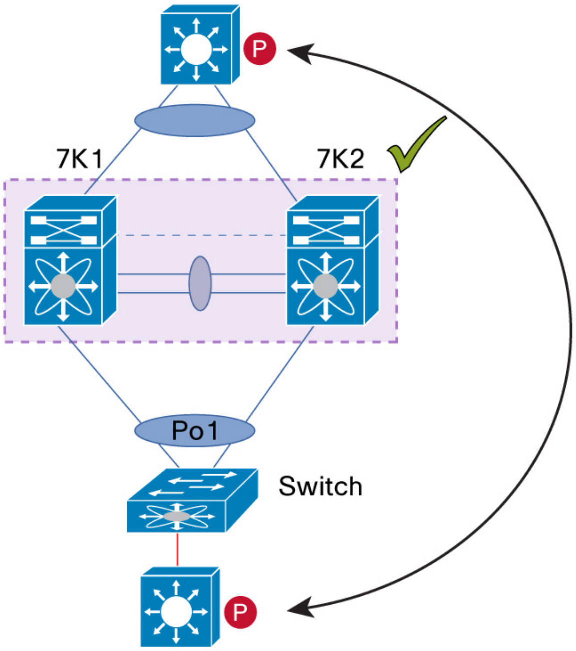
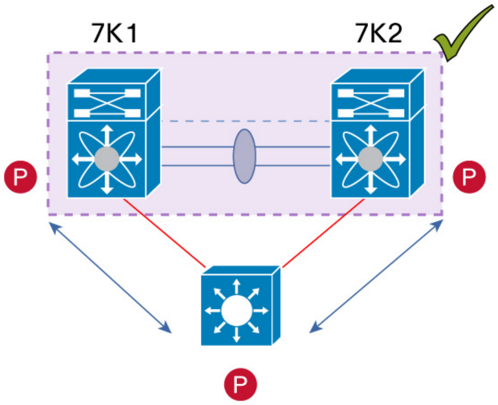


Figure 51. Supported Designs for Layer 3 and vPC: Peering Between ROUTERS



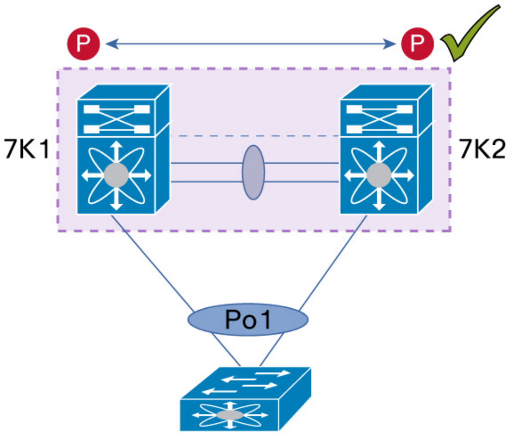
In this design, vPC is used as a pure L2 transit path. Because there is no direct routing protocol peering adjacency from L3 device to any vPC peer device, this topology is entirely valid and fully supported.

Figure 52. Supported Designs for Layer 3 and vPC: Peering with an External Router Using L3 Links



Using L3 links to connected a L3 device to vPC domain the highest recommended way to interconnect the 2 entities together.

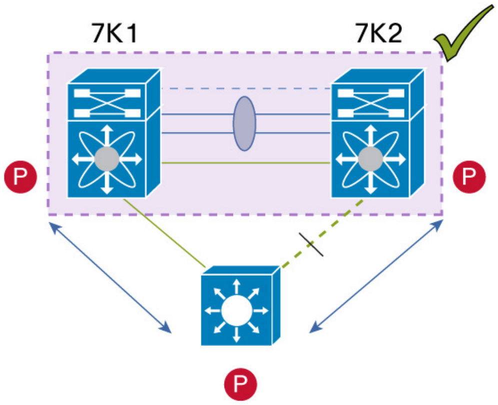
Figure 53. Supported Designs for Layer 3 and vPC: Peering Between vPC Devices (for Backup Routing Path)



Peering between the 2 vPC peer devices is fully functional and primary use case for this type of design deals with L3 backup routed path. In case L3 uplinks on vPC peer device 1 or peer device 2 fail down, the path between the 2 peer devices is used to redirect traffic to the switch having L3 uplinks in UP state.

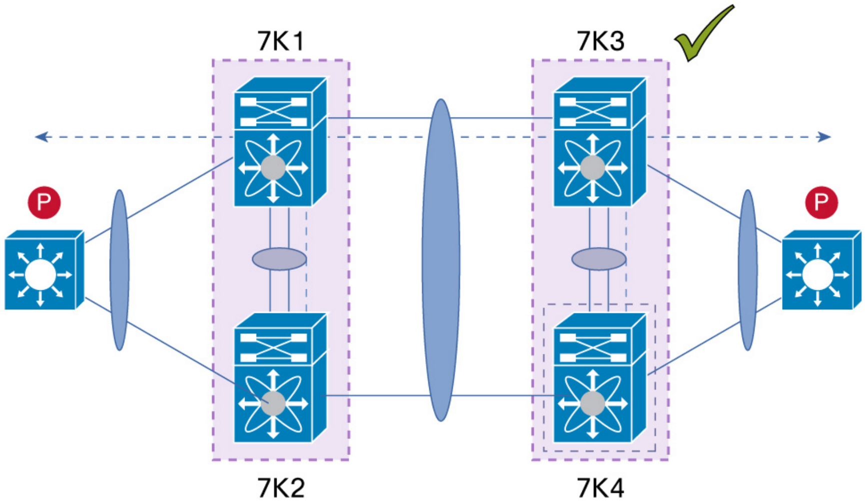
L3 backup routing path can be implemented using dedicated interface VLAN (i.e SVI) over vPC peer-link or using dedicated L2 or L3 links across the 2 vPC peer devices.

Figure 54. Supported Designs for Layer 3 and vPC: Peering over a Spanning Tree Protocol Interconnection Using a non-vPC VLAN



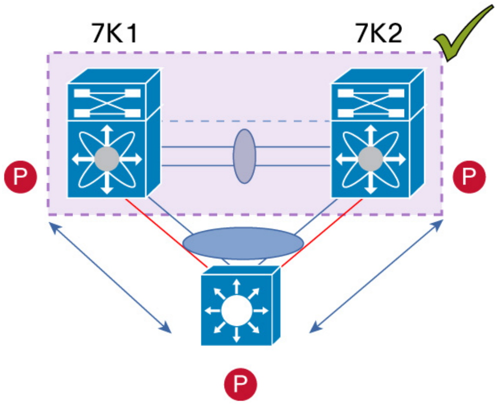
L3 device can be connected to vPC domain using STP interconnect links. Non-vPC VLAN must be used for this type of connection. A dedicated inter-switch link needs to be added across the 2 vPC peer devices to carry non-vPC VLAN traffic. vPC peer-link must not be used for this purpose.

Figure 55. Layer 3 and vPC Supported Designs: Peering between Two Routers with vPC Devices as Transit Switches



This type of design deals with DCI use-case. In term of functioning and operations, it is fully similar to topology shown in figure 51 (Peering between Routers). vPC domains are simply used as pure L2 transit path.

Figure 56. Layer 3 and vPC Supported Designs: Peering with an External Router on Parallel Interconnected Routed Ports



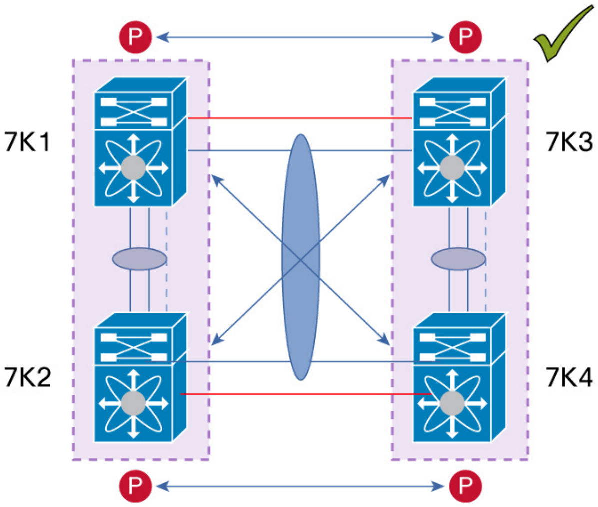
In this design, L3 device is attached to vPC domain through 2 different types of links: L2 links and L3 links.

L2 links are used for bridged traffic (traffic staying in the same VLAN) or inter-VLAN traffic (assuming vPC domain host the interface VLAN and associated HSRP configuration).

L3 links are typically used for routing protocol peering adjacency with each vPC peer device.

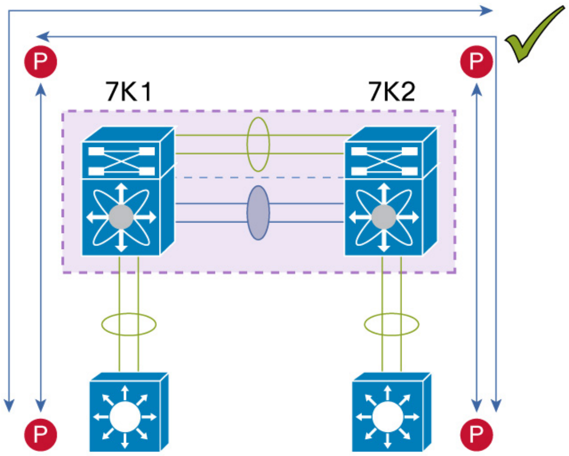
Purpose is to attract specific traffic to go through the L3 device (firewall in routed mode can be one application). L3 links are also used to carry routed traffic from L3 device to vPC domain.

Figure 57. Layer 3 and vPC Supported Designs: Peering over a vPC Interconnection (DCI case) on Parallel Interconnected Routed Ports



This type of design deals with DCI use-case. In case routing protocol peering adjacency needs to be established between the 2 data centers, recommendation is to add dedicated L3 links between the 2 sites as depicted in figure 59. vPC link between the 2 data centers still carry bridged traffic or inter-VLAN traffic while the dedicated L3 links carry routed traffic across the 2 sites.

Figure 58. Layer 3 and vPC Supported Designs: Peering over PC Interconnection and Dedicated Interswitch Link Using non-vPC VLAN



In case L3 device is single-attached to vPC domain, use non-vPC VLAN with a dedicated inter-switch link in order to establish routing protocol peering adjacency between the L3 device and each vPC peer device.

Do not use vPC VLAN (and vPC peer-link) for this purpose as this type of design is not supported.

Layer 3 and vPC Interactions: Unsupported Designs

The network topologies shown in figures 60 through 64 are unsupported designs for Layer 3 and vPC.

Figure 60 defines the symbols used in these diagrams.

Figure 59. Signification of Symbols Used in the Following Diagrams (Figures 60 Through 64)

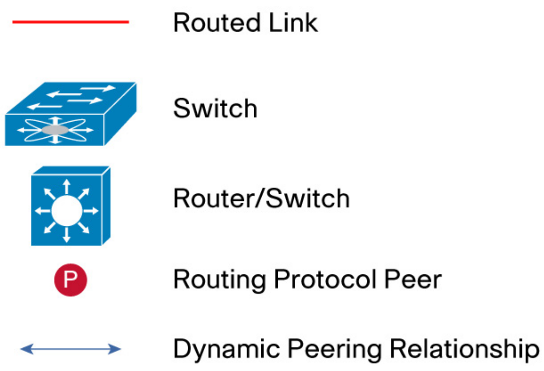
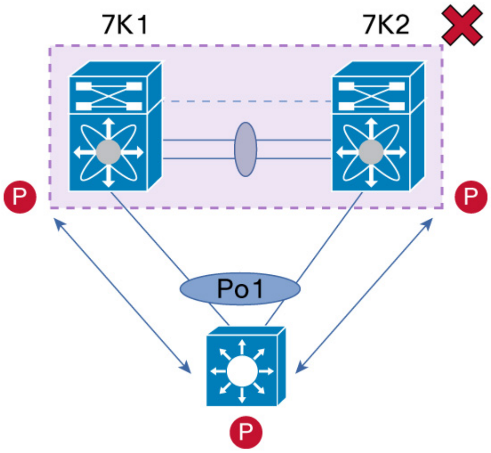
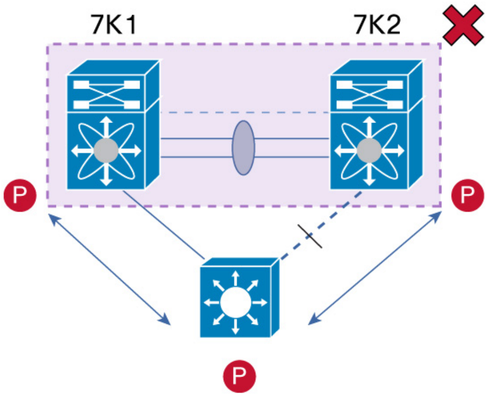


Figure 60. Layer 3 and vPC Unsupported Designs: Peering Over a vPC Interconnection



A design where L3 device is vPC-attached to vPC domain and establishes L3 routing protocol peering adjacency with each peer device is not supported. Reason is traffic may be blackholed because of vPC loop avoidance (L3 ECMP decision and L2 port-channel hashing algorithms are running independently so traffic may need to cross vPC peer-link in order to reach the next L3 hop which is a vPC peer device).

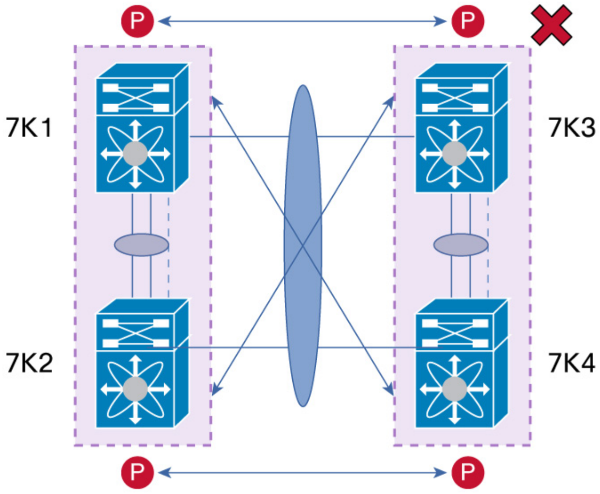
Figure 61. Layer 3 and vPC Unsupported Designs: Peering Over a Spanning Tree Protocol Interconnection Using a vPC VLAN



Connecting a L3 device to vPC domain using STP links with vPC VLAN is not a supported design.

Use rather non-vPC VLAN with dedicated inter-switch link to carry non-vPC VLAN to be fully compliant for this type of design (figure 55 of previous section)

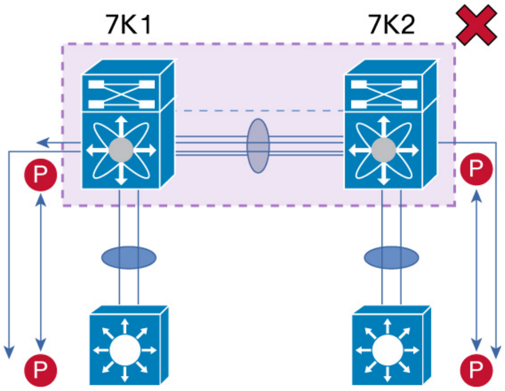
Figure 62. Layer 3 and vPC Unsupported Designs: Peering Over a vPC Interconnection (DCI Case)



In a DCI deployment, establishing routing protocol peering adjacency across the different vPC peer devices through DCI vPC (as depicted in figure 63, 7K1 peering with 7K3 and 7K2 peering with 7K4) is an unsupported design because it is subject to vPC loop avoidance rule.

Alternative supported topology is illustrated in figure 64 (previous section). By adding dedicated L3 link across the 2 data centers, this design is no more subject to vPC avoidance rule. L3 routed traffic will be carried over the L3 links and not over DCI vPC link.

Figure 63. Layer 3 and vPC Unsupported Designs: Peering Over PC Interconnection and Over vPC Peer-Link Using vPC VLAN



L3 device single-attached to vPC domain using vPC VLAN is not a supported design.

To change the design and make it fully officially supported, use non-vPC VLAN for the interconnection link between L3 device and vPC domain and add a dedicated inter-switch across the 2 vPC peer devices to carry this non-vPC VLAN. This topology is depicted in figure 59 of previous section.

vPC and L3 Backup Routing Path

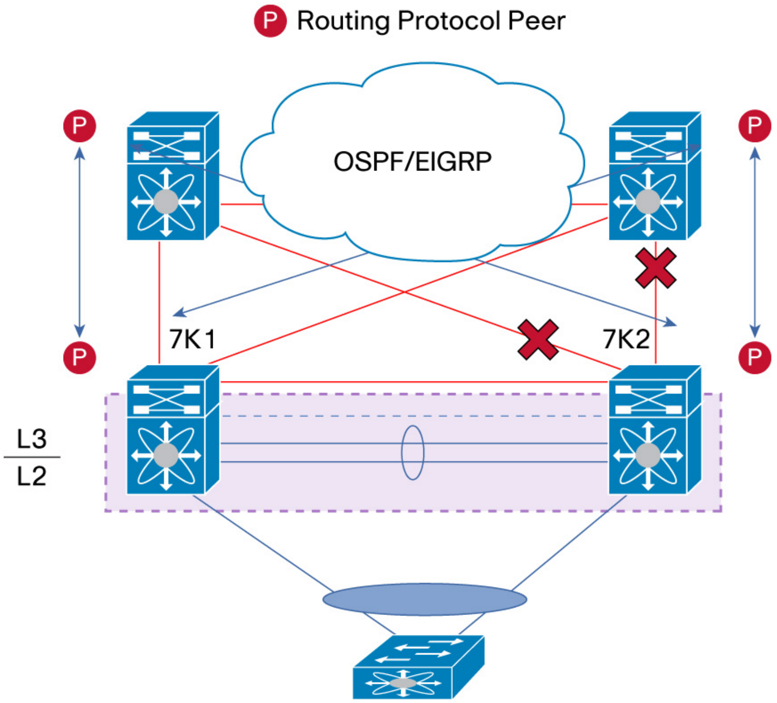
The purpose of setting up a L3 backup routing path is to establish an alternate Layer 3 path to the core in case of L3 uplinks failure (Figure 65).

L3 backup routing path (across the 2 vPC peer devices) is a point-to-point link where dynamic routing protocol peering adjacencies are established between the 2 switches/routers.

If L3 uplinks on 7K2 all fail down and routed traffic is sent from access switch to 7K2, then L3 backup routing path will be leveraged: 7K1 will receive the routed traffic from 7K2 and will be able to forward it out of its operating L3 uplinks.

L3 backup routing path offers an additional level of high availability for vPC domain in regards to L3 core reachability.

Figure 64. Establishing a L3 Backup Routing Path Between vPC Peer Devices



Strong Recommendation:

Always build L3 backup routed path for vPC domain in order to increase network resilience and availability. Use an OSPF point-to-point adjacency (or equivalent Layer 3 protocol) between the 2 vPC peer devices to establish a Layer 3 backup path to the core in case of uplink failures.

There are several ways to implement the L3 backup routing path.

Strong Recommendations:

To build L3 backup routing path, use the following options listed by descending order of preference:

- Use a dedicated Layer 3 point-to-point link between the vPC peer devices to establish a Layer 3 backup path to the core.
- Use the already existing Layer 2 port-channel trunk ISL (Inter Switch Link) for non-vPC VLAN and create dedicated VLAN/SVI to establish a Layer 3 neighborship
- Use vPC peer-link and create dedicated VLAN/SVI to establish a Layer 3 neighborship (least recommended solution)

Best Practices for HSRP/VRRP and vPC

This section describes best practices for using HSRP (Hot Standby Router Protocol) or VRRP (Virtual Router Redundancy Protocol) with vPC.

HSRP/VRRP active/active with vPC

HSRP (Hot Standby Router Protocol) and VRRP (Virtual Router Redundancy Protocol) are both network protocols that provides high availability for servers IP default gateway.

vPC domain at aggregation layer usually performs L2/L3 boundary so each vPC peer device is configure with interface VLAN (or SVI) and HSRP or VRRP runs on top of this interface.

HSRP and VRRP in the context of vPC have been improved from a functional and implementation standpoint to take full benefits of the L2 dual-active peer devices nature offered by vPC technology:

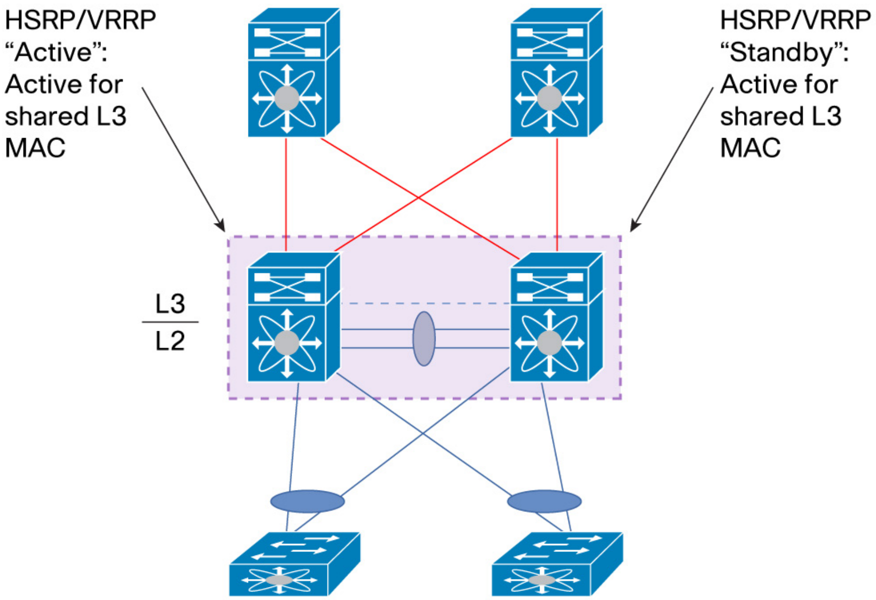
HSRP and VRRP operate in active-active mode from data plane standpoint, as opposed to classical active/standby implementation with STP based network.

No additional configuration is required. As soon as vPC domain is configured and interface VLAN with associated HSRP or VRRP group is activated, HSRP or VRRP will behave by default in active/active mode (on data plane side).

From a control plane standpoint, active-standby mode still applies for HSRP/VRRP in context of vPC; the active HSRP/VRRP intance responds to ARP request.

Figure 66 illustrates the active/active nature of HSRP or VRRP with vPC.

Figure 65. HSRP/VRRP Active-Active with vPC



Looking at the show hsrp group command, one vPC peer device appear as the active instance while the other vPC peer device appear as the standby instance. This is the HSRP/VRRP control plane information.

A particularity of the active HSRP/VRRP peer device is that it is the only one to respond to ARP requests for HSRP/VRRP VIP (Virtual IP). ARP response will contain the HSRP/VRRP vMAC which is the same on both vPC peer devices.

The standby HSRP/VRRP vPC peer device just relays the ARP request to active HSRP/VRRP peer device through vPC peer-link.

```
7K1# sh hsrp group 400
Vlan400 - Group 400 (HSRP-V2) (IPv4)
  Local state is Active, priority 100 (Cfged 100)
    Forwarding threshold(for vPC), lower: 1 upper: 100
  Hellotime 3 sec, holdtime 10 sec
  Next hello sent in 0.383000 sec(s)
  Virtual IP address is 40.40.40.254 (Cfged)
  Active router is local
  Standby router is 40.40.40.2, priority 100 expires in 7.386000 sec(s)
  Authentication text "cisco"
  Virtual mac address is 0000.0c9f.f190 (Default MAC)
  2 state changes, last state change 6d01h
  IP redundancy name is hsrp-Vlan400-400 (default)
```

```
7K2# sh hsrp group 400
Vlan400 - Group 400 (HSRP-V2) (IPv4)
  Local state is Standby, priority 100 (Cfged 100)
    Forwarding threshold(for vPC), lower: 1 upper: 100
  Hellotime 3 sec, holdtime 10 sec
  Next hello sent in 0.848000 sec(s)
  Virtual IP address is 40.40.40.254 (Cfged)
  Active router is 40.40.40.1, priority 100 expires in 7.852000 sec(s)
  Standby router is local
  Authentication text "cisco"
  Virtual mac address is 0000.0c9f.f190 (Default MAC)
  7 state changes, last state change 01:08:24
  IP redundancy name is hsrp-Vlan400-400 (default)
```

From the data plane perspective, both peer devices are forwarding. This is implemented by imposing the G bit (Gateway bit) for HSRP/VRRP vMAC in the MAC address table on both vPC peer devices, as shown below:

```
7K1# sh mac address-table address 0000.0c9f.f190
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
G 400      0000.0c9f.f190    static       -       F    F  sup-eth1(R)
```

```
7K2# sh mac address-table address 0000.0c9f.f190
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
G 400      0000.0c9f.f190    static       -       F    F  vPC Peer-Link(R)
```

Note that on active HSRP/VRRP instance, the vMAC points to sup-eth1(R) while on standby HSRP/VRRP instance, vMAC points to vPC peer-link. This is a quick way to recognize which switch is HSRP/VRRP active or standby (from control plane perspective).

HSRP/VRRP Guidelines and Restrictions

When using HSRP/VRRP within a vPC domain, follow these recommended best practices:

Strong Recommendations:

- When running HSRP/VRRP in active-active mode (data plane standpoint), aggressive timers can be relaxed: use the default HSRP/VRRP timers.
- Define the SVI associated with FHRP/VRRP as passive routing interface in order to avoid forming routing adjacency over vPC peer-link.
- Define vPC primary peer device as the active HSRP/VRRP instance and vPC secondary peer device as standby HSRP/VRRP (from control plane standpoint) for ease of operations.
- Disable ip redirect on the interface VLAN where HSRP/VRRP is configured (command is **no ip redirect**). This is a general best practice related to HSRP/VRRP.

Note: In the event of a data center outage, if the HSRP is enabled before the vPC has successfully come up, traffic loss can occur. You need to enable an HSRP delay timer to give the vPC enough time to stabilize. If you enable both HSRP delay timer and HSRP preemption delay, the Cisco Nexus 7000 Series devices will allow Layer 3 switching only after both these timers expire.

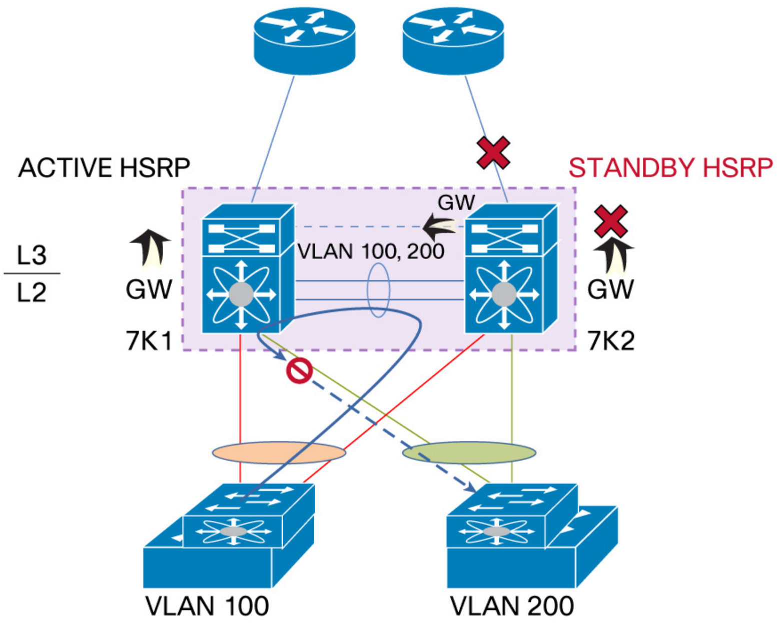
vPC and HSRP/VRRP Object Tracking

As Figure 67 suggests, it’s important not to use HSRP/VRRP link tracking in a vPC configuration.

Assume HSRP/VRRP object tracking is configured on both vPC peer devices and L3 uplink failure occurs on switch 7K2. This event triggers the HSRP/VRPP object tracking and the resulting SVI with associated HSRP/VRRP configuration is set to DOWN state. So everytime 7K2 receives a frame destined to HSRP/VRRP vMAC, it bridges this frame over vPC peer-link because the other vPC peer device is able to process this frame (as SVI with associated HSRP/VRRP configuration is still in UP state).

Using vPC with HSRP/VRRP object tracking may leads to traffic blackholing in case object tracking is triggered: the reason is that vPC systems will not forward a packet back on a vPC once it has crossed the peer-link (because of the vPC loop avoidance rule), except in the case of a remote vPC member port failure.

Figure 66. The Problem with HSRP/VRRP Object Tracking in a vPC Configuration



Strong Recommendation:

Do not use HSRP/VRRP object tracking in a vPC domain.

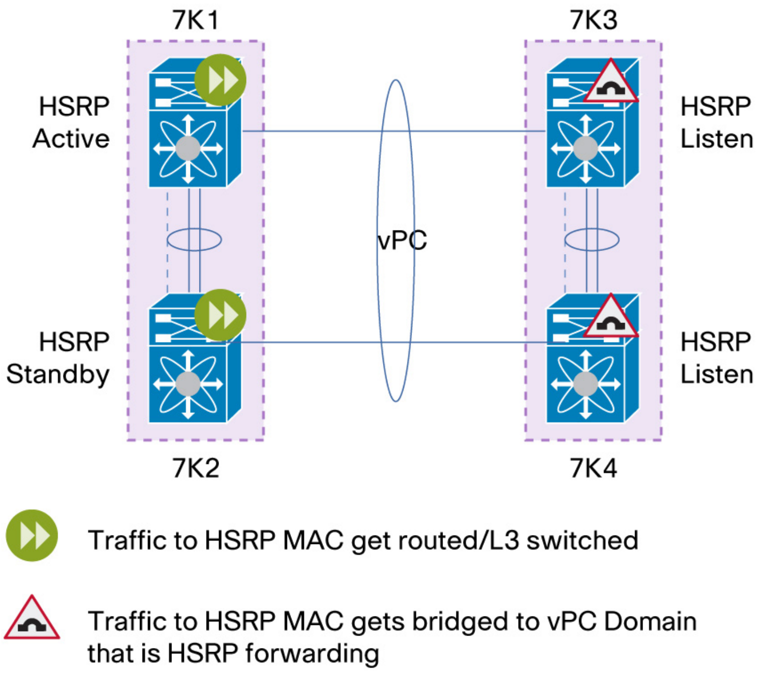
vPC and HSRP/VRRP in the Context of DCI

Some enhancements have been brought to HSRP/VRRP in the context of vPC for DCI.

A single HSRP/VRRP group can be created across the 2 data centers and HSRP/VRPP supports for Active/Active (data plane perspective) on one pair, and still allows normal HSRP behavior on other pair (all in one HSRP group): Inter-VLAN traffic and Layer 3 traffic will run across DCI vPC link for non Active/Active Layer 3 pair.

Figure 68 and 69 depicts 2 possible scenari for HSRP mode on each data center.

Figure 67. DCI with a Single HSRP/VRRP Group - Active/Active in DC1 and Listen/Listen in DC2

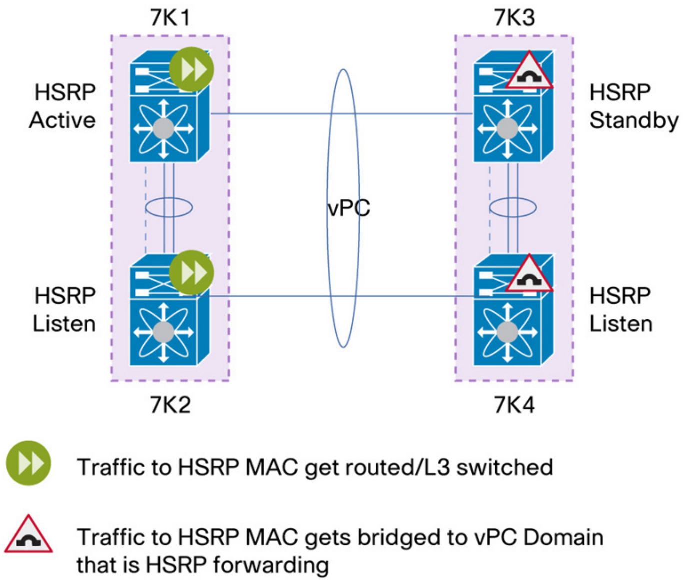

7K1
7K3
HSRP Active
HSRP Listen
vPC
HSRP Standby
HSRP Listen
7K2
7K4
Traffic to HSRP MAC get routed/L3 switched
Traffic to HSRP MAC gets bridged to vPC Domain that is HSRP forwarding

In scenario 1, 7K1 and 7K2 (part of data center 1) are configured with HSRP active/standby mode (control plane perspective). From data plane standpoint, they are both active/active.

7K3 and 7K4 (part of data center 2) are configured with HSRP listen/listen mode.

7K1 and 7K2 form vPC domain that is HSRP forwarding. Traffic to HSRP vMAC received by 7K3 or 7K4 are bridged over DCI vPC link in destination to 7K1 or 7K2 for L3 lookup and forwarding.

Figure 68. DCI with a Single HSRP/VRRP Group - Active/Active in DC1 and Listen/Listen in DC2


7K1
7K3
HSRP Active
HSRP Standby
vPC
HSRP Listen
HSRP Listen
7K2
7K4
Traffic to HSRP MAC get routed/L3 switched
Traffic to HSRP MAC gets bridged to vPC Domain that is HSRP forwarding

In scenario 2, 7K1 and 7K2 (part of data center 1) are configured with HSRP active/listen mode (control plane perspective). From data plane standpoint, they are both active/active.

7K3 and 7K4 (part of data center 2) are configured with HSRP standby/listen mode.

7K1 and 7K2 form vPC domain that is HSRP forwarding. Traffic to HSRP vMAC received by 7K3 or 7K4 are bridged over DCI vPC link in destination to 7K1 or 7K2 for L3 lookup and forwarding.

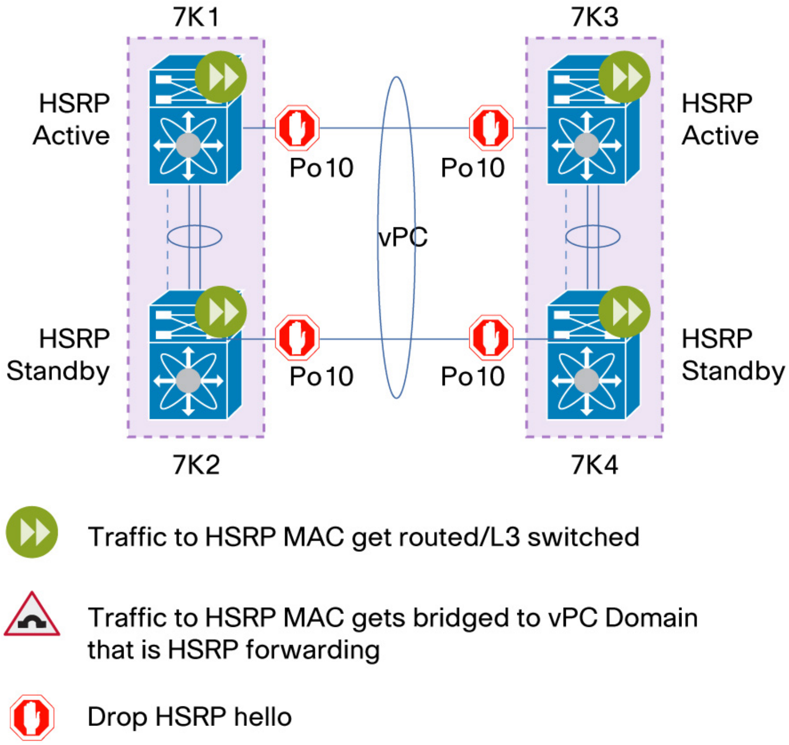
This mode of operation for HSRP/VRRP in the context of vPC for DCI is set by default (DCI with single HSRP/VRRP group). There is no specific configuration to implement in order to get this behavior.

In case HSRP (or VRRP) active/active (data plane perspective) is desired on both data centers (to avoid bridged traffic over DCI vPC link for packets destined to HSRP/VRRP vMAC), there is a technical solution to implement this type of operation. The solution is based on PACL (Port ACL) activation on DCI vPC link to stop propagation of HSRP/VRRP hello messages over this link.

Figure 70 illustrates the active/active (data plane perspective) behavior for HSRP/VRRP on both data centers and the application of PACL over the DCI vPC link to obtain this type of design.

Typical application leveraging HSRP (or VRRP) active/active on both data centers is VMOTION (virtual server move from 1 physical server to an another physical server) at L2.

Figure 69. DCI with a Single HSRP/VRRP Group - Active/Active in DC1 and Active/Active in DC2


7K1
7K3
HSRP Active
HSRP Active
Po10
Po10
vPC
HSRP Standby
HSRP Standby
Po10
Po10
7K2
7K4
Traffic to HSRP MAC get routed/L3 switched
Traffic to HSRP MAC gets bridged to vPC Domain that is HSRP forwarding
Drop HSRP hello

Note:

HSRP Hello causes vMAC to flap between local interface and layer 2 vPC DCI under the following conditions:

1. FHRP isolation PACL is configured on a Layer 2 vPC DCI interface.
2. Device acting as a Layer2 vPC DCI is not an FHRP gateway.

This can be overcome with the following two options depending on the design of the network:

- If the DCI device is only connected to the FHRP gateway, a VACL with an ARP inspection filter is recommended to isolate the data centers.
- If the DCI device has connections to other devices in the local data center, use the PACL with a static MAC entry. This will not stop the duplicate gateway gratuitous ARPs between the two sites.

PACL configuration to stop HSRPv1 hello messages:

```
ip access-list HSRPv1_Filtering
  10 deny udp any 224.0.0.2/32 eq 1985
  20 permit ip any any
```

PACL configuration to stop HSRPv2 hello messages:

```
ip access-list HSRPv2_Filtering
  10 deny udp any 224.0.0.102/32 eq 1985
  20 permit ip any any
```

PACL configuration to stop VRRP hello messages:

```
ip access-list VRRP_Filtering
  10 deny udp any 224.0.0.18/32 eq 1985
  20 permit ip any any
```

To apply the PACL to DCI vPC link, apply the PACL on each member ports (example with HSRPv1):

```
Interface Po10
  ip port access-group HSRPv1_Filtering
```

Best Practices for Network Services and vPC

This section describes best practices for network services integration with vPC. Network services devices include appliances or service modules like load balancers and firewalls.

Network Services Chassis with VDC Sandwich Design

Network Services can be deployed as part of the Cisco Catalyst 6500 Series Services chassis. A dedicated 6500 is used to host ASA service module or FWSM (FireWall Service Module) or ACE service module.

The 6500 service chassis provides firewalling and server load-balancing functionalities to the data center.

ASA or ACE can run in different modes (routed mode, transparent mode, one-arm mode).

However, for the design mentioned below, only transparent mode is supported.

The 6500 Services chassis provides port-channel (or etherChannel) capabilities for interaction with the vPC.

All services modules are configured in transparent mode. Multiple transparent contexts can be used if needed.

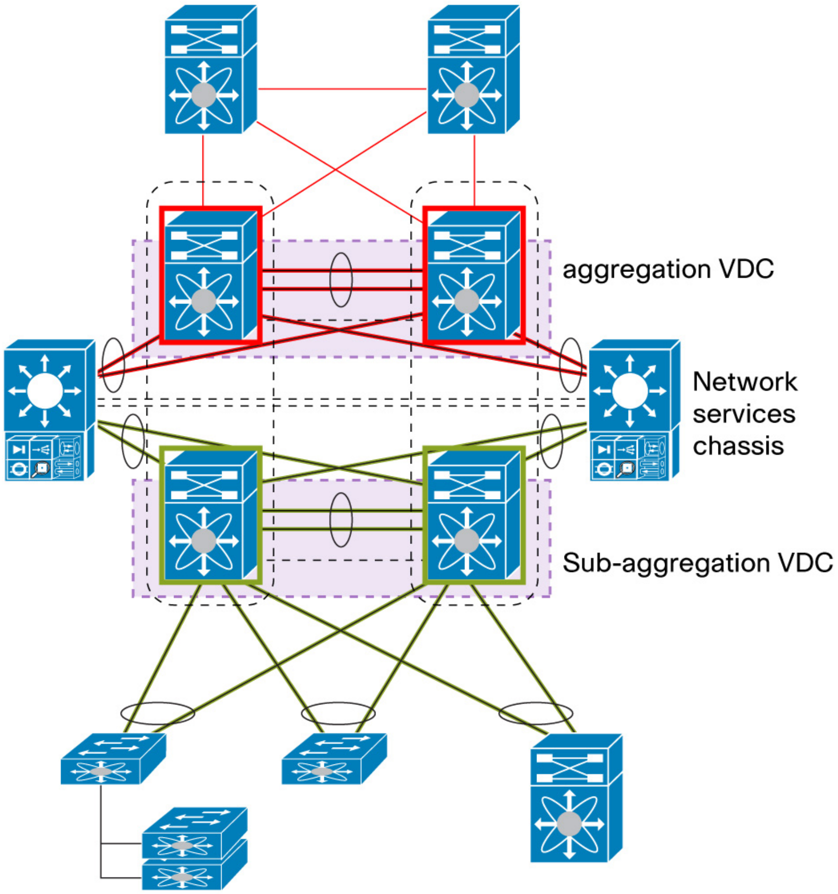
Two Cisco Nexus 7000 Series VDCs are used to “sandwich” services between switching layers. vPC runs in both VDC pairs to provide port-channels for both inside and outside interfaces to the Services chassis (Figure 71).

Here are some important design considerations for the Services chassis integration with vPC, as shown in Figure 71:

- Access switches requiring services are connected to the subaggregation VDC (vPC domain at the bottom [in green])
- Access switches not requiring services may be connected to the aggregation VDC (vPC domain at the top [in red])

If peering at Layer 3 between the network services chassis and the 2 vPC peer devices is required, an alternative design should be explored (for example, using Spanning Tree Protocol rather than vPC to attach Services chassis). This is typically the case if service modules within the network services chassis are configured in routed mode.

Figure 70. Network Services Chassis with VDC Sandwich Design



When configuring network service modules to run with vPC, follow these recommended best practices:

General Recommendations:

- Configure two Cisco Nexus 7000 Series VDCs to insert services between the virtual switching layers.
- Configure service modules within the network services chassis in transparent mode.
- Use port-channel (or etherchannel) capability provided by the services chassis for interaction with the vPC domains.
- Configure vPC domains to run in both VDC pairs to provide port-channel connections for both inside and outside interfaces to the services chassis.
- Connect access switches that require services to the subaggregation VDC layer.
- Connect access switches that do not require services to an aggregation VDC layer.
- If needed, use multiple VRF instances in the subaggregation VDC layer to extend support for multiple virtualized service contexts.
- Be aware of the following Layer 3 over vPC design caveat; if peering at Layer 3 is required across the two vPC layers, consider using an alternative solution, such as Spanning Tree Protocol rather than vPC to attach to the services chassis.

Network Services Appliances in Transparent Mode with vPC

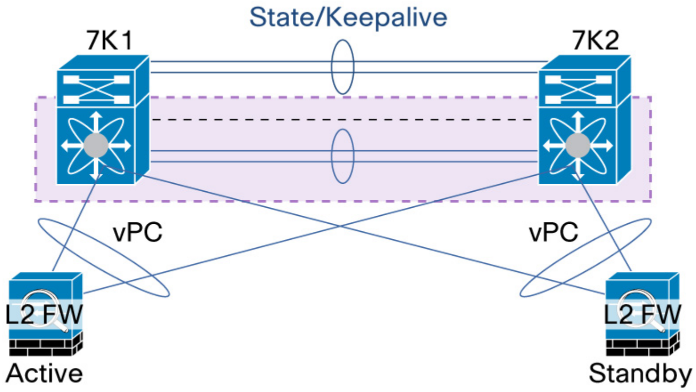
Services appliance in transparent mode integration with vPC is a straightforward implementation. There is no particular caveat for this type of design.

Service appliance must support port-channel capability as well as VLAN translation.

Because service appliance in transparent mode do not need to establish any kind of L3 peering adjacency with the 2 vPC peer devices, it makes this design very easy to deploy.

Figure 72 shows this kind of topology

Figure 71. Network Services Appliance Configured in Transparent Mode Connected to vPC Domain



Service appliance in transparent mode behaves like a bump in the wire and bridge traffic from inside interface to outside interface by swapping the VLAN id. Note that ingress VLAN and egress VLAN are associated to the same IP subnet.

Interface VLAN (i.e SVI) remain on vPC domain which still perform L2/L3 boundary (interface VLAN defined on vPC domain are used as default gateway for servers connected underneath.

When configuring network service appliance in transparent mode with vPC, follow these recommended best practices:

Strong Recommendations:

- Use port-channel capability provided by the services appliance for interaction with the vPC domains.
- Dedicate a Layer 2 port-channel for the service appliances state and keepalive VLANs (we recommend that you do not use a vPC peer-link)

Configuring Cisco ASA Service Appliance in Transparent Mode with vPC

Since Release 8.4, Cisco ASA 5500 Series Adaptive Security Appliance solution supports Link Aggregation Control Protocol (LACP). ASA port-channel contains up to eight active member ports.

Supported LACP modes are: ACTIVE, PASSIVE, and ON (ON means manual ports bundling i.e not using dynamic port-channeling control protocol).

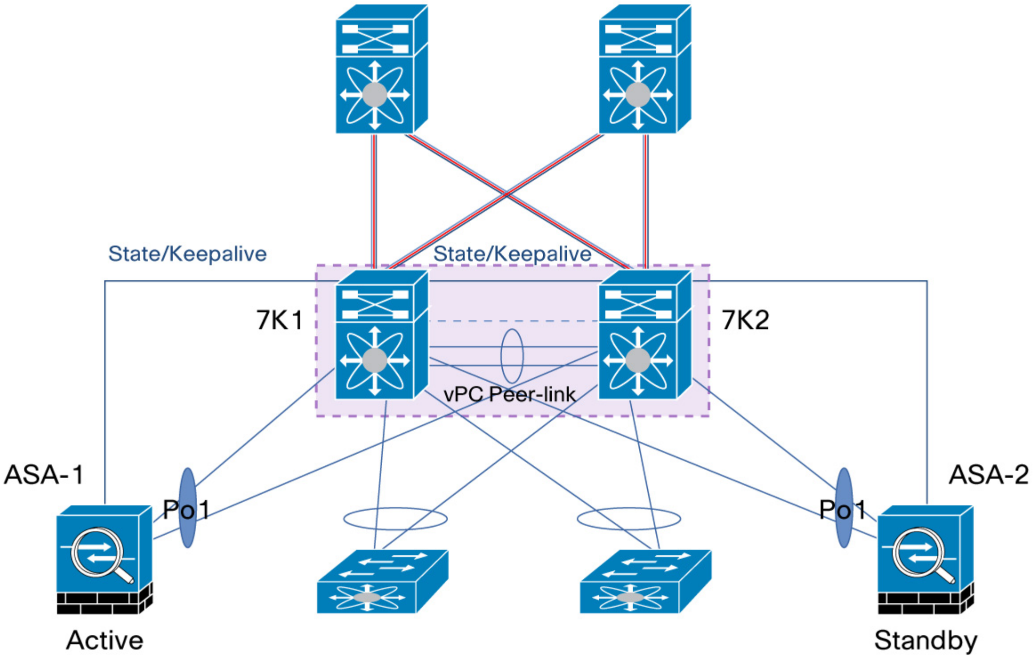
Port-channel (or EtherChannel) link is treated just like physical and logical interfaces on the Cisco ASA appliance.

ASA can be configured in transparent or routed mode. Both modes are supported when integrating ASA with Cisco Nexus 7000 Series vPC.

This section describes how to configure ASA in transparent mode in the context of vPC: ASA device is connected to a vPC domain using vPC link.

Let’s take topology represented in Figure 73 are reference.

Figure 72. ASA Services Appliance Configured in Transparent Mode and Connected to vPC Domain

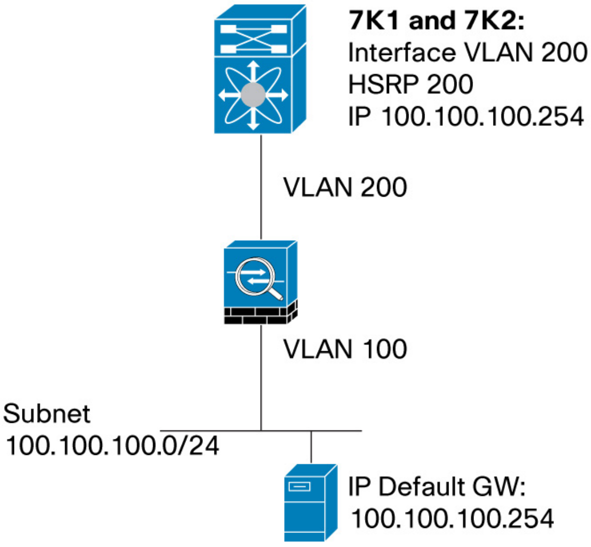


ASA-1 and ASA-2 run in HA (High Availability): ASA-1 is operating in active mode while ASA-2 is in standby mode. ASA-1 will process all flows coming from the network. If ASA-1 is down, ASA-2 will become active and then will be able to process subsequent flows.

VLAN 100 is used at the ingress side (or inside interface) and VLAN 200 is used at the egress side (outside interface). They are both associated to IP subnet 100.100.100.0/24.

Figure 74 represents the logical topology view of the network with ASA.

Figure 73. ASA in Transparent Mode Connected to vPC Domain - Logical View



As Figure 74 shows, the default gateway (HSRP) is hosted on Cisco Nexus 7000 Series.

If the server is on VLAN 100, the default gateway (GW) is hosted on VLAN 200 (FHRP on VLAN 200).

The reason is because ASA firewall performs VLAN translation from inside interface to outside interface.

Below is a sample port-channel configuration for ASA in transparent mode:

```
ASA configuration:

interface GigabitEthernet0/0
 channel-group 1 mode active
 no nameif
 no security-level
 no ip address
!
interface GigabitEthernet0/2
 channel-group 1 mode active
 no nameif
 no security-level
 no ip address
!
interface Port-channel1
 port-channel load-balance vlan-src-dst-ip
 no nameif
 no security-level
 no ip address
!
interface Port-channel1.100
 vlan 100
 nameif inside
 bridge-group 1
 security-level 99
!

interface Port-channel1.200
 vlan 200
 nameif outside
 bridge-group 1
 security-level 1
!

interface BVI1
 ip address 100.100.100.5 255.255.255.0 standby 100.100.100.6
!
```

Note: configuration for ASA state/keepalive link is not shown in the above example.

On NEXUS 7000 side (7K1 and 7K2), configurations for vPC member ports connected to ASA are the following:

```
7K1 and 7K2 configuration - vPC member ports connected to ASA-1:

interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 100,200
  vpc 1
```

```
7K1 and 7K2 configuration - vPC member ports connected to ASA-2:

interface port-channel2
```

```
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 100,200
  vpc 2
```

Interface VLAN 200 (i.e SVI 200) configuration on NEXUS 7000 (7K1 and 7K2) is the following:

```
7K1 configuration:

interface Vlan200
  ip address 100.100.100.1/24
  no ip redirect
  hsrp 200
    ip 100.100.100.254
  no shutdown
```

```
7K2 configuration:

interface Vlan200
  ip address 100.100.100.2/24
  no ip redirect
  hsrp 200
    ip 100.100.100.254
  no shutdown
```

It is recommended that ASA port-channel hashing algorithm and Cisco Nexus vPC hashing algorithm are the same on both sides. This allows upstream and downstream traffic to use the same port-channel member port (apply the same load balancing hashing algorithm all along the path [meaning up to L3 core] to have consistent result). Use the following show commands to check the configured load-balancing hashing algorithms on ASA firewall and NEXUS 7000 vPC peer devices:

```
ASA-1# sh port-channel 1 load-balance
EtherChannel Load-Balancing Configuration:
        vlan-src-dst-ip-port

EtherChannel Load-Balancing Addresses UsedPer-Protocol:
Non-IP: Source XOR Destination MAC address
  IPv4: Vlan ID and Source XOR Destination IP address and TCP/UDP (layer-4)port
number
  IPv6: Vlan ID and Source XOR Destination IP address and TCP/UDP (layer-4)port
number
```

```
7K1# sh port-channel load-balance

Port Channel Load-Balancing Configuration:
System: src-dst ip-l4port-vlan

Port Channel Load-Balancing Addresses Used Per-Protocol:
Non-IP: src-dst mac
IP: src-dst ip-l4port-vlan
```

Network Services Appliances in Routed Mode with vPC

Services appliance in routed mode connected to vPC domain using L3 links is straightforward design and works perfectly without special care about vPC configuration. This is the recommended way to connect service appliance in routed mode to a vPC domain.

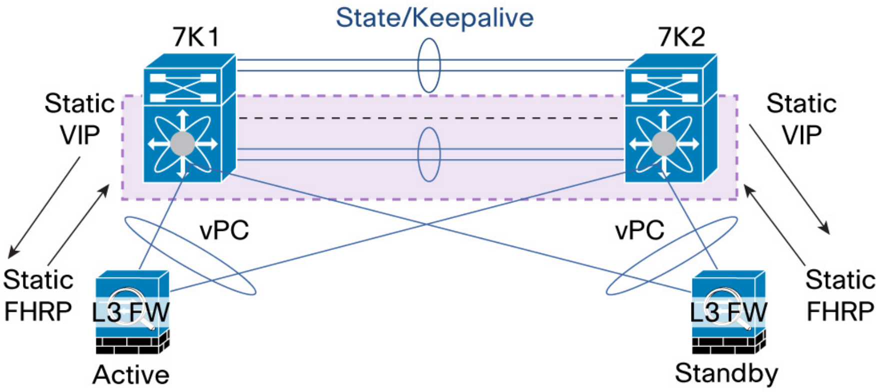
Strong Recommendation:

Connect a service appliance in routed mode to a vPC domain using L3 links if possible. This is the strongest recommended way to connect the service appliance to vPC domain.

However, if previous recommendation cannot be met, it is still possible to connect the service appliance to vPC domain using a vPC link or a single link between the 2 entities. This can be done only if specific design guidelines are followed. Design objective is to avoid the L3 over vPC issue.

There are 2 possible ways to connect the service appliance to vPC domain using vPC link or single link. They are illustrated in figure 75 and 76.

Figure 74. Service Appliance in Routed Mode Connected to vPC Domain Using vPC Link



In this design (figure 75), L3 service appliance is vPC-attached to each vPC peer device.

L3 service appliance is running in active-standby mode: one device is in active state while the other device is in standby mode.

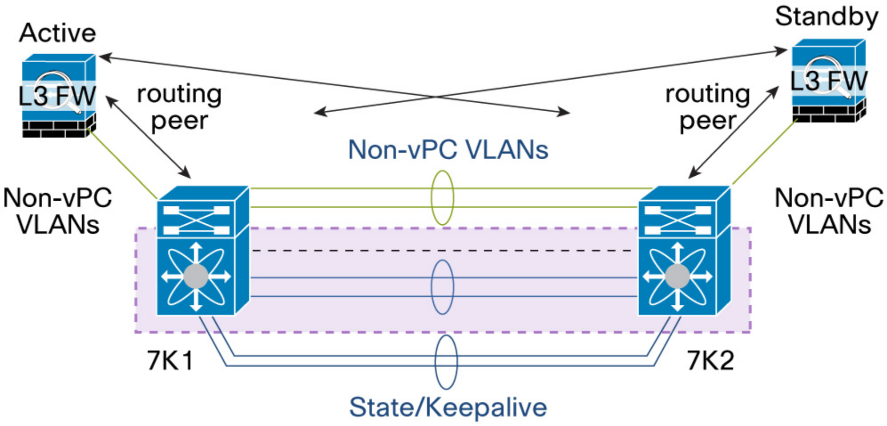
To avoid falling into the vPC loop avoidance situation, use the following rules to set correctly L3 adjacency between L3 service appliance and vPC domain:

On L3 service appliance, create a static route (it can be the default route) pointing to HSRP/VRRP VIP (Virtual IP) defined on vPC domain. This way, L3 service appliance (whichever one who is in active state) can send traffic to any vPC peer device. As both vPC peer devices are HSRP active (from data plane standpoint), they will be able to route traffic coming from the L3 service appliance.

For the return traffic from vPC domain to L3 service appliance, create a static route on each vPC peer device pointing to static VIP defined on L3 service appliance. L3 service appliance VIP is identical for both active and standby instances. Only the active L3 service appliance instance owns the VIP (i.e process packets destined to VIP address).

In this design, L3 service appliance usually host the default IP gateway for servers connected to vPC domain, HSRP/VRRP group created on vPC domain is just used for L3 connectivity between L3 service appliance and each vPC peer device.

Figure 75. Service Appliance in Routed Mode Connected to vPC Domain Using Single Link



In this design (figure 76), L3 service appliance is single-attached to each vPC peer device.

As commonly recommended with device single-attached to vPC domain, use non-vPC VLAN for this purpose and add a dedicated inter-switch link to carry non-vPC VLAN traffic across the 2 vPC peer devices.

L3 service appliance can establish L3 routing protocol peering adjacency with both vPC peer devices over this dedicated infrastructure without any issues.

In this design, both vPC peer devices host the default IP gateway for servers connected to the vPC domain.

L3 service appliance will then process traffic from or to servers depending on the L3 path defined by routing policy.

When configuring network service appliance in routed mode connected to vPC domain, follow these recommended best practices:

Strong Recommendations:

- Dedicate a Layer 2 port-channel for the service appliances state and keepalive VLANs (we recommend that you do not use a vPC peer-link).
- Connect service appliances to vPC domain via vPC and configure static routes to the HSRP/VRRP address on the service appliance side. On the Cisco Nexus 7000 Series side, create a static route pointing to the VIP of service appliance (figure 75).
- Connect service appliances to each vPC peer device through a single link using non-vPC VLAN. Implement a separate inter-switch layer 2 port-channel for non-vPC VLAN traffic across the 2 vPC peer devices (figure 76).

Configuring Cisco ASA Service Appliance in Routed Mode with vPC

Since Release 8.4, Cisco ASA 5500 Series Adaptive Security Appliance solution supports Link Aggregation Control Protocol (LACP). ASA port-channel contains up to eight active member ports.

Supported LACP modes are: ACTIVE, PASSIVE, and ON (ON means manual ports bundling i.e not using dynamic port-channeling control protocol).

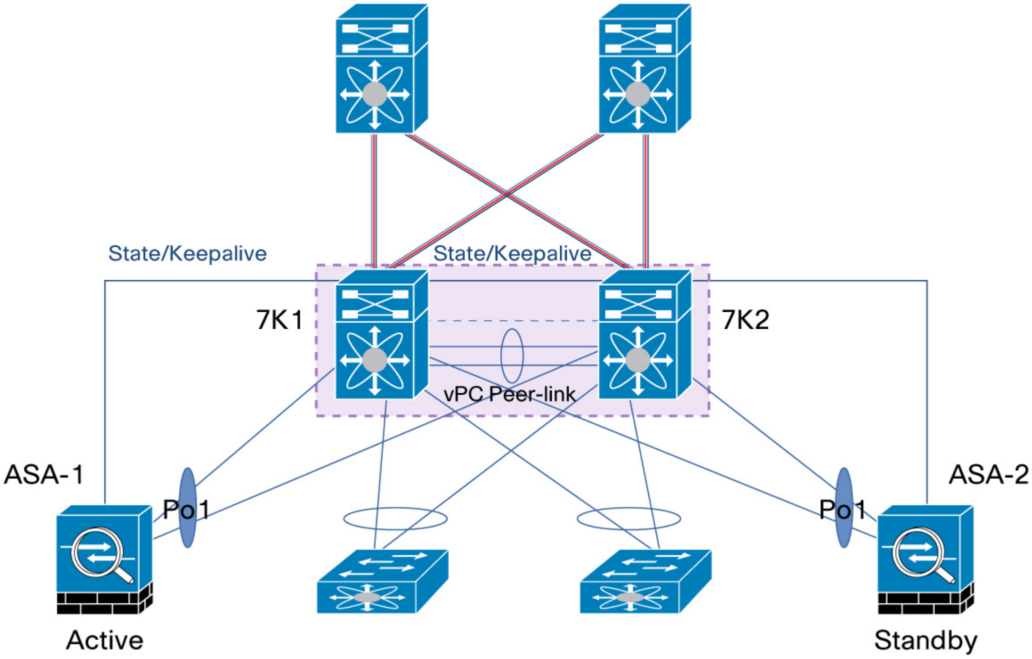
Port-channel (or EtherChannel) link is treated just like physical and logical interfaces on the Cisco ASA appliance.

ASA can be configured in transparent or routed mode. Both modes are supported when integrating ASA with Cisco Nexus 7000 Series vPC.

This section describes how to configure ASA in routed mode in the context of vPC: ASA device is connected to a vPC domain using vPC link.

Let’s take topology represented in figure 77 are reference

Figure 76. ASA Services Appliance Configured in Routed Mode and Connected to vPC Domain



ASA-1 and ASA-2 run in HA (High Availability): ASA-1 is operating in active mode while ASA-2 is in standby mode. ASA-1 will process all flows coming from the network. If ASA-1 is down, ASA-2 will become active and then will be able to process subsequent flows.

VLAN 100 is used at the ingress side (or inside interface) and VLAN 200 is used at the egress side (outside interface). VLAN 100 is associated to IP subnet 100.100.100.0/24 while VLAN 200 is associated to IP subnet 200.200.200.0/24.

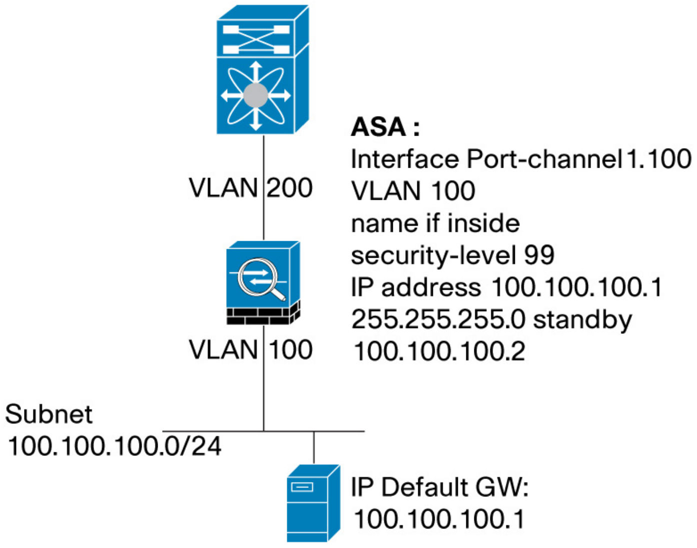
ASA host the default IP gateway for servers connected to vPC domain. In this design, port-channel 1 is configured with sub-interface Po1.100. VLAN 100 is carried inside Po1.100 and the sub-interface is configured with IP address 100.100.100.1/24 which is the IP address that servers will use as default gateway.

Note that sub-interface Po1.100 is defined as the inside interface (i.e the most secured interface on the ASA firewall).

Port-channel 1 is also configured with sub-interface Po1.200. VLAN 200 is carried inside Po1.200 and the sub-interface is configured with IP address 200.200.200.1/24. Sub-interface Po1.200 is defined as the outside interface (i.e the less secured interface on the ASA firewall).

Figure 78 represents the logical topology view of the network with ASA.

Figure 77. ASA in Routed Mode Connected to vPC Domain - Logical View



Below is a sample port-channel configuration for ASA in routed mode:

```
ASA configuration:

interface GigabitEthernet0/0
 channel-group 1 mode active
 no nameif
 no security-level
 no ip address
!
interface GigabitEthernet0/2
 channel-group 1 mode active
 no nameif
 no security-level
 no ip address
!
interface Port-channel1
 port-channel load-balance vlan-src-dst-ip
 no nameif
 no security-level
 no ip address
!
interface Port-channel1.100
 vlan 100
 nameif inside
 security-level 99
 ip address 100.100.100.1 255.255.255.0 standby 100.100.100.2
!
interface Port-channel1.200
 vlan 200
 nameif outside
 security-level 1
 ip address 200.200.200.1 255.255.255.0 standby 200.200.200.2
!
```

Note: Configuration for ASA state/keepalive link is not shown in the above example.

On NEXUS 7000 side (7K1 and 7K2), configurations for vPC member ports connected to ASA are the following:

```
7K1 and 7K2 configuration - vPC member ports connected to ASA-1:

interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 100,200
  vpc 1
```

```
7K1 and 7K2 configuration - vPC member ports connected to ASA-2:
```

```
interface port-channel2
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 100,200
  vpc 2
```

L3 configuration for ASA and NEXUS 7000 (7K1 and 7K2) are shown below:

```
ASA configuration:

route outside 0.0.0.0 0.0.0.0 200.200.200.200 1
```

```
7K1 configuration:

interface Vlan200
  ip address 200.200.200.10/24
  no ip redirect
  hsrp 200
    ip 200.200.200.200

ip route 100.100.100.0/24 Vlan200 200.200.200.1 name ASA
```

(subnet 100.100.100.0/24 is the subnet that is serviced by the ASA firewall)

```
7K2 configuration:

interface Vlan200
  ip address 200.200.200.11/24
  no ip redirect
  hsrp 200
    ip 200.200.200.200

ip route 100.100.100.0/24 Vlan200 200.200.200.1 name ASA
```

(subnet 100.100.100.0/24 is the subnet that is serviced by the ASA firewall)

It is recommended that ASA port-channel hashing algorithm and Cisco Nexus vPC hashing algorithm are the same on both sides. This allows upstream and downstream traffic to use the same port-channel member port (apply the same load balancing hashing algorithm all along the path [meaning up to L3 core] to have consistent result). Use the following show commands to check the configured load-balancing hashing algorithms on ASA firewall and NEXUS 7000 vPC peer devices:

```
ASA-1# sh port-channel 1 load-balance
```

```
EtherChannel Load-Balancing Configuration:
        vlan-src-dst-ip-port

EtherChannel Load-Balancing Addresses UsedPer-Protocol:
Non-IP: Source XOR Destination MAC address
  IPv4: Vlan ID and Source XOR Destination IP address and TCP/UDP (layer-4)port
number
  IPv6: Vlan ID and Source XOR Destination IP address and TCP/UDP (layer-4)port
number
```

```
7K1# sh port-channel load-balance

Port Channel Load-Balancing Configuration:
System: src-dst ip-l4port-vlan

Port Channel Load-Balancing Addresses Used Per-Protocol:
Non-IP: src-dst mac
IP: src-dst ip-l4port-vlan
```

Best Practices for Multicast and vPC

Multicast traffic can be carried over a vPC domain in multiple scenari:

- Multicast source can be connected in the L2 domain behind a vPC link or in the L3 core out of vPC domain
- Multicast receiver can be connected in the L2 domain behind a vPC link or in the L3 core out of vPC domain

From a L3 multicast routing protocol standpoint, vPC fully supports PIM SM (Protocol Independent Multicast Sparse Mode). Each vPC peer device can be configured with PIM SM related commands and vPC domain can interact with L3 core configured with PIM SM as well. Multicast traffic coming from one VLAN can be routed properly to other VLAN where multicast receivers are connected.

Cisco Nexus 7000 Series support PIM Spare Mode forwarding in hardware, both for (*, G) and (S, G) mroute entries.

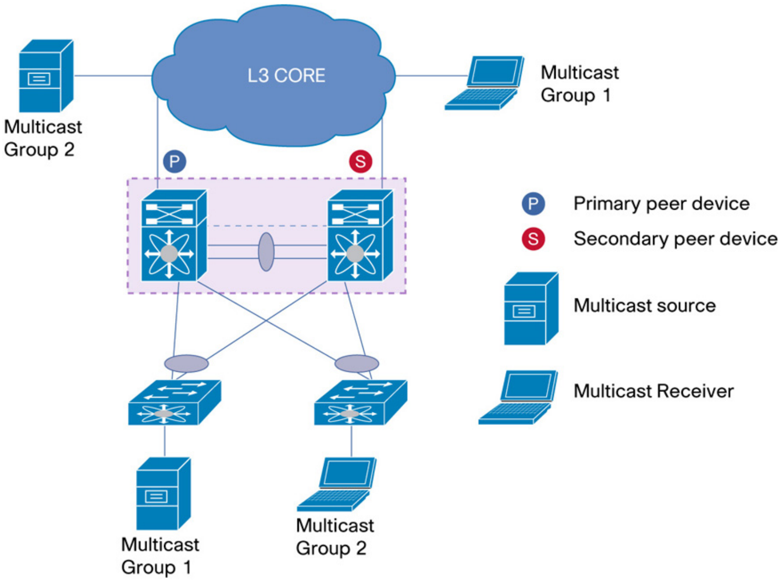
Figure 79 illustrates possible scenari for multicast with vPC (multicast sources and receivers inside or outside the vPC domain).

Note: If multicast source and multicast receivers are located on the same VLAN, there is no need to turn on dynamic multicast routing protocol.

Be aware PIM-SSM (Source Specific Multicast) and PIM Bidir (BiDirectional) are not supported with vPC.

In case one of the two multicast routing protocols is mandatory in the network, consider another implementation of L2 network design not based on vPC technology.

Figure 78. Multicast with vPC



vPC uses Cisco Fabric Services (CFS) to synchronize Internet Group Management Protocol (IGMP) states. IGMP packets are sent with information about the source interface so that the equivalent state can be established on vPC members on both the vPC peers.

For multicast sources in the vPC domain, both vPC peer devices are active forwarders; duplicates are avoided via vPC loop-avoidance logic.

Note: If one of the vPC member links goes down, then multicast receivers (configured in the vPC cloud) will receive duplicate traffic.

For multicast source is in the Layer 3 core, vPC peer device with the unicast best metric to the multicast source becomes active forwarder. Cisco Fabric Services allows the communication between the vPC peer devices to determine the active forwarder (unicast best metric to the multicast source wins). If metric is same for both vPC peer devices, the operationally primary vPC peer device becomes the active forwarder.

Note that with vPC technology, active forwarder concept is per multicast source (multicast source in the L3 core case). If multicast sources S1 and S2 do have not same unicast routing metric in respect to the vPC devices, the active forwarder will change accordingly (vPC peer device 1 can be active forwarder for S1 while vPC peer device 2 can be active forwarder for S2 as an example).

Required Recommendation:

- Use PIM SM (Protocol Independent Multicast Sparse Mode) only with vPC. PIM SSM (Source Specific Mode) and PIM BiDir (Bidirectional) do not interoperate with vPC.

General Recommendation:

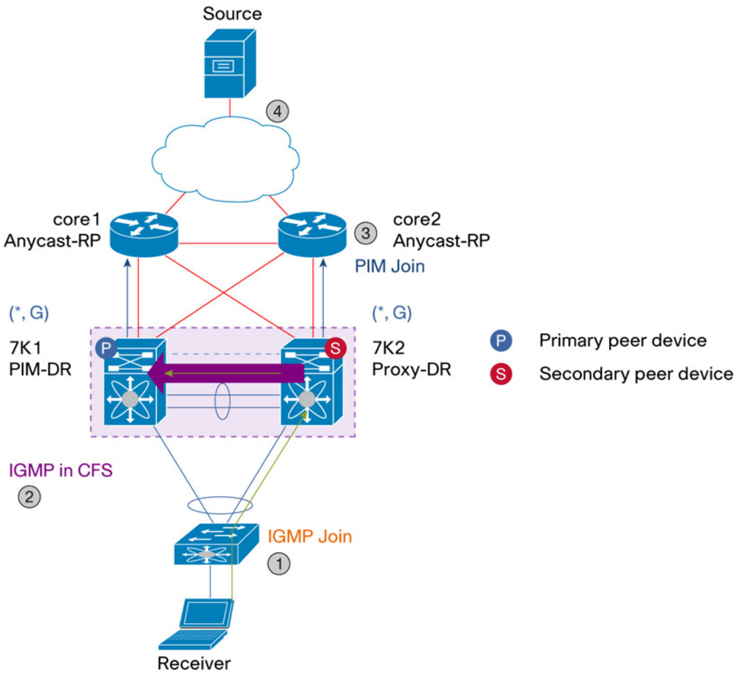
- For ease of operations, configure PIM DR (Designated Router) on vPC primary peer device.

To understand how multicast works with vPC, let's take the topology illustrated in Figure 80.

Multicast source is in L3 core and multicast receiver is in vPC domain.

Multicast PIM RP (Rendez-vous Point) is located in L3 core and routers connected to vPC peer devices are configured with anycast-RP.

Figure 79. How Multicast Works with vPC (Multicast Packet Flow)



As Figure 80 indicates, the multicast packet flow is as follows:

1. Receiver sends IGMP join, port-channel hash on access switch happens to select link to vPC peer device 2 (7K2)

   7K2 creates snooping, IGMP, and (*,G) mroute state with VPC VLAN as OIF (Outgoing Interface List)

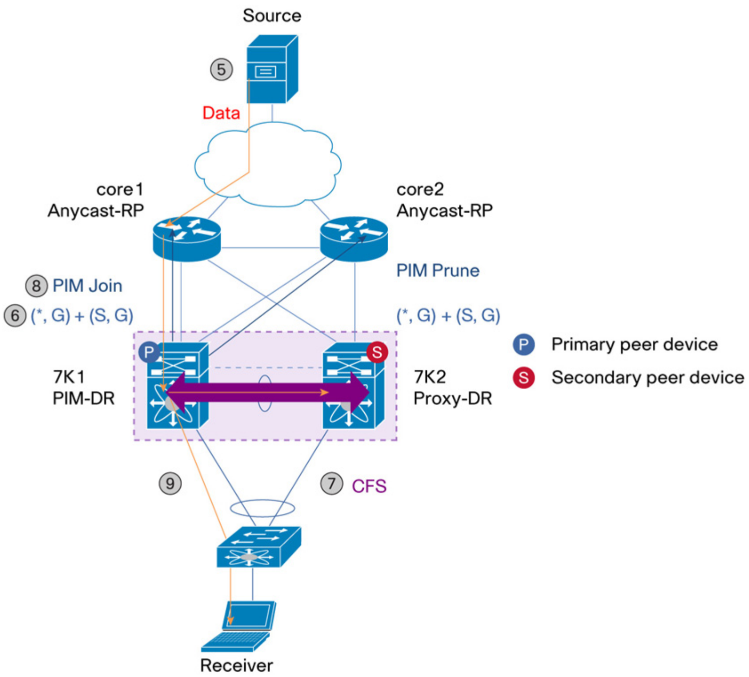
2. vPC peer device 2 (7K2) sends IGMP packet encapsulated in CFS to vPC peer device 1 (7K1)

   7K1 creates identical multicast states to vPC peer device 2 (7K2)

3. Both vPC peer devices send PIM (*,G) joins to the RP to join the RPT (RP tree - RPT is also called shared tree in the context of PIM SM)

4. If ECMP to RP, hash selects the RPF interface

Figure 80. How Multicast Works with vPC (Multicast Packet Flow) - Contd



5. Source begins transmitting. First Hop Router registers source to RPs
6. One or both VPC peers receive (S,G) traffic on shared tree (depends on upstream state)
7. vPC peer devices negotiates for active forwarder role: CFS (CISCO Fabric Services) messages exchanged to determine forwarder; best unicast routing metric to multicast source wins and in case of tie, vPC operational primary peer device wins. Let's assume vPC peer device 1 (7K1) is selected as active forwarder for this multicast source.
8. vPC peer device 1 (7K1), elected active forwarder for (S,G), sends PIM (S,G) joins toward source. It joins SPT (Shortest Path Tree), prunes RPT (RP Tree aka Shared Tree) and adds SVI associated to vPC VLAN as L3 OIF.
9. Multicast data traffic flows down the source tree to active forwarding vPC peer device.

   vPC peer device 1 (7K1) sends a copy of the multicast traffic out of the vPC member port connected to access switch. One another copy is sent out of vPC peer-link. As there is no single-attached device (or orphan port) interested by this multicast traffic on the other side, vPC peer device 2 (7K2) will drop the multicast traffic received from vPC peer-link.

Pre-building Shorted Path for Multicast with vPC (PIM pre-build-spt)

Default operation of multicast with vPC has the following characteristics:

- Convergence delay upon active forwarder change involves the new forwarder triggering upstream joins to rebuild the SPT (Shortest Path Tree). This occurs when active forwarder fails down for instance.
- May causes periodic duplicates from non-forwarder, due to forwarding on (*,G) during periodic (S,G) state expiry

Enhancement to multicast with vPC called PIM Pre-build SPT was developed to address these issues.

Objective is to create an activate/active multicast packet forwarding (Live/Live data stream) on both vPC peer devices, leveraging the dual forwarders behavior with vPC technology.

PIM Pre-build SPT on non-forwarder attracts multicast traffic by triggering upstream PIM J/Ps (Join/Prune) without setting any interfce in the OIF (Outgoing Interface) list. Multicast traffic is then always pulled to non-active forwarder and finally dropped due to no OIFs. PIM Pre-build SPT feature can be turned on/off via the following global configuration knob:

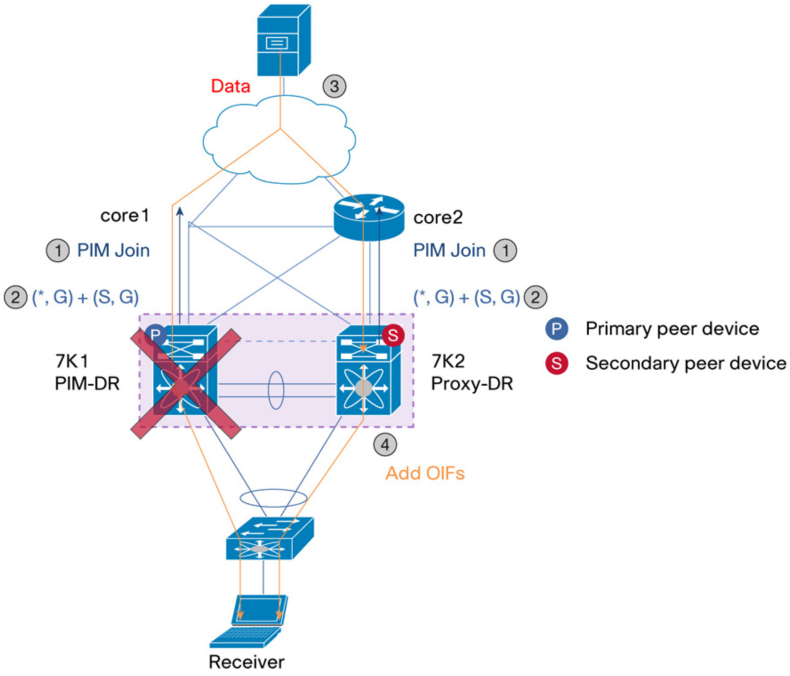
```
7K(config)# ip pim pre-build-spt
```

Immediate effect of PIM Pre-build SPT is to improve convergence time upon active forwarder failure (1 to 3 seconds of convergence time as a result). The other vPC peer device (which is non active forwarder) does not need to create any new upstream multicast state and can quickly transition to active forwarder role by programming properly the OIF (Outgoing InterFace) list.

Impact with enabling PIM prebuild SPT is to consume bandwidth and replication capacity on primary and secondary data path (i.e on vPC primary and secondary peer devices) in steady state.

Figure 82 illustrates how PIM prebuild SPT works and how vPC system reacts upon active forwarder failure event.

Figure 81. Multicast with vPC - PIM Prebuild SPT



1. Both forwarder and non-forwarder vPC peer devices join SPT (Shortes Path Tree) for new multicast sources.
2. Multicast Data traffic flows down from source tree to both vPC peer devices (7K1 and 7K2). 7K1 is the active forwarder so it propagates multicast traffic out of its vPC member port. The other vPC peer device (7K2) is the non-active forwarder so it drops multicast traffic it receives from the upstream network (OIF list is empty).

3. On failure of active forwarder (vPC peer device 1 aka 7K1), the other vPC peer device (7K2) will become the new active forwarder. 7K2 has already (S,G) state programmed in hardware and is already receiving multicast traffic, so the only action it needs to perform is to program vPC member port in the OIF list.

Strong Recommendation:

- Always enable PIM prebuild SPT when using multicast with vPC.

Best Practices for FEX and vPC

FEX (Fabric Extender aka NEXUS 2000) is a port extension technology that allows to place TOR (Top Of Rack) devices (i.e the FEX) closed to servers while maintaining a unique configuration management point which is the NEXUS 7000 device itself.

Connecting a FEX to NEXUS 7000 switch brings the following benefits:

- Benefit of TOR in term of server cabling cost as FEX is located closed to servers
- Benefit of EOR (End Of Row) in term of management as configuration and monitoring of FEX is done on the parent switch (1 unique point of management). Ports on the FEX appear on parent switch interface list and SNMP/SYSLOG information related to FEX is directly managed by NEXUS 7000 device.

FEX can be connected to the following parent line cards:

- 32 X 10-Gbps ETH ports line card (N7K-M132XP-12 and N7K-M132XP-12L)
- 48 X 1/10-Gbps ETH ports line card (N7K-F248XP-25)

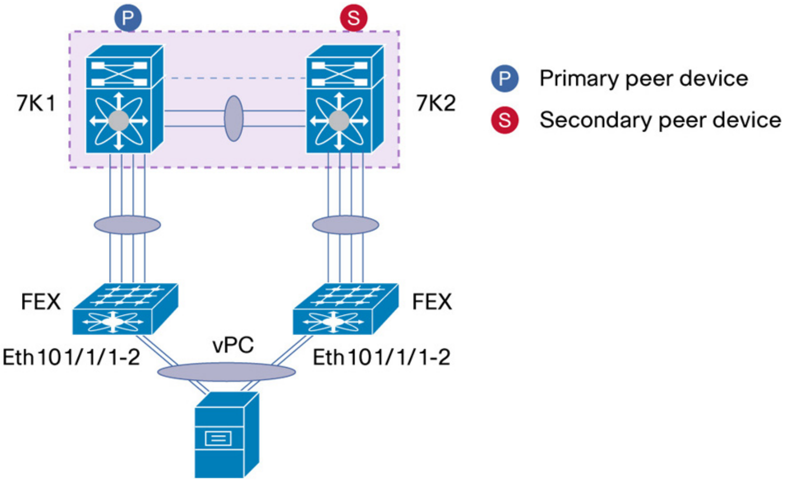
Models of FEX supported as of NX-OS 6.0 are:

- 2232PP (32 X 1/10-Gbps fiber Host Interface and 8 X 10-Gbps FEX uplinks)
- 2248 (48 X 100-Mbps/1-Gbps copper Host Interface and 4 X 10-Gbps FEX uplinks)
- 2224 (24 X 100-Mbps/1-Gbps copper Host Interface and 4 X 10-Gbps FEX uplinks)

FEX does not need any NX-OS license to operate.

NX-OS 5.2 code release is the first to support vPC functionality with FEX. Feature is also called **host vPC**: server dual-attached to 2 different FEX through a port-channel. Each FEX is connected to 1 parent switch and the 2 parent switches forming a vPC domain. Host vPC is illustrated in figure 83.

Figure 82. FEX with vPC - Host vPC



All models of FEX (2224, 2248 and 2232) support 8 port-channel member ports max.

Total number of member ports in a host vPC configuration can go up to 16.

In a host vPC configuration, there is no requirement that both FEX models match on both sides of the vPC.

FEX 2224 can be used in the left side and FEX 2248 can be used on the right side for instance.

Note: FEX 2232PP supports fiber connectivity on HIF (Host InterFace port) so mixing a 2232PP on one side and 2248 on the other side is not technically possible as 2248 supports copper connectivity (cannot mix fiber and copper ports in the same port-channel bundle).

For ease of operations, recommendation is to use the same FEX models on both sides of the VPC.

Host facing ports on FEX can be configured as switchport mode access or switchport mode trunk (so carrying multiple VLAN). Configure these ports as spanning-tree port type edge or edge trunk to force the ports to transition to forwarding state quickly.

Note: By default, FEX host facing ports are configured with spanning-tree BPDU guard and there is no knob to disable the feature. FEX host interface is not capable of processing STP BPDU and in case it occurs, the port is set to err-disabled state.

Host vPC configuration looks like classical vPC configuration and is shown below:

```
7K1:
interface eth101/1/1-2
channel-group 10 mode active

Int port-channel10
switchport
```

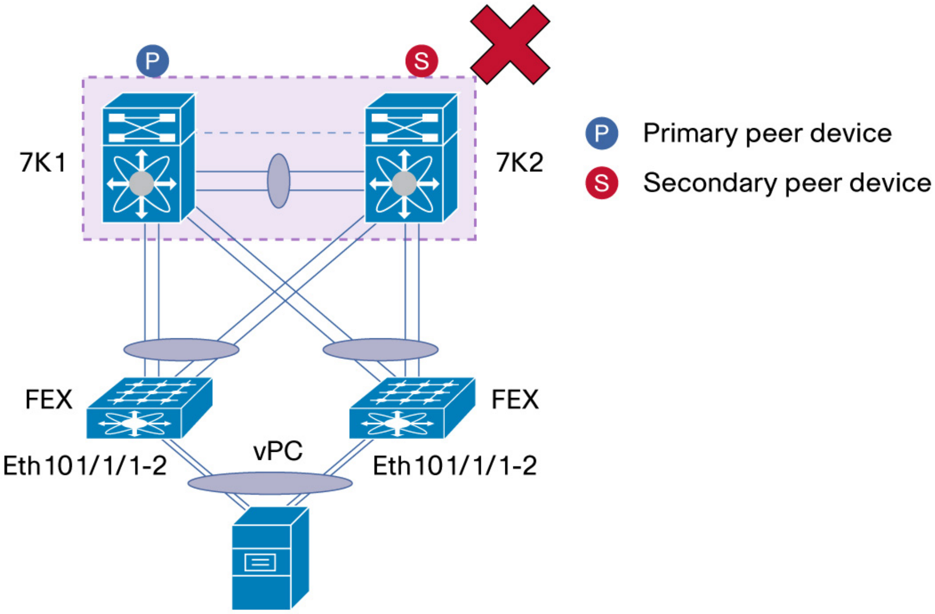
```
  switchport mode trunk
  switchport trunk allowed vlan 1-20
vpc 10
```

```
7K2:
interface eth101/1/1-2
channel-group 10 mode active

Int port-channel10
switchport
  switchport mode trunk
  switchport trunk allowed vlan 1-20
vpc 10
```

Note: The following designs are not supported with FEX and vPC as of NX-OS 6.0:

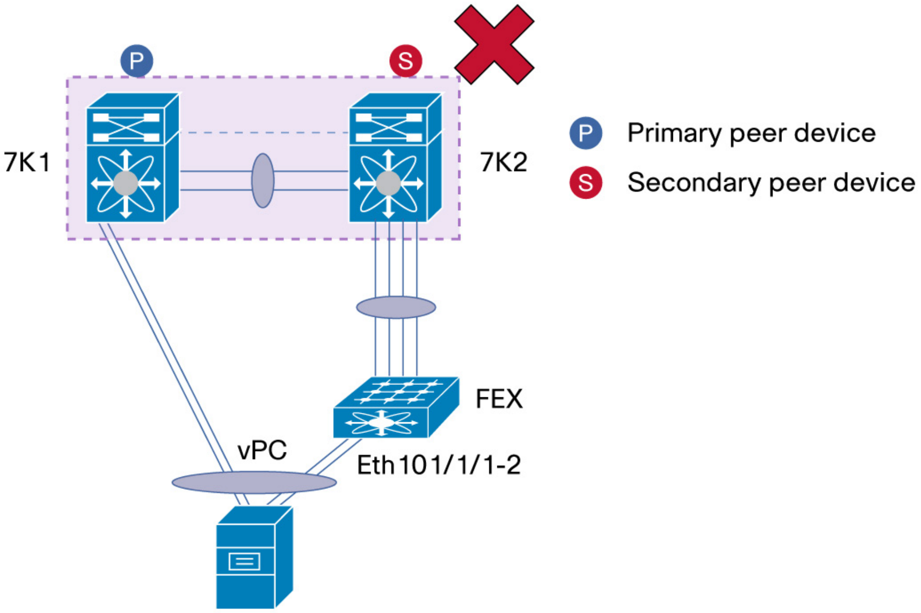
- eVPC (enhanced vPC) - Figure 84
- vPC with 1 leg on FEX and the other leg on NEXUS 7000 line card

Figure 83. FEX with vPC - eVPC (Enhanced vPC) - Not Supported



eVPC (enhanced vPC) is the configuration where FEX is vPC-attached to 2 different parent switches and server is vPC-attached to the 2 FEX (this design is also called 2-layer vPC). This is not supported on FEX with NEXUS 7000 series switches in vPC mode.

Figure 84. FEX with vPC - vPC with 1 Leg on FEX and the Other Leg On NEXUS 7000 Line Card - Not Supported



In this design, server is vPC-attached to vPC domain using this type of connectivity:
Left vPC leg connected to NEXUS 7000 integrated line card and right vPC leg connected to FEX.

This type of design is not supported.

Server must be vPC-attached to 2 different FEX devices or vPC-attached to 2 different NEXUS 7000 integrated line cards. Mixing FEX and integrated line card for vPC connectivity is not allowed.

Use the following best practices/recommendations to build properly host vPC configuration:

Strong Recommendations:

- Dual-attach the server to 2 different FEX, each FEX connected to 1 vPC peer device. Do not use other type of connections (for instance 1 vPC leg to NEXUS 7000 integrated line cards and the other vPC leg to FEX).
- Check vPC scalability number (link provided at the beginning of the document) to stay within tested and supported vPC configuration.

General Recommendations:

- Use same FEX model on both sides of the vPC (i.e 2224-2224, 2248-2248, 2232-2232).

Best Practices for VDC and vPC

Virtual Device Context (VDCs) is a virtual instance of a switch running on the Cisco Nexus 7000 Series.

As of NX-OS 6.0 release, a Cisco Nexus 7000 Series chassis supports up to 4 VDC.

VDC instances within the same chassis are independent from each other in the sense that any feature can be turned on inside a VDC without impacting operations of the other VDC.

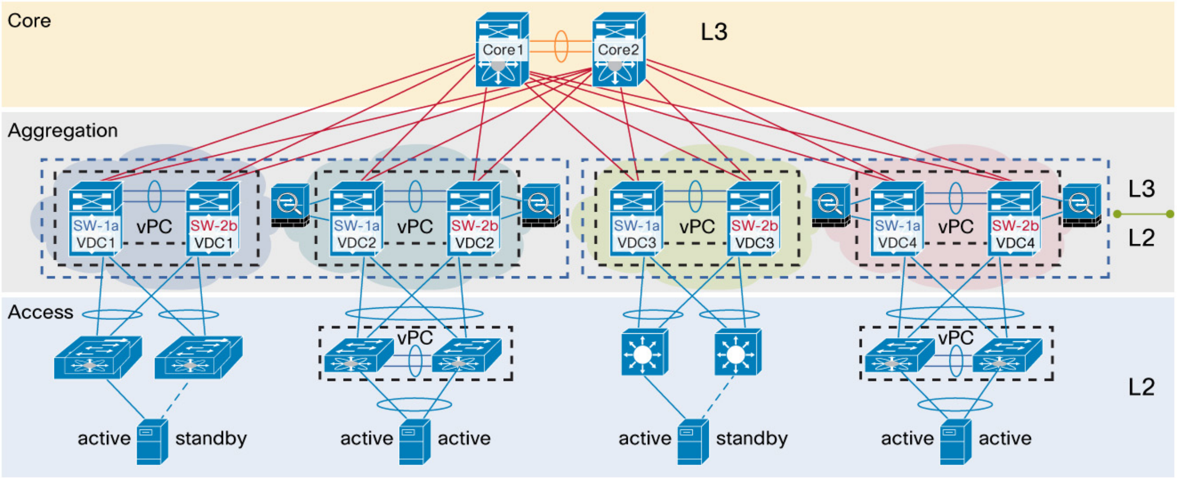
vPC technology works seamlessly with VDC. There is no particular constraint when vPC feature is enabled in a VDC. On top of vPC feature, all other L2 or L3 or security or QOS features can be enabled without creating collision with other VDC. For instance, interface vlan feature, OSPF feature, PIM feature and QOS configuration can be applied to VDC where vPC feature was enabled.

To use VDC capability on the Cisco Nexus 7000 Series chassis, the LAN_ADVANCED_SERVICES_PKG license must be installed.

A VDC can be configured with only 1 vPC domain. It is not possible to define multiple vPC domain within the same VDC (the same statements hold true for NEXUS 7000 series chassis without VDC).

Powerful topologies can be deployed with vPC using VDC. Figure 86 illustrates a classical vPC topology used at the aggregation layer providing L2/L3 demarcation point. 2 physical NEXUS 7000 chassis is enough to build 4 independent vPC domains at aggregation layer as depicted on figure 86.

Figure 85. vPC and VDC Interaction - Creating 4 Independent vPC Domains with 2 NEXUS 7000 Series Chassis



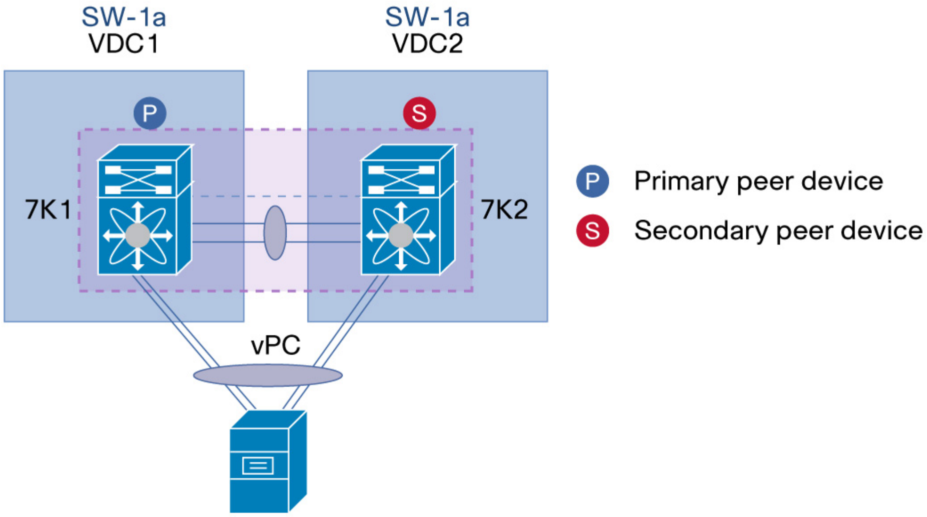
First Cisco Nexus 7000 Series chassis (SW-1a) host left vPC peer device for each vPC domain (using VDC1 for vPC domain 1, VDC2 for vPC domain 2 and so on).

Second Cisco Nexus 7000 Series chassis (SW-2b) host right vPC peer device for each vPC domain (using VDC1 for vPC domain 1, VDC2 for vPC domain 2 and so on).

VDC brings lots of flexibility in term of deployment in the context of vPC.

However, there is one particular design that needs to be cautious about: vPC domain created within the same NEXUS 7000 chassis using 2 different VDC as depicted in Figure 87.

Figure 86. vPC Domain Created Within the Same NEXUS 7000 Chassis Using 2 Different VDC



Although this design works perfectly well, it is not officially supported by Cisco. Justification is High Availability aspects of vPC are not garanteed with this type of design (if the NEXUS 7000 chassis goes down, the whole vPC domain is out of service and ISSU - In Service Software Upgrade - cannot efficiently operate in this particular design). Strong recommendation is not to use this type of configuration for production environment (it can however be used in LAB facility to test and verify vPC functional behavior).

Use the following best practices/recommendations to deploy successfully vPC with VDC.

Strong Recommendations:

- One vPC domain per VDC is supported, up to the maximum number of VDCs supported in the NEXUS 7000 series chassis.
- For each VDC that is deployed, a separate vPC peer-link and vPC peer-keepalive link infrastructure is needed.
- Running a vPC domain between VDCs on the same NEXUS 7000 series chassis is not officially supported and it is not recommended for production environments.
- 8 GB of RAM on SUPERVISOR module is recommended to support multiple VDC running vPC feature.

Note: All general recommendations for VDCs, regardless of vPC, apply for VDC. Use the default VDC for administrative management purposes only and dedicate a full line card to the VDC if possible.

Best Practices for ISSU (In-Service Software Upgrade) with vPC

This section describes best practice for performing a nondisruptive software upgrade using Cisco In-Service Software Upgrade (ISSU) when a vPC domain is configured.

vPC System NX-OS Upgrade (or Downgrade)

vPC feature is fully compatible with Cisco ISSU (In-Service Software Upgrade) or ISSD (In-Service Software Downgrade). A Cisco NX-OS code upgrade (or downgrade) on a vPC system can be performed without any packet loss.

For simplicity of reading, whatever is stated for ISSU is also true for ISSD from this point of the document.

In a vPC environment (with or without use of VDC), an ISSU is the recommended method to upgrade the system. The vPC system can be independently upgraded with no disruption to traffic. The upgrade is serialized and must be run one at the time. The configuration lock during ISSU prevents synchronous upgrades on both vPC peer devices to happen (configuration is automatically locked on other vPC peer device when ISSU is initiated).

To perform ISSU operation, 1 single knob is needed. From the default VDC, type the following command:

```
install all kickstart <bootflash_kickstart-image> system <bootflash_system-image>
```

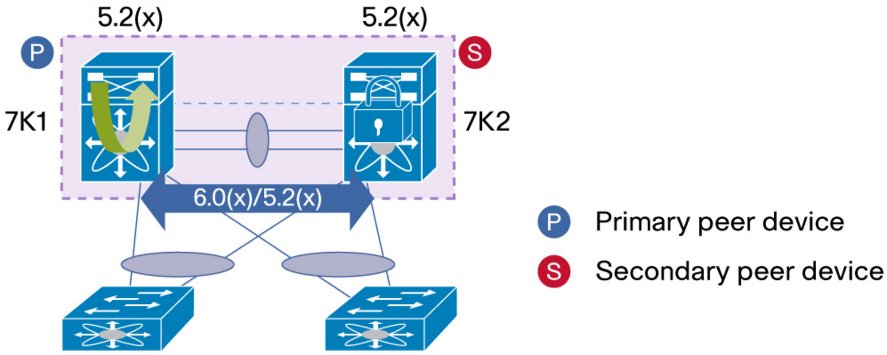
vPC presents the advantage to run seamlessly even if the 2 vPC peer devices operate with different NX-OS code.

For instance, vPC peer device 1 can run with NX-OS 5.2 release while the other vPC peer device runs with NX-OS 6.0 code. This is the key to support hitless upgrade for the vPC domain.

Note: vPC with FEX (host vPC) also fully supports ISSU. There is zero packet loss when upgrading the vPC domain containing FEX. Server dual-attached to 2 different FEX through a standard port-channel is not aware of the upgrade operation occuring in the network.

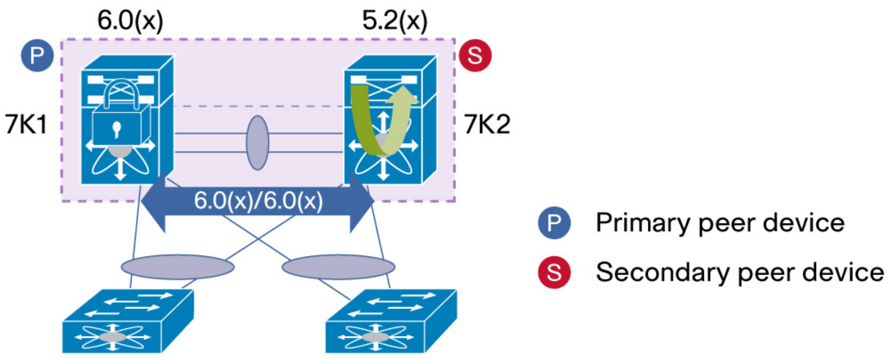
Figures 88 through 90 illustrate vPC system upgrade sequences from Cisco NX-OS Software Release 5.2 to Release 6.0.

Figure 87. ISSU with vPC - Step 1



In Step 1, as Figure 88 shows, both vPC peer devices run the Release 5.2 code. Release 6.0 code is loaded on vPC peer device 1, 7K1 (loading the code first on primary or secondary vPC peer device has no importance) using ISSU. Note that other vPC peer device (7K2) has its configuration locked to protect against any operation on the switch.

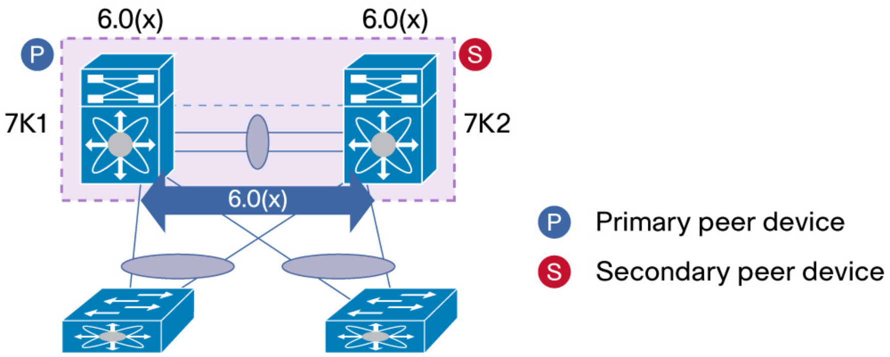
Figure 88. ISSU with vPC - Step 2



In Step 2, vPC peer device 1 (7K1) is now loaded with NX-OS Release 6.0. vPC peer device 2 (7K2) now loads the Release 6.0 code using ISSU. 7K1 configuration is then locked to prevent any change during the upgrade.

During this transition phase, vPC system works fine even if both peer devices have different NX-OS code.

Figure 89. ISSU with vPC - Step 3



In Step 3, both vPC peer devices run the Release 6.0 code. ISSU process is now completed.

Strong Recommendations:

- Use ISSU (In-Service Software Upgrade) or ISSD (In-Service Software Downgrade) to change NX-OS code release for vPC domain. Perform the operation sequentially, one vPC peer device at a time.
- Refer to NX-OS release notes to select correctly the target NX-OS code release based on running code (ISSU compatibility matrix)

Note: All general recommendations for ISSU and ISSD, regardless of vPC, still apply here.

Check carefully NX-OS release notes and follow the recommended guidelines to perform a successful ISSU or ISSD operation.

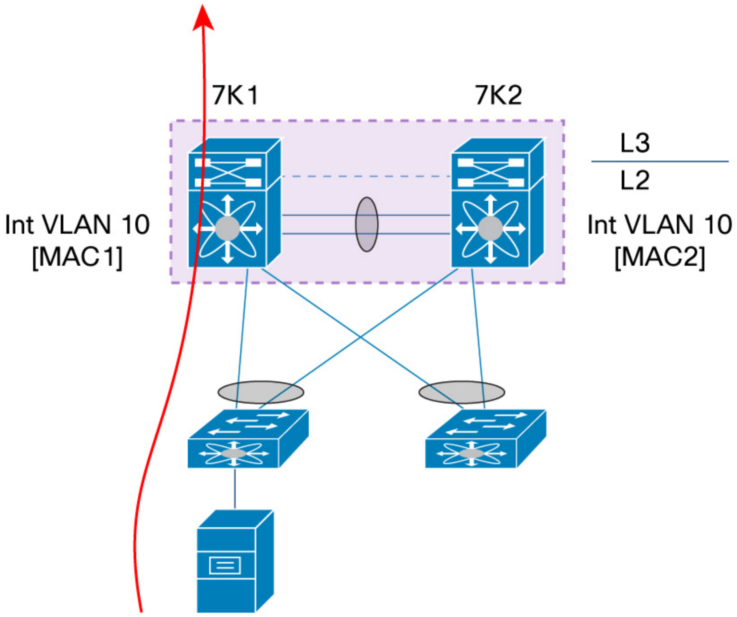
vPC Enhancements

This section describes some enhancements brought to vPC and the recommended way to use them.

vPC Peer-Gateway

The vPC Peer-Gateway enhancement (Figure 91) allows vPC interoperability with some network-attached storage (NAS) or load-balancer devices that do not perform a typical default gateway ARP request at boot up. vPC Peer-Gateway allows a vPC peer device to act as the active gateway for packets addressed to the other peer device router MAC. It keeps the forwarding of traffic local to the vPC peer device and avoids use of the peer-link (by not bridging the traffic to the other vPC peer device). There is no impact on traffic and existing functionality when activating the Peer-Gateway capability.

Figure 90. vPC Peer-Gateway



As depicted in figure 91, a NAS device (not running standard ARP request for IP default gateway) is connected to an access switch which is vPC-attached to vPC peer devices 7K1 and 7K2.

7K1 and 7K2 are configured with interface VLAN (i.e SVI) for the VLAN where NAS device is connected to (let's say VLAN 10). As NAS device don't perform standard ARP request to retrieve MAC address of the default gateway, it uses an another method to learn this MAC address. This can be done by listening to the network traffic and selecting the first received source MAC address as default gateway MAC address.

Let's assume NAS device receives its first packet from vPC peer device 7K2. In this case, it will use MAC address of interface VLAN 10 on 7K2 as default gateway MAC address. All routed traffic sent by NAS device need then to reach 7K2 in order to be routed correctly (L3 traffic destined out of vPC domain or inter-VLAN traffic). For inter-VLAN traffic, there is a risk to hit the vPC loop avoidance issue: NAS device send routed traffic, access switch hashes the traffic in direction to 7K1; 7K1 has to bridge the traffic over vPC peer-link because 7K2 MAC address (more exactly MAC address of interface VLAN 10) is the L2 destination of this traffic. Now, if traffic needs to exit out a vPC member port, it will be dropped in hardware because of vPC loop avoidance rule.

By enabling vPC Peer-Gateway functionality, each vPC peer device will replicate locally MAC address of interface VLAN defined on the other vPC peer device with the G flag (Gateway flag). In the above figure, 7K1 will program MAC2 (MAC address of interface VLAN 10) in its MAC table and set G flag for this MAC address. 7K2 will do the same for MAC1.

To activate vPC peer-gateway capability, use the following command line (under vPC configuration context mode):

```
    N7k(config-vpc-domain)# peer-gateway
```

Both vPC peer devices need to be configured with this command.

Strong Recommendation:

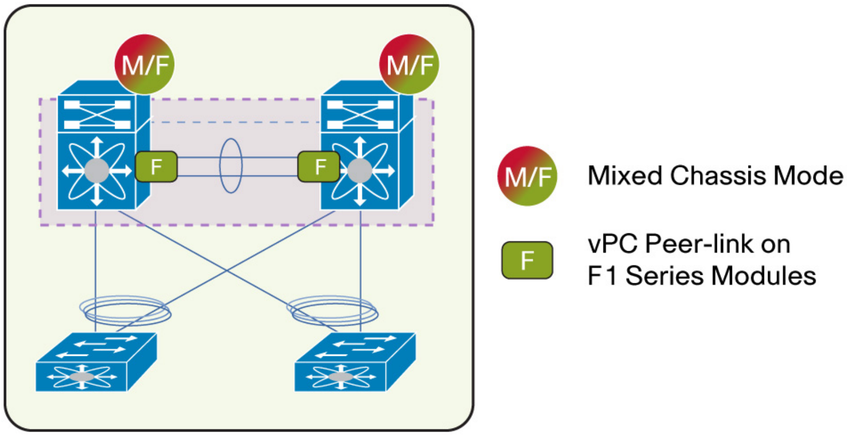
Always enable vPC peer-gateway in the vPC domain (i.e configure peer-gateway on both vPC peer devices), even if there is no end device using this feature (devices that don't perform standard ARP request for their default IP gateway). There is no side-effects enabling it.

vPC Peer-Gateway Exclude-Vlan

vPC peer-gateway exclude-vlan feature has been introduced since NX-OS 5.1.3 to address specific topology with vPC peer-link on F1 ports in mixed chassis mode (M1/F1).

The feature is relevant only for this type of topology (Figure 92). If vPC domain is a full F1 topology (not mixed chassis mode M1/F1) or a full M1 topology or a full F2 topology, this feature does not apply. Even for mixed chassis mode with vPC peer-link on M1 ports, the feature does not apply.

Figure 91. vPC Peer-Gateway Exclude-VLAN - Unique Topology Where the Feature Apply



vPC peer-gateway exclude-vlan feature was developed to avoid transit traffic between vPC peer devices using vPC peer-link to be punted to CPU, allowing direct HW switching.

Typical application for vPC peer-gateway exclude-vlan is with L3 backup routed path when a dedicated VLAN over vPC peer-link (also called transit VLAN) is used for this purpose. If all L3 uplinks on the vPC peer device fail down, backup routed path is used to carry traffic to the other vPC peer device which still has connectivity to the L3 core.

In this case, transit VLAN carrying the traffic from one vPC peer device to the other vPC peer device must be declared using the command below to work properly (otherwise transit traffic will be limited to some Mbps):

```
N7k(config-vpc-domain)# peer-gateway exclude-vlan <vlan list>
```

Required Recommendation:

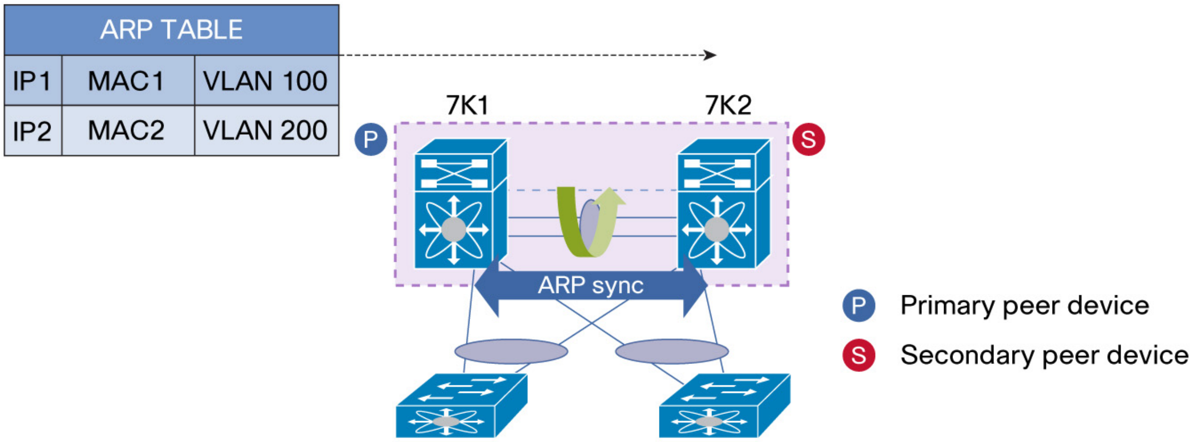
Always use vPC peer-gateway exclude-vlan when transit VLAN (over vPC peer-link) is used in the vPC domain. This is applicable only for mixed chassi mode (M1/F1) with vPC peer-link on F1 ports.

vPC ARP Sync

vPC ARP Sync improves convergence time for Layer 3 flows (North to South traffic).

When vPC peer-link fails and then recovers, vPC ARP Sync performs an ARP bulk synchronization over Cisco Fabric Services (CFS) from vPC primary peer device to vPC secondary peer device.

Figure 92. vPC ARP Sync - vPC Peer-Link Fails and Recovers



ARP TABLE		
IP1	MAC1	VLAN 100
IP2	MAC2	VLAN 200

vPC ARP Sync needs to be enabled on both vPC peer devices using the command:

```
N7k(config-vpc-domain)# ip arp synchronize
```

Strong Recommendation:

Always enable vPC ARP Sync on both vPC peer devices.

vPC Delay Restore

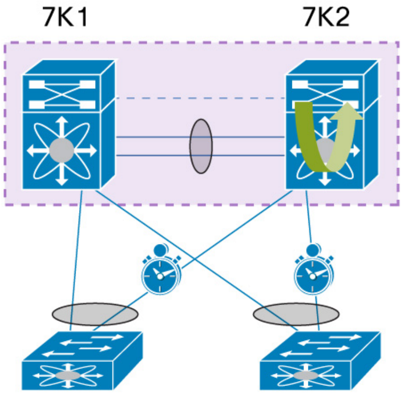
After a vPC peer device reloads and comes back up, the routing protocol needs time to reconverge.

The recovering vPCs leg may black-hole routed traffic from access to core until Layer 3 connectivity is reestablished.

vPC Delay Restore feature delays vPCs leg bringup on the recovering vPC peer device. vPC Delay Restore allows for Layer 3 routing protocols to converge before allowing any traffic on vPC leg. Result is a more graceful restoration and zero packet loss during the recovery phase (traffic still get diverted on the alive vPC peer device).

This feature is enabled by default with a vPC restoration default timer of 30 seconds. The timer can be tuned according to a specific Layer 3 convergence baseline from 1 to 3600 seconds.

Figure 93. vPC Delay Restore - 7K2 Fails and Then Recovers



Command to configure vPC Delay Restore is:

```
N7K(config-vpc-domain)# delay restore <1-3600 sec>
```

Both vPC peer devices need to be configured with this command.

A similar command - **delay restore interface-vlan <1-3600 sec >** - can be used to delay the SVI bring-up timing upon vPC peer device failure and recovery.

Strong Recommendation:
Always enable vPC delay restore (on both vPC peer devices) and the tune the timer accordingly based on network profile.

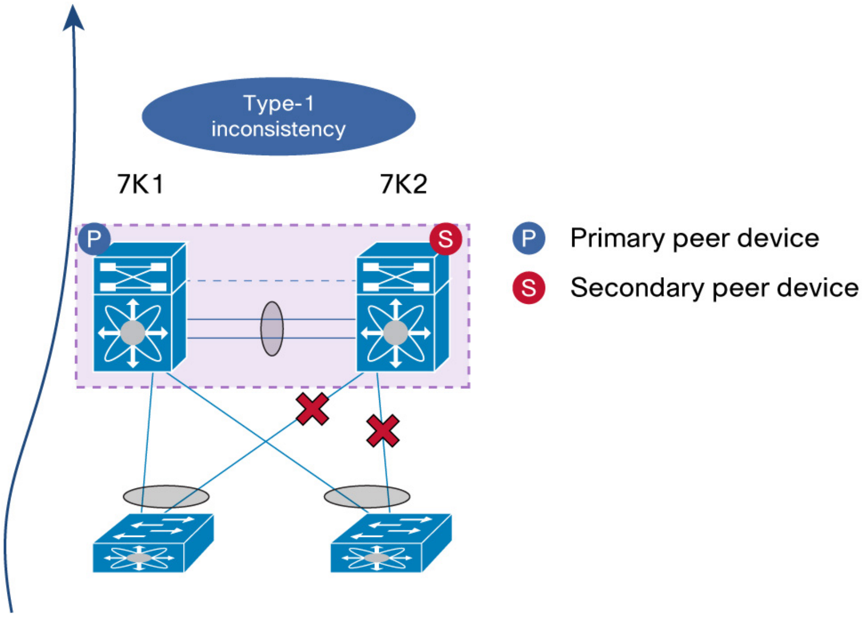
vPC Graceful Type-1 Checks

vPC member ports on both vPC peer devices should have identical parameters (MTU, speed, …).

Any inconsistency in such parameters is Type1. As a consequence, all vlans on both vpc member ports are brought down in such Inconsistency.

With vPC graceful type-1 check capability, only member ports on secondary vPC peer device are brought down. vPC member ports on primary vPC peer device remain up and process all traffic coming from (or going out to) the access device.

Figure 94. vPC Graceful Type-1 Checks



vPC graceful type-1 check is enabled by default. The associated command is:

```
N7K(config-vpc-domain)# graceful consistency-check
```

Both vPC peer devices need to be configured with this command.

Strong Recommendation:
Always enable vPC graceful type-1 check on both vPC peer devices.

vPC Auto-Recovery

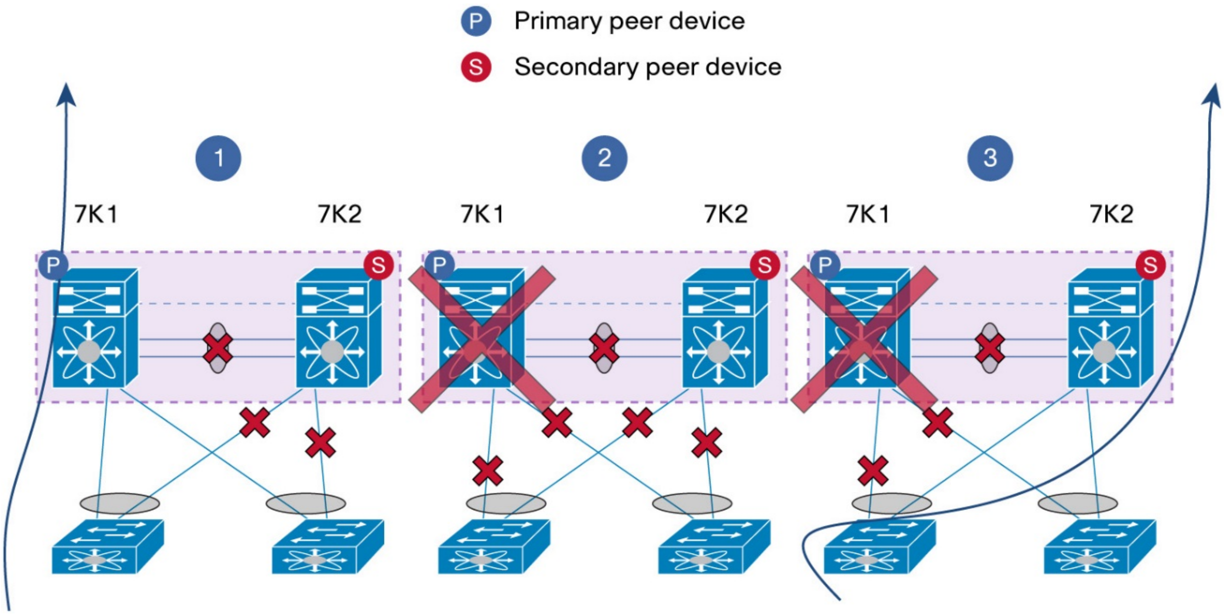
vPC auto-recovery feature was designed to address 2 enhancements to vPC.

The first one is to provide a backup mechanism in case of vPC peer-link failure followed by vPC primary peer device failure (vPC auto-recovery feature).

The second is to handle a specific case where both vPC peer devices reload but only one comes back to life (vPC auto-recovery reload-delay feature).

Let’s see the first enhancement. Figure 96 illustrates the different failure phases and resulting behavior with vPC auto-recovery.

Figure 95. vPC Auto-Recovery - vPC Peer-Link Failure Then vPC Primary Peer Device Failure



1. vPC peer-link goes down: vPC secondary peer device (7K2) shuts all its vPC member ports.
2. Primary vPC peer-device (7K1) then goes down. 7K2 receive no more any messages on vPC peer-keepalive link.
3. After 3 consecutive keepalive timeouts, vPC secondary peer device (7K2) changes role to operational primary peer device and then brings backs its vPC member ports to UP state.

vPC auto-recovery for the purpose of previous case is not enabled by default. Activate vPC auto-recovery by using this command:

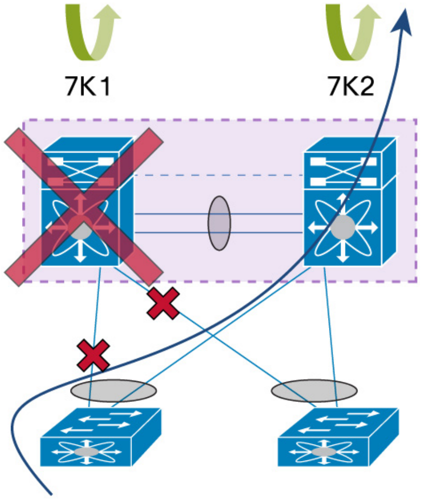
```
N7K(config-vpc-domain)# auto-recovery
```

Both vPC peer devices need to be configured with this command.

Strong Recommendation:
Always enable vPC auto-recovery on both vPC peer devices

Second enhancement is displayed in figure 97. Both vPC peer devices reload but only 7K2 comes back to life.

Figure 96. vPC Auto-Recovery Reload-Delay - Both vPC Peer Devices Reload But Only 7K2 Comes Back to Life



If both vPC peer devices reload, by default all vPC member ports are suspended until peer adjacency is reestablished between vPC devices. If only one vPC peer device becomes operational, its local vPC ports will remain suspended.

The vPC auto-recovery reload-delay feature allows the unique alive vPC peer device to assume the vPC primary role and bring up all local vPCs ports after the expiration of the delay timer. The delay can be tuned from 240 seconds to 3600 seconds.

vPC auto-recovery reload-delay is not enabled by default. Activate it by using this command:

```
N7K(config-vpc-domain)# auto-recovery reload-delay <240-3600 seconds>
```

Both vPC peer devices need to be configured with this command.

Strong Recommendation:
Always enable vPC auto-recovery reload-delay on both vPC peer devices.

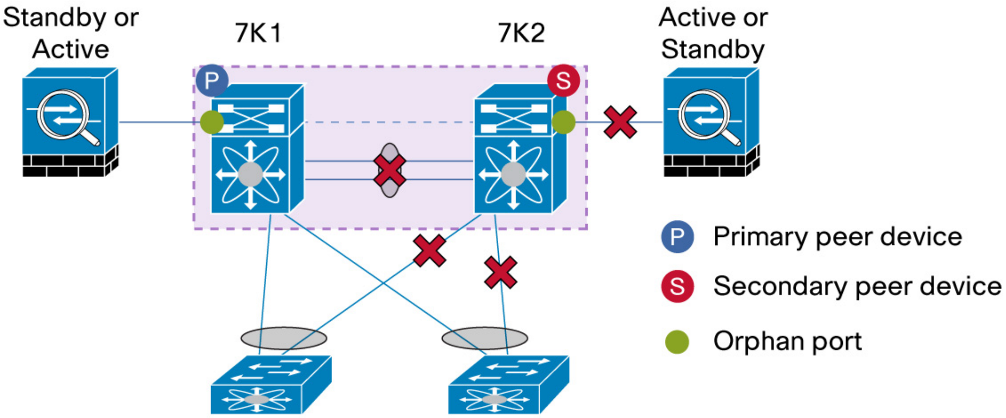
Note: vPC auto-recovery reload-delay deprecates previous feature called vPC reload restore.

vPC Orphan Ports Suspend

vPC orphan ports suspend feature was developed for single-attached devices to vPC domain and optionally working in active/standby mode (firewall or load-balancer for instance).

When a vPC peer-link goes down, the vPC secondary peer device shuts all of its vPC member ports, but it does not shut down vPC orphan ports. With vPC **orphan-ports suspend** configured, an orphan port is also shut down along with the vPC member ports when the peer-link goes down (figure 98). When the vPC peer-link is restored, configured vPC orphan ports on the secondary vPC peer device are brought up along with vPC member ports.

Figure 97. vPC Orphan Ports Suspend



vPC orphan port that must be suspended when vPC peer-link fails must be explicitly configured using the command:

```
N7K (config)# int eth 1/1
N7K (config-if)# vpc orphan-ports suspend
```

vPC orphan-port suspend CLI is available only on physical ports, not on port-channels. To configure orphan ports suspend for the port-channel, apply the above configuration for all member ports of the port-channel.

Strong Recommendation:
Use vPC orphan port suspend when single-attached devices connected to vPC domain need to be disconnected from network when vPC peer-link fails.

vPC Failure Scenarios

- **vPC member port fails**

  When one vPC member port fails, the host MAC detects a link failure on one of the port-channel members, it redistributes the affected flows to the remaining port channel members. Before the failure, MAC pointed to primary port and after the failure, it points to secondary port. 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 required for link failure scenarios.

- **vPC 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 a scenario when 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 best practice, we recommend that you configure port channels that have at least two physical 10 Gigabit-Ethernet ports as the vPC peer link.

- **vPC Primary switch Failure**

  In a vPC topology, if a failure occurs on a primary switch, then the secondary switch becomes the operational primary switch.

  If the Primary switch comes back again it will take the role of vPC operational secondary.

- **vPC keepalive link failure followed by a peer link failure**

  If the vPC keepalive link fails first and then a peer link fails, vPC primary switch continues to be primary but the vPC secondary switch becomes the operational primary switch and keeps its vPC member ports up (this is also known as dual active scenario). This can occur when both the vPC switches are healthy but, the failure is occurred because of a connectivity issue between the switches. This situation is known as a split-brain scenario. There is no loss of traffic for existing flows but new flows can be effected as the peer link is not available, the two vPC switches cannot synchronize the unicast MAC address and the IGMP groups and therefore they cannot maintain the complete unicast and multicast forwarding table and there may be some duplicate packet forwarding.

- **vPC peer keepalive link failure**

  During a vPC peer keepalive link failure there is no impact on traffic flow. We recommend that you restore the peer keepalive link at the earliest to avoid a dual active scenario.

- **Primary and secondary vPC roles in sticky bit**

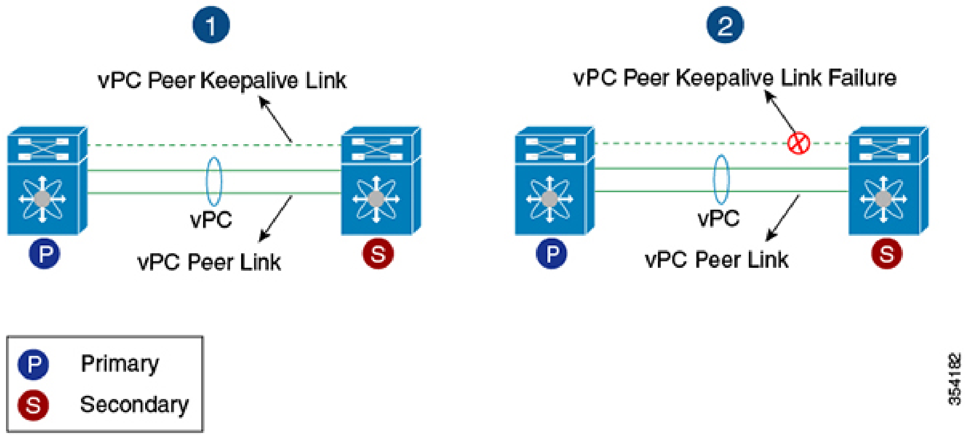
  When the two vPC systems are joined to form a vPC domain, the priority decides which device is the vPC primary and which is the vPC secondary. Lower priority means it is preferred compared to a higher priority. Also, these roles are non-preemptive, meaning a device can be operational primary but may be configured as secondary switch. To understand the operational role of a vPC member, you need to consider the status of the peer-keepalive link and the peer link.

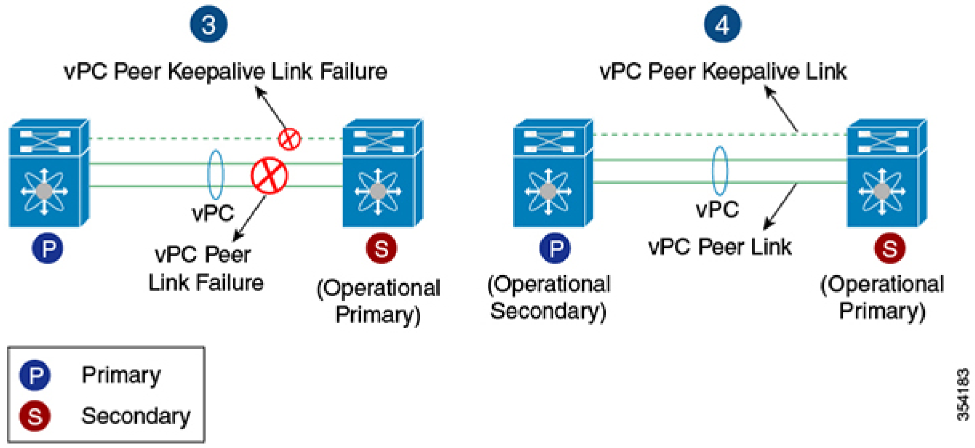
  If a primary device is reloaded and when the system comes back online and connects to the vPC secondary device (now the operational primary), the operational role of the secondary device (operational primary) will not change, to avoid unnecessary disruptions. This behavior is achieved with the sticky-bit method, whereby the sticky information is not saved in the startup configuration, thus making the device that is up and running win over the reloaded device. Hence, the vPC primary becomes the vPC operational secondary.

  In case of the peer-keepalive link fails followed by the peer-link, it would result to dual-active or split brain scenario which will make both the switches to act as vPC primary. In this case the vPC primary switch will remain as the primary and the vPC secondary switch will become operational primary

  On restoring the peer-link, the VPC operational primary switch will remain as the operational primary and the vPC secondary switch will become operational secondary.

Figure 99 illustrates the different failure phases and resulting behavior with vPC sticky-bit.





1. In the above scenario, Switch A is the primary switch and Switch B is the secondary primary.
2. When a vPC peer keepalive link fails, there is no impact on the vPC device roles.
3. When the vPC peer link fails after the vPC peer keepalive link failure, the secondary switch assumes the role of operational primary. This leads to split brain scenario.
4. When the vPC peer link is restored, the vPC device roles are reversed. Primary switch now assumes the role of operational secondary and secondary switch assumes the role of operational primary.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**