了解虚拟端口通道 (vPC) 增强功能

目录

简介

本文档介绍在vPC域中的Cisco Nexus交换机上配置的常见虚拟端口通道(vPC)增强功能。

先决条件

要求

思科建议您了解有关虚拟端口通道 (vPC) 使用案例、配置和实施的基本信息。有关此功能的详细信息，请参阅以下适用文档之一：

- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.3(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.2(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.1(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 9.3(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 9.2(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 7.x 版
- Cisco Nexus 7000 系列 NX-OS 接口配置指南 8.x 版
- Cisco Nexus 7000 系列 NX-OS 接口配置指南 7.x 版
- 设计和配置指南：Cisco Nexus 7000系列交换机虚拟端口通道(vPC)的最佳实践

使用的组件

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

背景信息

从 Cisco Nexus 数据中心交换机使用 Cisco NX-OS 以来，虚拟端口通道 (vPC) 功能获得了全方位增强，可在出现故障时提高 vPC 连接设备的可靠性，并且优化两台 vPC 对等交换机的转发行为。通过了解每个增强功能的用途、所引入的行为变化以及能解决的故障情形，您可以更清晰地理解应该在 vPC 域中配置增强功能的原因和时机，从而帮助您充分满足业务需求和要求。

适用硬件

本文档中介绍的操作过程适用于所有支持 vPC 的 Cisco Nexus 数据中心交换机。

vPC 对等交换机

本部分介绍 vPC 对等网关增强功能；可通过 peer-switch vPC 域配置命令启用此功能。

概述

在很多环境中，vPC 域中的一对 Nexus 交换机是汇聚交换机或核心交换机，充当第 2 层交换以太网域和第 3 层路由域之间的边界。这两台交换机都配置了多个 VLAN，负责路由 VLAN 间东西向流量和南北向流量。在这些环境中，从生成树协议的角度来看，Nexus 交换机通常还充当根网桥。

通常，通过将一个vPC对等体的生成树优先级设置为低值（例如0），将其配置为生成树的根网桥。另一个vPC对等配置有稍高的生成树优先级（例如4096），如果充当根网桥的vPC对等设备发生故障，它可以接管生成树中的根网桥角色。通过此配置，充当根网桥的 vPC 对等体会生成“生成树网桥协议数据单元 (BPDU)”，包含带有其系统 MAC 地址的网桥 ID。

但是，如果充当根网桥的vPC对等体发生故障并导致另一个vPC对等体接管生成树根网桥，则另一个vPC对等体会使用包含其系统MAC地址的网桥ID创建生成树BPDU，该网桥ID不同于原始根网桥的系统MAC地址。根据下游网桥的连接方式，此更改的影响会有所不同，将在以下小节中介绍。

冗余连接的非 vPC 网桥

通过冗余链路连接到两个vPC对等设备的非vPC连接网桥（因此，从生成树协议角度而言，一个链路处于阻塞状态），可检测BPDU中的更改（因此，根网桥中的更改），从而观察根端口中的更改。其他指定转发接口立即转换到阻塞状态，然后通过生成树协议有限状态机（阻塞、学习和转发），暂停时间介于所配置的生成树协议转发延迟计时器（默认情况下为15秒）之间。

根端口的变化以及随之发生的生成树协议有限状态机的遍历可能会导致网络中发生大量中断。引入 vPC 对等交换机增强功能的主要目的是，防止在其中一个 vPC 对等体离线时以上问题导致网络中断。通过vPC对等交换机增强功能，未连接vPC的网桥仍有一个处于阻塞状态的冗余链路，但如果现有根端口因链路故障而关闭，该接口会立即转换到转发状态。当脱机vPC对等重新联机时，也会发生同样的情况 — 具有最低根网桥开销的接口会抢占“根端口”角色，并且冗余链路会立即转换到“阻塞”状态。观察到的唯一数据平面影响是vPC对等设备离线时，在途的数据包不可避免地丢失。

vPC 连接的网桥

生成树域中与vPC连接的网桥会检测BPDU中的更改（以及根网桥中的更改），并刷新从本地MAC地址表中动态获取的MAC地址。在不依赖生成树协议实现无环拓扑的与vPC连接的设备拓扑中，此行为效率低下，且没有必要。从生成树协议的角度来看，vPC就像普通端口通道一样，被视为单个逻辑接口，因此vPC对等体的丢失类似于端口通道成员内单个链路的丢失。在任一场景中，生成树都不会变化，因此不必刷新生成树域中的网桥上动态获取的 MAC 地址（刷新的目的是借助以太网的“洪泛和自适应”行为重新获取生成树的新转发接口的 MAC 地址）。

此外，刷新动态获取的 MAC 地址可能会造成中断。请思考如下场景：两台主机传输大量基于 UDP 的单向流（例如 TFTP 客户端将数据发送到 TFTP 服务器）。在此流中，数据主要从 TFTP 客户端流向 TFTP 服务器，而 TFTP 服务器很少向 TFTP 客户端发回数据包。因此，在生成树域中动态获取的MAC地址刷新后，TFTP服务器的MAC在一段时间内无法获取。这意味着，发送到TFTP服务器的TFTP客户端数据在整个VLAN中泛洪，因为流量是未知单播流量。这可能会导致大量数据流传输到网络内预期之外的地方；如果流经网络的超订用部分，则可能会导致性能问题。

这种行为效率低下且不必要，思科为此引入了 vPC 对等交换机增强功能，以防在重新加载或关闭充当一个或多个 VLAN 的生成树根网桥的 vPC 对等体时发生类似情况。

要启用vPC对等交换机增强功能，两个vPC对等必须拥有相同的生成树协议配置（包括所有vPC VLAN的生成树优先级值），并且必须是所有vPC VLAN的根网桥。满足这些必备条件后，必须使用 peer-switch vPC 域配置命令启用 vPC 对等交换机增强功能。

注意:vPC对等交换机增强功能仅在包含所有VLAN根的vPC域上受支持。

启用vPC对等交换机增强功能后，两个vPC对等设备开始生成相同的生成树BPDU，其中网桥ID包含两个vPC对等设备共享的vPC系统MAC地址。如果重新加载vPC对等体，则其余vPC对等体发起的生成树BPDU不会更改，因此生成树域中的其他网桥不会看到根网桥中的任何更改，也不会对网络中的更改做出次优反应。

注意事项

在生产环境中配置 vPC 对等交换机增强功能之前，您应该了解一些注意事项。

在 vPC 对等体之间生成树优先级值必须匹配

在启用 vPC 对等交换机增强功能之前，必须修改所有 vPC VLAN 的生成树优先级配置，使两个 vPC 对等体的配置完全相同。

考虑此处的配置，其中N9K-1配置为优先级为0的VLAN 1、10和20的生成树根网桥。N9K-2是VLAN 1、10和20的辅助生成树根网桥，优先级为4096。

```
<#root>

N9K-1#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 0
interface port-channel1
  spanning-tree port type network

N9K-2#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 4096
interface port-channel1
  spanning-tree port type network
```

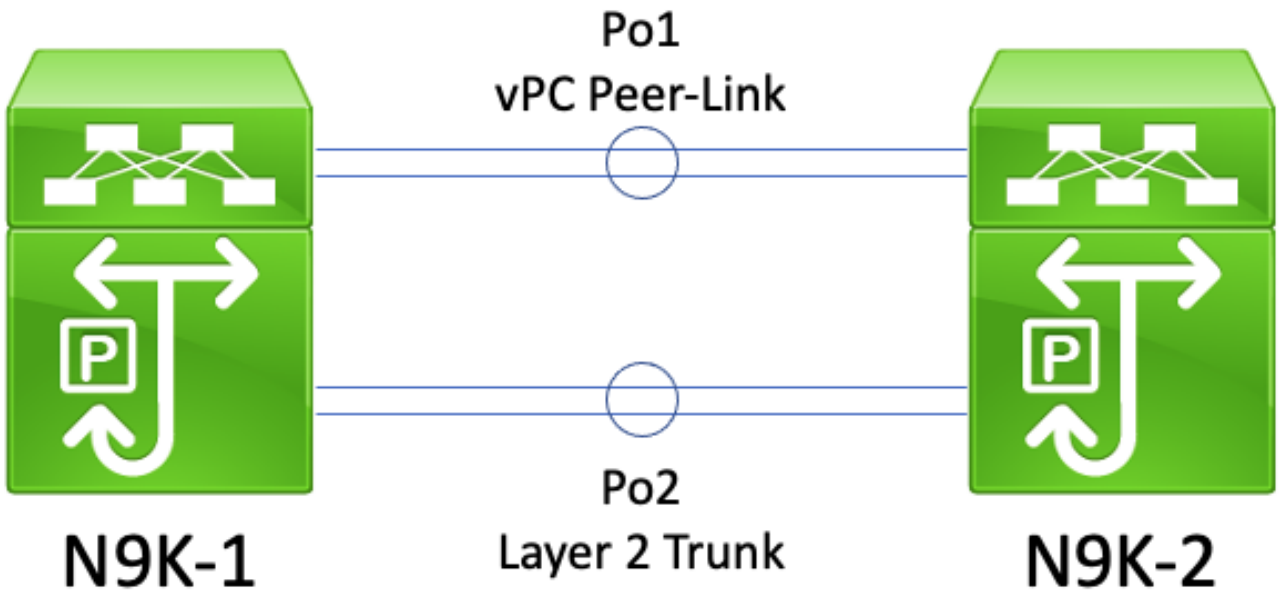
在启用vPC对等交换机增强功能之前，您必须修改N9K-2上VLAN 1、10和20的生成树优先级配置，以匹配N9K-1上相同VLAN的生成树优先级配置。此处显示修改示例。

```
<#root>

N9K-2#

configure terminal
```

```
Enter configuration commands, one per line. End with CNTL/Z.
N9K-2(config)#

spanning-tree vlan 1,10,20 priority 0

N9K-2(config)#

end

N9K-2#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 0
interface port-channel1
  spanning-tree port type network

N9K-1#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 0
interface port-channel1
  spanning-tree port type network
```

vPC 对等交换机对非 vPC VLAN 的影响

请思考以下拓扑：



在此拓扑中，两个vPC对等体（N9K-1和N9K-2）之间有两个第2层中继 — Po1和Po2。Po1是传输vPC VLAN的vPC对等链路，而Po2是传输所有非vPC VLAN的第2层中继。如果N9K-1和N9K-2上通过Po2传输的非vPC VLAN的生成树优先级值相同，则每个vPC对等体会从两台交换机上相同的vPC系统MAC地址生成生成树BPDU帧。因此，尽管N9K-2是生成树BPDU的源交换机，但N9K-1似乎在Po2上为每个非vPC VLAN接收自己的生成树BPDU。从生成树的角度来看，N9K-1将Po2置于所有非vPC VLAN的阻塞状态。

这是预料之中的现象。为了防止发生此行为或解决此问题，两个 vPC 对等体必须为所有非 vPC VLAN 配置不同的生成树优先级值。这允许一个vPC对等体成为非vPC VLAN的根网桥，并将vPC对等体之间的第2层中继转换到指定转发状态。同样，远程vPC对等体将第2层中继在vPC对等体之间转换为指定根状态。这允许非vPC VLAN中的流量通过第2层中继通过两个vPC对等体流动。

配置

有关如何配置 vPC 对等交换机功能，请参阅以下示例。

在本示例中，N9K-1被配置为优先级为0的VLAN 1、10和20的生成树根网桥。N9K-2是VLAN 1、10和20的辅助生成树根网桥，优先级为4096。

<#root>

```
N9K-1#

show running-config vpc

<snip>
vpc domain 1
  role priority 150
  peer-keepalive destination 10.122.190.196

interface port-channel1
  vpc peer-link

N9K-2#

show running-config vpc

<snip>
vpc domain 1
  peer-keepalive destination 10.122.190.195

interface port-channel1
  vpc peer-link

N9K-1#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 0
interface port-channel1
  spanning-tree port type network

N9K-2#

show running-config spanning-tree

spanning-tree vlan 1,10,20 priority 4096
interface port-channel1
  spanning-tree port type network
```

首先，必须把 N9K-2 的生成树优先级配置改为与 N9K-1 完全相同。这是使 vPC 对等交换机功能按预期运行的前提条件。如果N9K-2的系统MAC地址低于N9K-1的系统MAC地址，则N9K-2会取代生成树域的根网桥角色，从而导致生成树域中的其他网桥刷新所有受影响的VLAN的本地MAC地址表

。此现象的一个示例如下所示。

<#root>

```
N9K-1#

show spanning-tree vlan 1


VLAN0001
  Spanning tree enabled protocol rstp
  Root ID    Priority    1
             Address     689e.0baa.dea7
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    1      (priority 0 sys-id-ext 1)
             Address     689e.0baa.dea7
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Desg FWD 1         128.4096 (vPC peer-link) Network P2p
Po10             Desg FWD 1         128.4105 (vPC) P2p
Po20             Desg FWD 1         128.4115 (vPC) P2p

N9K-2#

show spanning-tree vlan 1


VLAN0001
  Spanning tree enabled protocol rstp
  Root ID    Priority    1
             Address     689e.0baa.dea7
             Cost        1
             Port        4096 (port-channel1)
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    4097   (priority 4096 sys-id-ext 1)
             Address     689e.0baa.de07
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Root FWD 1         128.4096 (vPC peer-link) Network P2p
Po10             Desg FWD 1         128.4105 (vPC) P2p
Po20             Desg FWD 1         128.4115 (vPC) P2p

N9K-2#

configure terminal

Enter configuration commands, one per line. End with CNTL/Z.
N9K-2(config)#

spanning-tree vlan 1,10,20 priority 0

N9K-2(config)#

end

N9K-2#
```

```
show spanning-tree vlan 1


VLAN0001
  Spanning tree enabled protocol rstp
  Root ID    Priority    1
             Address     689e.0baa.de07
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    1      (priority 0 sys-id-ext 1)
             Address     689e.0baa.de07
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Desg FWD 1         128.4096 (vPC peer-link) Network P2p
Po10             Desg FWD 1         128.4105 (vPC) P2p
Po20             Desg FWD 1         128.4115 (vPC) P2p
```

接下来，我们可以通过 peer-switch vPC 域配置命令启用 vPC 对等交换机功能。这会更改两个vPC对等体发起的生成树BPDU中的网桥ID，从而导致生成树域中的其他网桥刷新所有受影响的VLAN的本地MAC地址表。

```
<#root>

N9K-1#

configure terminal

N9K-1(config)#

vpc domain 1

N9K-1(config-vpc-domain)#

peer-switch

N9K-1(config-vpc-domain)#

end

N9K-1#

N9K-2#

configure terminal

N9K-2(config)#

vpc domain 1

N9K-2(config-vpc-domain)#

peer-switch

N9K-2(config-vpc-domain)#

end

N9K-2#
```

您可以使用 show spanning-tree summary 命令验证两个声明自己为 vPC VLAN 根网桥的 vPC 对等体，从而验证 vPC 对等交换机功能是否按预期运行。此输出还应显示 vPC 对等交换机功能已启用且正常运行。

<#root>

N9K-1#

show spanning-tree summary

```
Switch is in rapid-pvst mode
Root bridge for: VLAN0001, VLAN0010, VLAN0020
L2 Gateway STP                          is disabled
Port Type Default                       is disable
Edge Port [PortFast] BPDU Guard Default  is disabled
Edge Port [PortFast] BPDU Filter Default is disabled
Bridge Assurance                        is enabled
Loopguard Default                       is disabled
Pathcost method used                    is short
vPC peer-switch                         is enabled (operational)
STP-Lite                                is disabled

Name                   Blocking Listening Learning Forwarding STP Active
---------------------- -------- --------- -------- ---------- ----------
VLAN0001                    0         0        0          3          3
VLAN0010                    0         0        0          3          3
VLAN0020                    0         0        0          3          3
---------------------- -------- --------- -------- ---------- ----------
3 vlans                     0         0        0          9          9
```

N9K-2#

show spanning-tree summary

```
Switch is in rapid-pvst mode
Root bridge for: VLAN0001, VLAN0010, VLAN0020
L2 Gateway STP                          is disabled
Port Type Default                       is disable
Edge Port [PortFast] BPDU Guard Default  is disabled
Edge Port [PortFast] BPDU Filter Default is disabled
Bridge Assurance                        is enabled
Loopguard Default                       is disabled
Pathcost method used                    is short
vPC peer-switch                         is enabled (operational)
STP-Lite                                is disabled

Name                   Blocking Listening Learning Forwarding STP Active
---------------------- -------- --------- -------- ---------- ----------
VLAN0001                    0         0        0          3          3
VLAN0010                    0         0        0          3          3
VLAN0020                    0         0        0          3          3
---------------------- -------- --------- -------- ---------- ----------
3 vlans                     0         0        0          9          9
```

使用 show spanning-tree vlan{x} 命令可查看特定 VLAN 的更多详细信息。具有主或运行主vPC角色的交换机的所有接口都处于指定转发状态。具有辅助或运行辅助的vPC角色的交换机的所有接口都处于指定转发状态，但vPC对等链路除外，该链路处于根转发状态。请注意，show vpc role 输出

中显示的 vPC 系统 MAC 地址与每个 vPC 对等体的根网桥 ID 和网桥 ID 完全相同。

<#root>

```
N9K-1#

show vpc role


vPC Role status
----------------------------------------------------
vPC role                          : primary
Dual Active Detection Status      : 0
vPC system-mac                    : 00:23:04:ee:be:01
vPC system-priority               : 32667
vPC local system-mac              : 68:9e:0b:aa:de:a7
vPC local role-priority           : 150
vPC local config role-priority    : 150
vPC peer system-mac               : 68:9e:0b:aa:de:07
vPC peer role-priority            : 32667
vPC peer config role-priority     : 32667

N9K-1#

show spanning-tree vlan 1


VLAN0001
  Spanning tree enabled protocol rstp
  Root ID    Priority    1
             Address     0023.04ee.be01
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    1      (priority 0 sys-id-ext 1)
             Address     0023.04ee.be01
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Desg FWD 1         128.4096 (vPC peer-link) Network P2p
Po10             Desg FWD 1         128.4105 (vPC) P2p
Po20             Desg FWD 1         128.4115 (vPC) P2p

N9K-2#

show vpc role


vPC Role status
----------------------------------------------------
vPC role                          : secondary
Dual Active Detection Status      : 0
vPC system-mac                    : 00:23:04:ee:be:01
vPC system-priority               : 32667
vPC local system-mac              : 68:9e:0b:aa:de:07
vPC local role-priority           : 32667
vPC local config role-priority    : 32667
vPC peer system-mac               : 68:9e:0b:aa:de:a7
vPC peer role-priority            : 150
vPC peer config role-priority     : 150
```

```
N9K-2#

show spanning-tree vlan 1


VLAN0001
  Spanning tree enabled protocol rstp
  Root ID    Priority    1
             Address     0023.04ee.be01
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    1      (priority 0 sys-id-ext 1)
             Address     0023.04ee.be01
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Root FWD 1         128.4096 (vPC peer-link) Network P2p
Po10             Desg FWD 1         128.4105 (vPC) P2p
Po20             Desg FWD 1         128.4115 (vPC) P2p
```

最后，我们可以在任一 vPC 对等体上使用 Ethanalyzer 控制平面数据包捕获实用程序，以确认两个 vPC 对等体都生成了生成树 BPDU，该 BPDU 的网桥 ID 和根网桥 ID 包含两个 vPC 对等体之间共用的 vPC 系统 MAC 地址。

<#root>

```
N9K-1#

ethanalyzer local interface inband display-filter stp limit-captured-frames 0

<snip>
Capturing on inband
2021-05-13 01:59:51.664206 68:9e:0b:aa:de:d4 -> 01:80:c2:00:00:00 STP RST. Root = 0/1/00:23:04:ee:be:01

N9K-2#

ethanalyzer local interface inband display-filter stp limit-captured-frames 0

<snip>
Capturing on inband
2021-05-13 01:59:51.777034 68:9e:0b:aa:de:34 -> 01:80:c2:00:00:00 STP RST. Root = 0/1/00:23:04:ee:be:01
```

影响

启用vPC对等交换机增强功能的影响因生成树域中的其他网桥是否通过vPC连接到两个vPC对等体而异，或者是否冗余地连接到两个没有vPC的vPC对等体。

冗余连接的非 vPC 网桥

如果非 vPC 连接的网桥通过冗余链路（从生成树协议角度来看，一个链路处于阻塞状态）连接到两个 vPC 对等体，在检测到生成树 BPDU 中通告的生成树根网桥发生变化时，两个冗余接口之间的

网桥根端口可能会发生变化。其他指定转发接口将立即转换为阻塞状态，然后遍历生成树协议有限状态机（“阻塞”“学习”和“转发”），暂停时间与已配置的生成树协议转发延迟计时器相同，默认为 15 秒。根端口的变化以及随之发生的生成树协议有限状态机的遍历可能会导致网络中发生大量中断。

值得一提的是，当前作为生成树域的根网桥的vPC对等设备离线时（例如发生电源故障、硬件故障或重新加载时），就会发生这种影响。并非只有 vPC 对等交换机增强功能会导致此行为；从生成树角度来看，启用 vPC 对等交换机增强功能只是导致与 vPC 对等体离线类似的行为。

vPC 连接的网桥

如果vPC连接的网桥检测到生成树BPDU中通告的生成树根网桥发生变化，网桥会从其MAC地址表中刷新动态获取的MAC地址。配置vPC对等交换机功能时，您可以在以下两种情况下观察此行为：

1. 当在两个 vPC 对等体之间配置一致的生成树优先级值时，如果之前不是根网桥的 vPC 对等体的系统 MAC 地址比之前是根网桥的 vPC 对等体的系统 MAC 地址低，则生成树根网桥可能会从一个 vPC 对等体更改为另一个 vPC 对等体。本文档的“vPC 对等交换机配置”部分显示了此场景的示例。
2. 通过peer-switch vPC域配置命令启用vPC对等交换机功能后，两个vPC对等交换机开始作为生成树域的根网桥运行。两台vPC对等设备开始生成相同的生成树BPDU，并声明自己是生成树域的根网桥。

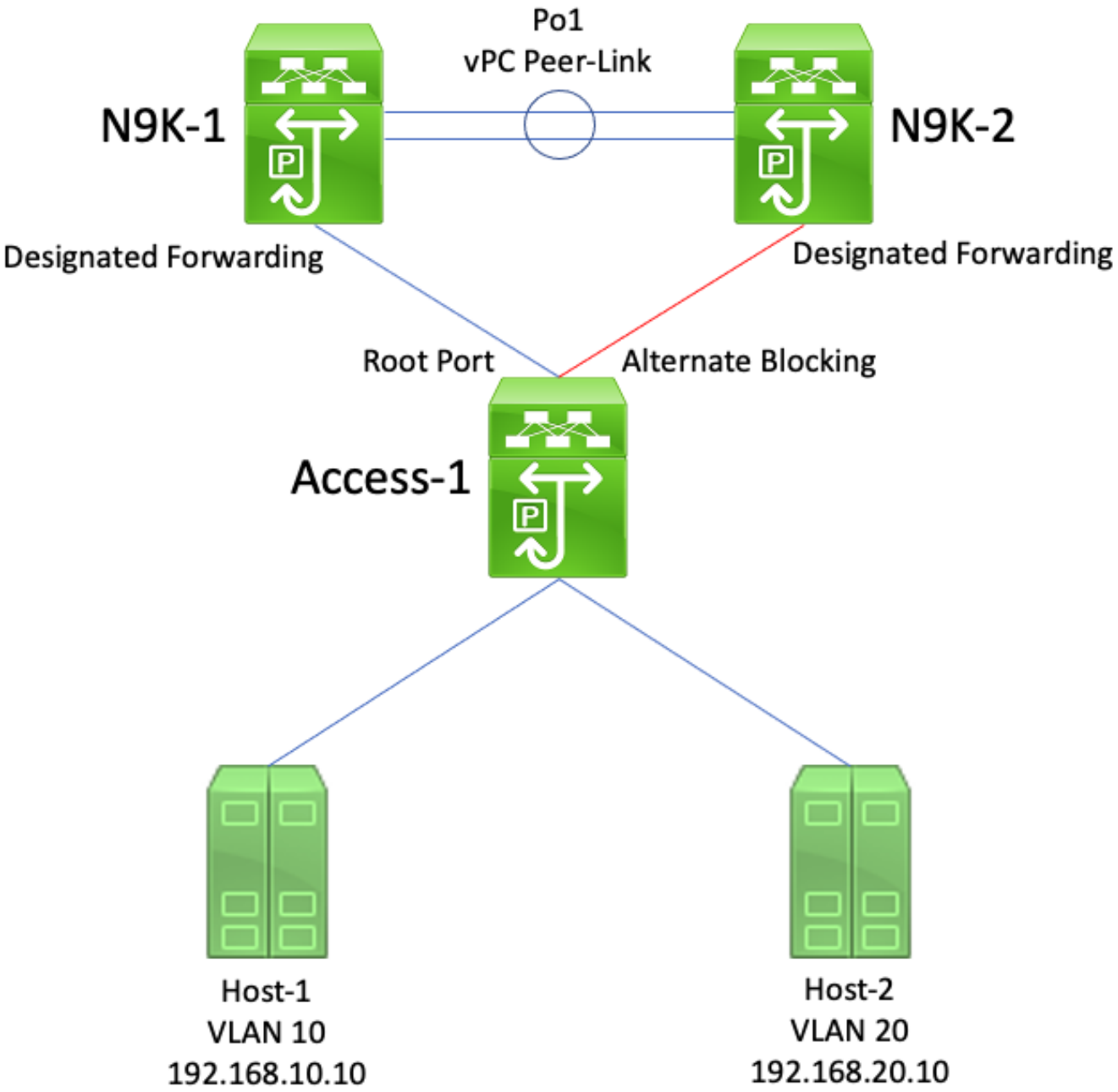
在大多数场景和拓扑中，不会因为这两个场景之一而观察到任何数据平面影响。但是，在短时间内，由于未知单播泛洪，数据平面流量在VLAN内泛洪，因为动态获取的MAC地址直接导致帧的目的MAC地址无法通过任何交换机端口获知。在某些拓扑中，如果数据平面流量泛洪到 VLAN 内超订用的网络设备，可能会导致短暂的性能问题或丢包。这也会导致带宽密集型单向流量流或静默主机（主要接收数据包但很少发送数据包的主机）出现问题，因为此类流量会长时间泛洪到VLAN中，而不是正常情况下直接交换到目标主机。

值得一提的是，此影响与从受影响的VLAN内网桥的MAC地址表中动态获取的MAC地址刷新有关。并非只有 vPC 对等交换机增强功能或根网桥的变化会导致此行为，由于 VLAN 中出现非边缘端口而生成的拓扑变化通知也会导致此行为。

故障情形示例

冗余连接的非 vPC 网桥重新启动有限状态机

请思考以下拓扑：



在此拓扑中，N9K-1 和 N9K-2 是 vPC 域中的 vPC 对等体。N9K-1 为所有 VLAN 配置了生成树优先级值为 0，使 N9K-1 成为所有 VLAN 的根网桥。N9K-2 为所有 VLAN 配置了生成树优先级值为 4096，使 N9K-2 成为所有 VLAN 的辅助根网桥。Access-1 是通过第 2 层交换机端口冗余连接到 N9K-1 和 N9K-2 的交换机。这些交换机端口未捆绑到端口通道中，因此生成树协议将连接到 N9K-1 的链路置于指定根状态，将连接到 N9K-2 的链路置于备用阻塞状态。

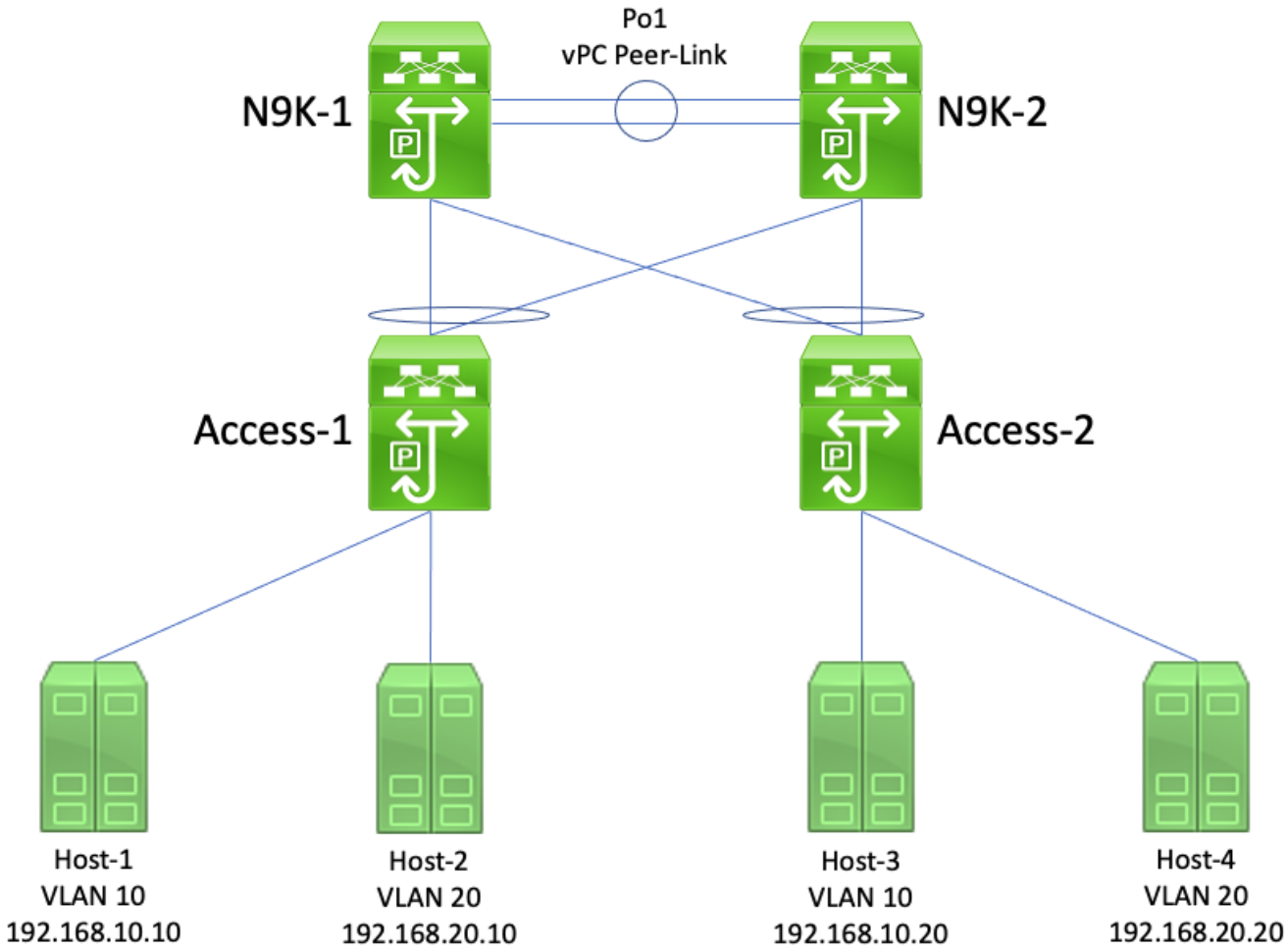
请考虑 N9K-1 因硬件故障、电源故障或交换机重新加载而离线的故障情形。N9K-2使用自己的系统MAC地址作为网桥ID通告生成树BPDU，从而声明自己是所有VLAN的根网桥。Access-1发现根网桥的ID发生更改。此外，其指定根端口转换到关闭/关闭状态，这意味着新的指定根端口是处于面向N9K-2的备用阻塞状态的链路。

指定根端口的变化导致所有非边缘生成树端口逐步通过生成树协议有限状态机（阻塞、学习和转发），暂停时间介于所配置的生成树协议转发延迟计时器（默认情况下为15秒）之间。此过程可能对网络造成极大中断。

在启用了vPC对等交换机增强功能的同一故障场景中，N9K-1和N9K-2都使用共享vPC系统MAC地址作为网桥ID传输相同的生成树BPDU。如果N9k-1发生故障，N9K-2将继续传输此相同的生成树BPDU。因此，Access-1立即将指向N9K-2的备用阻塞链路转换为指定根状态，并开始通过该链路转发流量。此外，生成树根网桥 ID 不会发生变化，这会阻止非边缘端口遍历生成树协议有限状态机，从而减少网络中断。

vPC 连接的网桥刷新动态获取的 MAC 地址

请思考以下拓扑：



在此拓扑中，N9K-1和N9K-2是vPC域中的vPC对等体，在VLAN 10和VLAN 20之间执行VLAN间路由。N9K-1为VLAN 10和VLAN 20配置了生成树优先级值0，使N9K-1成为两个VLAN的根网桥。N9K-2 配置了 VLAN 10 和 VLAN 20 的生成树优先级值 4096，使 N9K-2 成为两个 VLAN 的辅助根网桥。Host-1、Host-2、Host- 3 和 Host-4 都在相互持续通信。

请考虑 N9K-1 因硬件故障、电源故障或交换机重新加载而离线的故障情形。N9K-2使用自己的系统MAC地址作为网桥ID通告生成树BPDU，从而宣称自己是VLAN 10和VLAN 20的根网桥。Access-1和Access-2会看到根网桥的ID发生变化，尽管生成树保持不变（即面向N9K-1和N9K-2的vPC仍然是指定根端口），但Access-1和Access-2都会刷新VLAN 10和VLAN 20中所有动态获知的MAC地址的MAC地址。

在大多数环境中，刷新动态获取的 MAC 地址只会产生极小的影响。除了在 N9K-1 离线时正好传输给它而丢失的数据包，没有任何其他数据包丢失，但流量会作为未知单播流量在每个广播域中暂时

泛洪，而广播域中的所有交换机都重新获取动态 MAC 地址。

在同一故障情形中，当启用 vPC 对等交换机增强功能时，N9K-1 和 N9K-2 会传输使用共享 vPC 系统 MAC 地址作为网桥 ID 的相同的生成树 BPDU。如果N9k-1发生故障，N9K-2将继续传输此相同的生成树BPDU。因此，Access-1和Access-2不知道生成树拓扑发生了任何变化 — 从它们的角度来看，根网桥的生成树BPDU相同，因此无需刷新相关VLAN中动态获取的MAC地址。这可以防止在此故障情形中未知单播流量在每个广播域中泛洪。

VPC 对等网关

本部分介绍使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能。

概述

默认情况下，在 vPC 域中配置的 Nexus 交换机根据双主用第一跳冗余协议 (FHRP) 转发流量。这意味着，如果任一vPC对等设备收到一个数据包，该数据包的目的MAC地址属于交换机上配置的热备份路由器协议(HSRP)或虚拟路由器冗余协议(VRRP)组，则无论其HSRP或VRRP控制平面状态如何，交换机都将根据其本地路由表路由该数据包。换句话说，处于 HSRP 备用或 VRRP 备份状态的 vPC 对等体的预期行为是路由发往 HSRP 或 VRRP 虚拟 MAC 地址的数据包。

当vPC对等设备路由发往FHRP虚拟MAC地址的数据包时，它会使用新的源和目标MAC地址重写该数据包。源MAC地址是数据包路由到的VLAN中vPC对等体的交换虚拟接口(SVI)的MAC地址。根据vPC对等体的本地路由表，目的MAC地址是与数据包目的IP地址的下一跳IP地址关联的MAC地址。在VLAN间路由场景中，数据包被重写后的目的MAC地址是数据包最终发送到的主机的MAC地址。

某些主机不遵循标准转发行为以优化功能。在此种行为中，主机在应答传入数据包时不会查找路由表和/或 ARP 缓存。相反，主机会为应答数据包翻转传入数据包的源 MAC 地址和目的 MAC 地址。换言之，传入数据包的源 MAC 地址成为应答数据包的目的 MAC 地址，而传入数据包的目的 MAC 地址成为应答数据包的源 MAC 地址。此行为不同于遵循标准转发行为的主机；这些主机将查找本地路由表和/或 ARP 缓存，并将应答数据包的目的 MAC 地址设置为 FHRP 虚拟 MAC 地址。

如果主机生成的应答数据包应发往一个 vPC 对等体，但是此 vPC 把数据包出口转发给另一个 vPC，这种非标准主机行为可能违反 vPC 环路规避规则。另一台vPC对等设备收到发往其vPC对等设备的MAC地址的数据包，并将数据包从vPC对等链路转发到拥有数据包目的MAC地址字段中的MAC地址的vPC对等设备。拥有MAC地址的vPC对等设备尝试在本地路由数据包。如果数据包需要传出vPC，vPC对等体会因违反vPC环路避免规则而丢弃此数据包。因此，针对采取此非标准行为的源主机或目的主机的某些流，您可能会观察到连接问题或丢包现象。

我们引入了 vPC 对等网关增强功能，以消除主机采取此非标准行为造成的丢包。实现的方法是允许一个 vPC 对等体在本地路由发往另一个 vPC 对等体 MAC 地址的数据包，这样发往远程 vPC 对等体的数据包就无需从 vPC 对等链路上出口转发，即可路由。换言之，vPC 对等网关增强功能允许一个 vPC 对等体“代表”远程 vPC 对等体路由数据包。我们可以使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能。

注意事项

vPC 或 vPC VLAN 上单播路由协议邻接抖动

如果在两个 vPC 对等体与 vPC 连接的路由器（或通过 vPC 孤立端口连接的路由器）之间形成动态单播路由协议邻接，那么在启用了 vPC 对等网关增强功能而没有立即配置路由/第 3 层的时候，路由协议邻接可能会持续抖动。本文档在"带 vPC 对等网关的 vPC 上的单播路由协议邻接故障情形示例"和"带 vPC 对等网关的 vPC VLAN 上的单播路由协议邻接"等部分详描述了这些故障情形。

若要解决此问题，请在使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能后，立即使用 Layer3 peer-router vPC 域配置命令启用路由/第 3 层 vPC 增强功能。

自动禁用 ICMP 和 ICMPv6 重定向

当vPC对等网关增强功能启用时，所有vPC VLAN SVI（即，与通过vPC对等链路中继的VLAN关联的任何SVI）上都会自动禁用ICMP和ICMPv6重定向数据包的生成。交换机通过在所有 vPC VLAN SVI 上配置 no ip redirects 和 no ipv6 redirects 来实现此目的。这可防止交换机生成 ICMP 重定向数据包，以响应传入了交换机但目的 MAC 和 IP 地址指向交换机 vPC 对等体的的数据包。

如果特定VLAN内的环境中需要ICMP或ICMPv6重定向数据包，您需要使用peer-gateway exclude-vlan <vlan-id> vPC domain configuration命令排除此VLAN以利用vPC对等网关增强功能。

注意：Nexus 9000系列交换机不支持peer-gateway exclude-vlan <vlan-id> vPC域配置命令。

配置

有关如何配置 vPC 对等网关功能，请参阅此处。

在本例中，N9K-1 和 N9K-2 是 vPC 域中的 vPC 对等体。两台vPC对等设备都为VLAN 10配置了一个HSRP组。N9K-1是优先级为150的HSRP活动路由器，而N9K-2是默认优先级为100的HSRP备用路由器。

<#root>

```
N9K-1#

show running-config vpc

<snip>
vpc domain 1
  role priority 150
  peer-keepalive destination 10.82.140.43

interface port-channel1
  vpc peer-link

N9K-2#

show running-config vpc

<snip>
vpc domain 1
  peer-keepalive destination 10.82.140.42

interface port-channel1
  vpc peer-link
```

```
N9K-1#

show running-config interface vlan 10

<snip>
interface Vlan10
  no shutdown
  ip address 192.168.10.2/24
  hsrp 10
    preempt
    priority 150
    ip 192.168.10.1

N9K-2#

show running-config interface vlan 10

<snip>
interface Vlan10
  no shutdown
  ip address 192.168.10.3/24
  hsrp 10
    ip 192.168.10.1

N9K-1#

show hsrp interface vlan 10 brief

*:IPv6 group   #:group belongs to a bundle
                      P indicates configured to preempt.
                      |
 Interface   Grp  Prio P State    Active addr      Standby addr     Group addr
  Vlan10      10   150  P Active   local            192.168.10.3     192.168.10.1    (conf)

N9K-2#

show hsrp interface vlan 10 brief

*:IPv6 group   #:group belongs to a bundle
                      P indicates configured to preempt.
                      |
 Interface   Grp  Prio P State    Active addr      Standby addr     Group addr
  Vlan10      10   100    Standby  192.168.10.2     local            192.168.10.1    (conf)
```

N9K-1 的 VLAN 10 SVI 的 MAC 地址为 00ee.ab67.db47，N9K-2 的 VLAN 10 SVI 的 MAC 地址为 00ee.abd8.747f。VLAN 10 的 HSRP 虚拟 MAC 地址为 0000.0c07.ac0a。在此状态下，每台交换机的 MAC 地址表中都包括了每台交换机的 VLAN 10 SVI MAC 地址和 HSRP 虚拟 MAC 地址。每台交换机的VLAN 10 SVI MAC地址和HSRP虚拟MAC地址都存在网关(G)标志，这表示交换机在本地路由发往此MAC地址的数据包。

请注意，在 N9K-1 的 MAC 地址表中，N9K-2 的 VLAN 10 SVI MAC 地址不带网关标记。同样，在 N9K-2 的 MAC 地址表中，N9K-1 的 VLAN 10 SVI MAC 地址不带网关标记。

<#root>

```
N9K-1#

show mac address-table vlan 10
```

```
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
G   10     0000.0c07.ac0a   static   -         F      F    sup-eth1(R)
G   10     00ee.ab67.db47   static   -         F      F    sup-eth1(R)
*   10     00ee.abd8.747f   static   -         F      F    vPC Peer-Link(R)

N9K-2#
```

show mac address-table vlan 10

```
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
G   10     0000.0c07.ac0a   static   -         F      F    vPC Peer-Link(R)
*   10     00ee.ab67.db47   static   -         F      F    vPC Peer-Link(R)
G   10     00ee.abd8.747f   static   -         F      F    sup-eth1(R)
```

我们可以通过 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能。这使交换机能够在本地路由收到的数据包，其中目标MAC地址属于在vPC对等链路上获知的vPC对等设备的MAC地址。这可以通过为交换机的 MAC 地址表中的 vPC 对等体 MAC 地址设置网关标记来完成。

<#root>

```
N9K-1#
```

configure terminal

```
Enter configuration commands, one per line. End with CNTL/Z.
N9K-1(config)#
```

vpc domain 1

```
N9K-1(config-vpc-domain)#
```

peer-gateway

```
N9K-1(config-vpc-domain)#
```

end

```
N9K-1#

N9K-2#
```

configure terminal

```
Enter configuration commands, one per line. End with CNTL/Z.
N9K-2(config)#
```

vpc domain 1

```
N9K-2(config-vpc-domain)#
```

peer-gateway

```
N9K-2(config-vpc-domain)#
end
N9K-2#
```

您可以通过验证 MAC 地址表中的 vPC 对等体 MAC 地址是否带有网关标记来验证 vPC 对等网关增强功能是否按预期运行。

<#root>

```
N9K-1#
show mac address-table vlan 10

Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
G   10     0000.0c07.ac0a   static    -         F      F    sup-eth1(R)
G   10     00ee.ab67.db47   static    -         F      F    sup-eth1(R)
G   10     00ee.abd8.747f   static    -         F      F    vPC Peer-Link(R)

N9K-2#
show mac address-table vlan 10

Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link,
        (T) - True, (F) - False, C - ControlPlane MAC, ~ - vsan
   VLAN     MAC Address      Type      age     Secure NTFY Ports
---------+-----------------+--------+---------+------+----+------------------
G   10     0000.0c07.ac0a   static    -         F      F    vPC Peer-Link(R)
G   10     00ee.ab67.db47   static    -         F      F    vPC Peer-Link(R)
G   10     00ee.abd8.747f   static    -         F      F    sup-eth1(R)
```

影响

启用vPC对等网关增强功能的影响可能因周围的拓扑和所连接主机的行为而异，如下文小节所述。如果以下两个子部分均不适用于您的环境，则启用vPC对等网关增强功能不会造成中断，也不会对您的环境产生影响。

vPC 或 vPC VLAN 上单播路由协议邻接抖动

如果在两个 vPC 对等体与 vPC 连接的路由器（或通过 vPC 孤立端口连接的路由器）之间形成动态单播路由协议邻接，那么在启用了 vPC 对等网关增强功能而没有立即配置路由/第 3 层的时候，路由协议邻接可能会持续抖动。本文档在“带 vPC 对等网关的 vPC 上的单播路由协议邻接故障情形示例”和“带 vPC 对等网关的 vPC VLAN 上的单播路由协议邻接”等部分详描述了这些故障情形。

若要解决此问题，请在使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能后，立即使

用 Layer3 peer-router vPC 域配置命令启用路由/第 3 层 vPC 增强功能。

自动禁用 ICMP 和 ICMPv6 重定向

当vPC对等网关增强功能启用时，所有vPC VLAN SVI（即，与通过vPC对等链路中继的VLAN关联的任何SVI）上都会自动禁用ICMP和ICMPv6重定向数据包的生成。交换机通过在所有 vPC VLAN SVI 上配置 no ip redirects 和 no ipv6 redirects 来实现此目的。这可防止交换机生成 ICMP 重定向数据包，以响应传入了交换机但目的 MAC 和 IP 地址指向交换机 vPC 对等体的的数据包。

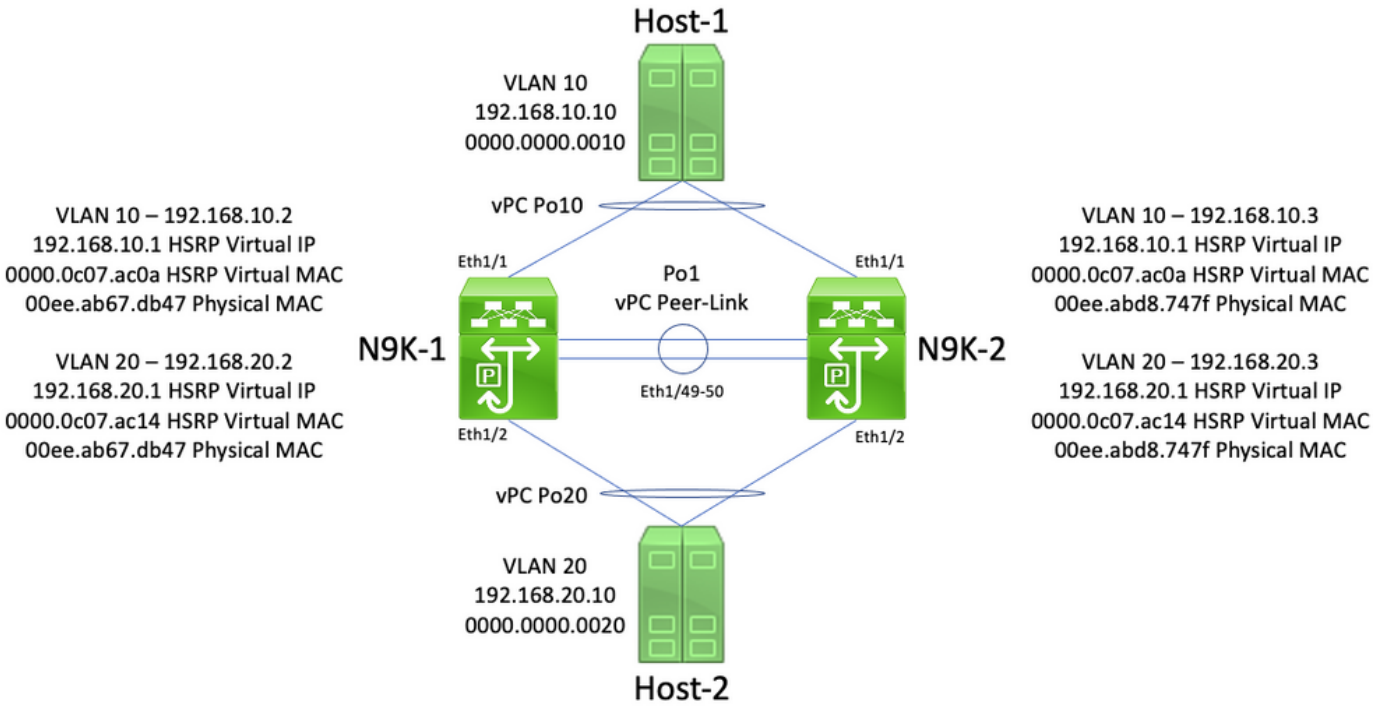
如果特定VLAN内的环境中需要ICMP或ICMPv6重定向数据包，您需要使用peer-gateway exclude-vlan <vlan-id> vPC domain configuration命令排除此VLAN以利用vPC对等网关增强功能。

注意：Nexus 9000系列交换机不支持peer-gateway exclude-vlan <vlan-id> vPC域配置命令。

故障情形示例

vPC 连接的主机出现非标准转发行为

请思考以下拓扑：

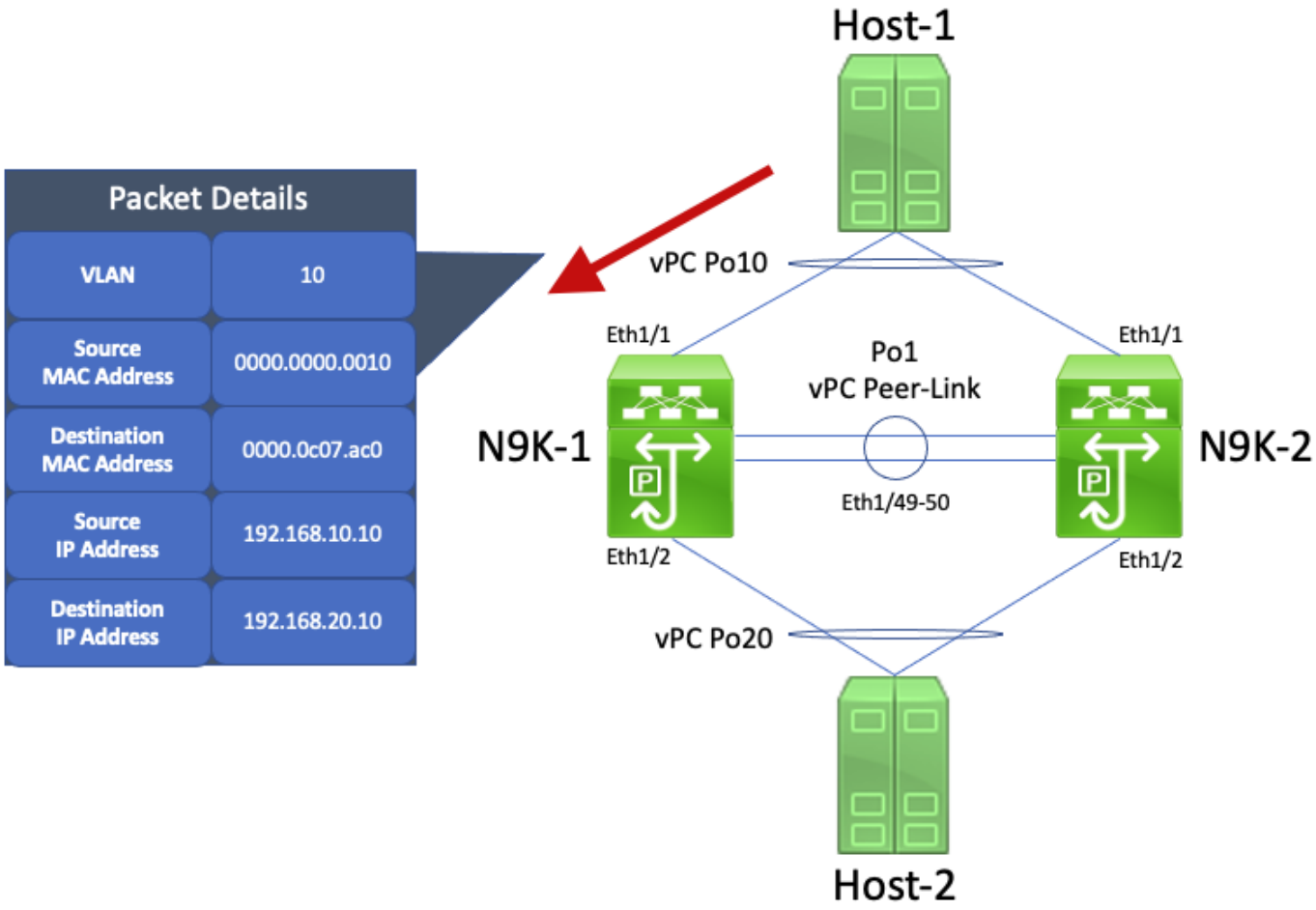


在此拓扑中，N9K-1和N9K-2是vPC域中的vPC对等体，在VLAN 10和VLAN 20之间执行VLAN间路由。接口 Po1 是 vPC 对等链路。名为Host-1的主机通过vPC Po10连接到VLAN 10中的N9K-1和N9K-2。Host-1拥有的IP地址为192.168.10.10,MAC地址为0000.0000.0010。名为Host-2的主机通过vPC Po20连接到VLAN 20中的N9K-1和N9K-2。Host-2拥有的IP地址为192.168.20.10,MAC地址为0000.0000.0020。

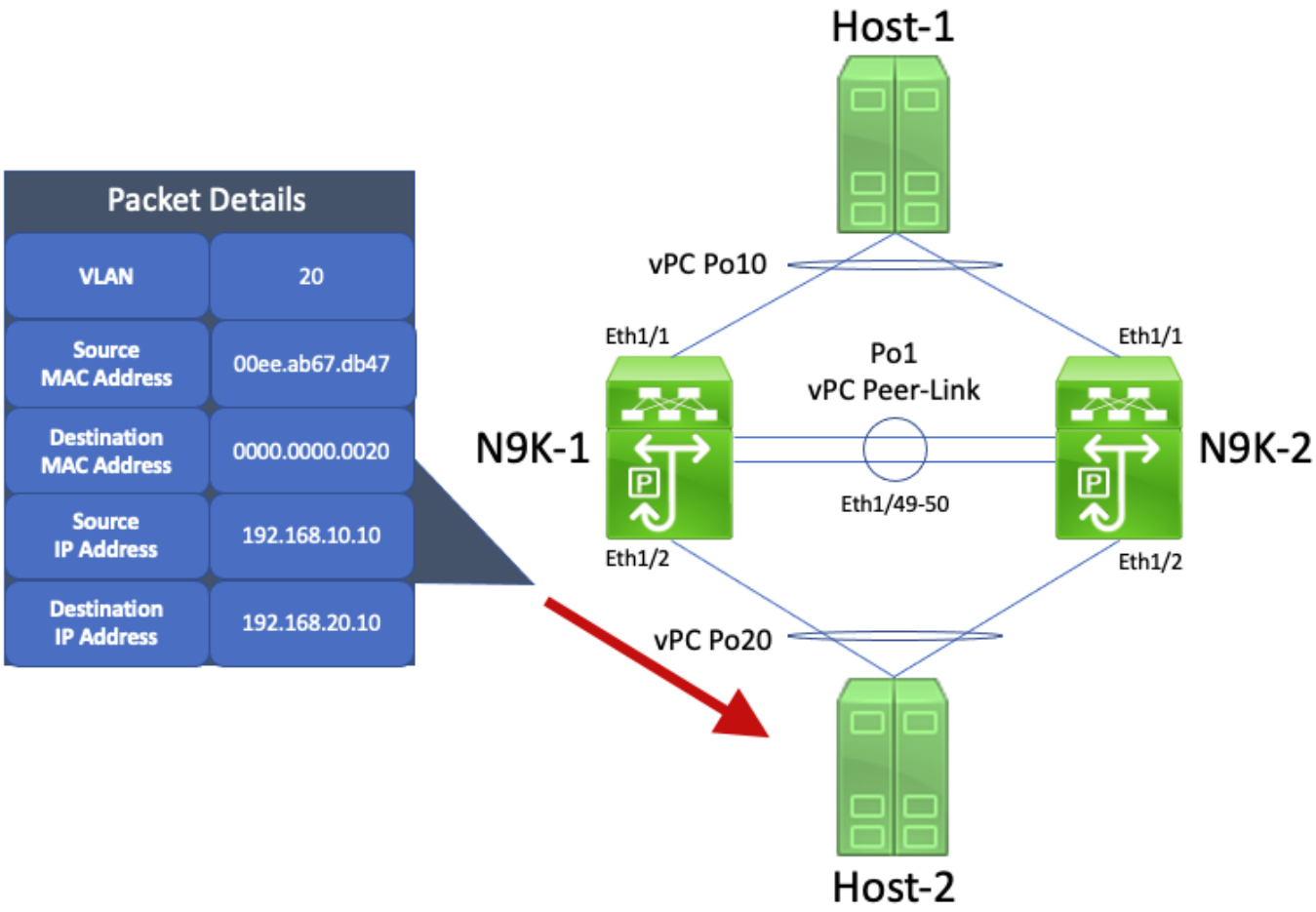
N9K-1 和 N9K-2 在 VLAN 10 和 VLAN 20 中都有 SVI，并在每个 SVI 下都激活了 HSRP。N9K-1的VLAN 10接口的IP地址为192.168.10.2,N9K-1的VLAN 20接口的IP地址为192.168.20.2。两个N9K-

1的SVI的物理MAC地址都是00ee.ab67.db47。N9K-2的VLAN 10接口的IP地址为1921.68.10.3,N9K-2的VLAN 20接口的IP地址为192.168.20.3。两个N9K-2的SVI的物理MAC地址均为00ee.abd8.747f。VLAN 10 的 HSRP 虚拟 IP 地址为 192.168.10.1，HSRP 虚拟 MAC 地址为 0000.0c07.ac0a。VLAN 20 的 HSRP 虚拟 IP 地址为 192.168.20.1，HSRP 虚拟 MAC 地址为 0000.0c07.ac14。

考虑以下场景：Host-1向Host-2发送ICMP回应请求数据包。在Host-1为其默认网关（HSRP虚拟IP地址）解析ARP后，Host-1遵循标准转发行为并生成一个ICMP回应请求数据包，其中源IP地址为192.168.10.10，目的IP地址为192.168.20.10，源MAC地址为000.000.0010，目的MAC地址为000.0c07.ac0a。此数据包从N9K-1流出。此处显示了一个直观示例。

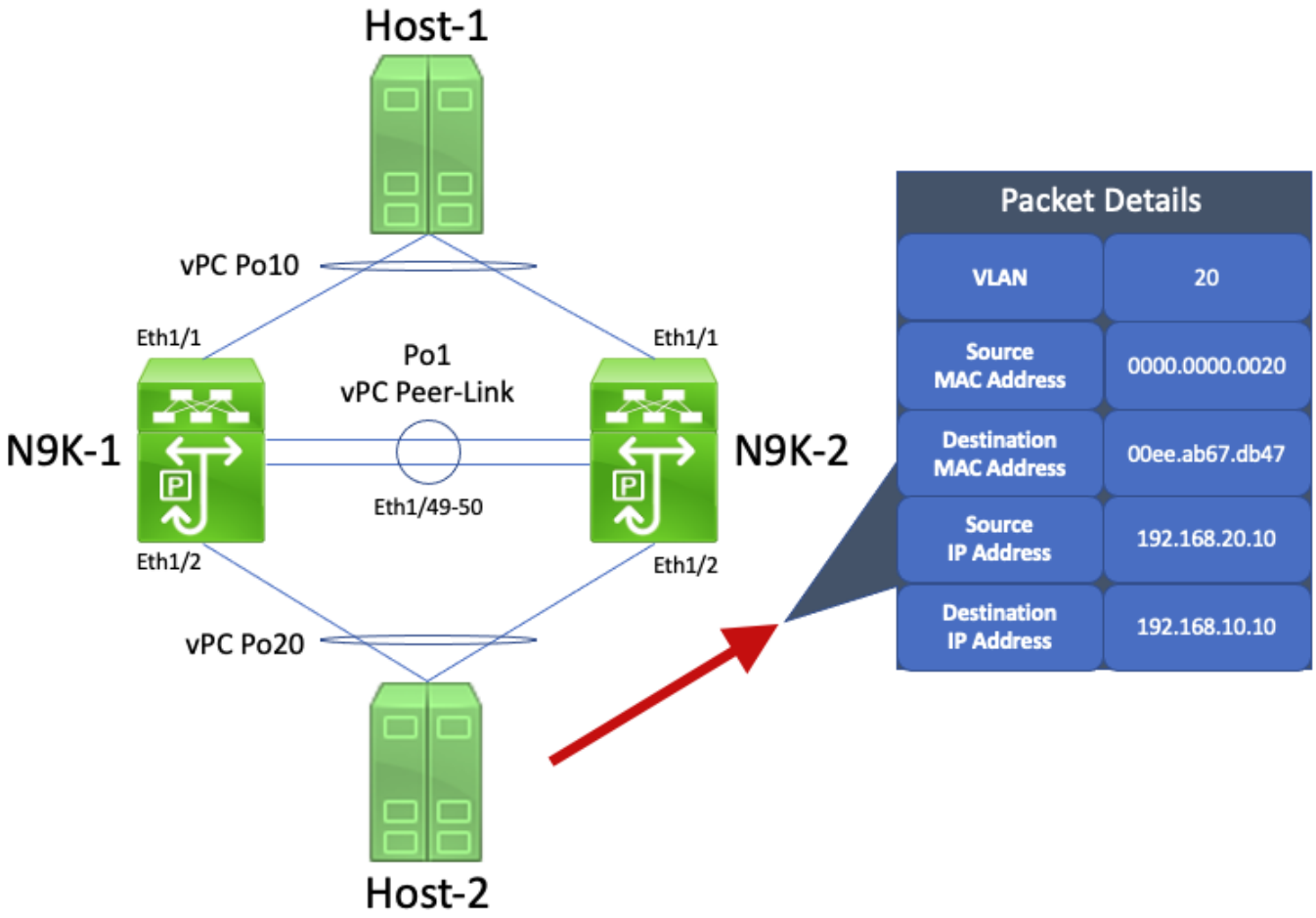


N9K-1 收到此数据包。由于此数据包发往 HSRP 虚拟 MAC 地址，因此无论其 HSRP 控制平面状态如何，N9K-1 能够根据本地路由表路由此数据包。此数据包从VLAN 10路由到VLAN 20。作为数据包路由的一部分，N9K-1通过重新寻址数据包的源MAC地址和目的MAC地址字段来执行数据包重写。数据包的新源MAC地址是与N9K-1的VLAN 20 SVI关联的物理MAC地址(00ee.ab67.db47)，新的目标MAC地址是与Host-2关联的MAC地址(0000.0000.0020)。此处显示了一个直观示例。

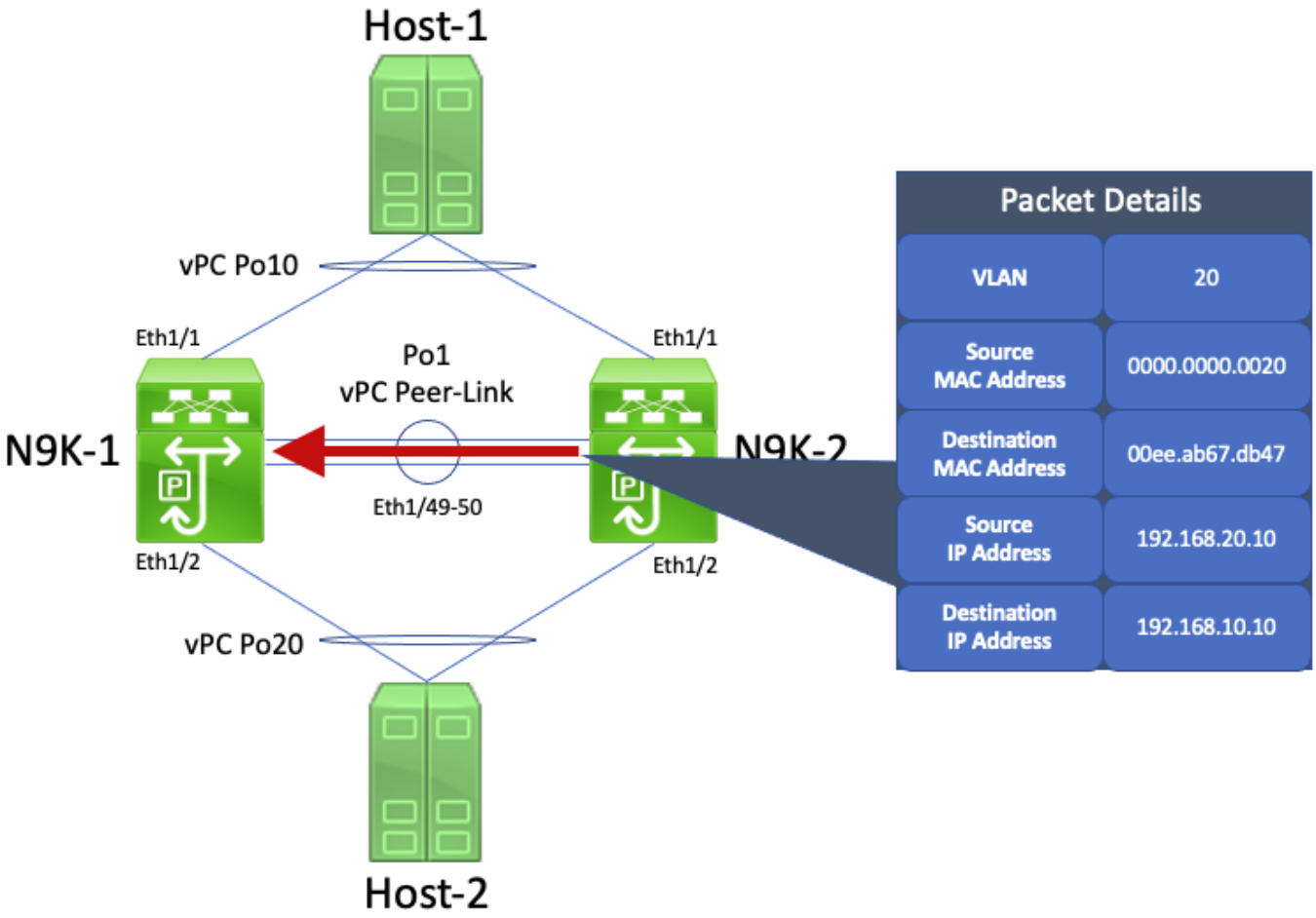


Host-2 收到此数据包并生成 ICMP 回应应答数据包，以响应 Host-1 的 ICMP 回应请求数据包。但是，当 Host-2 不遵循标准转发行为时，为了优化其转发，Host-2 不会为 Host-1 的 IP 地址 (192.168.10.10) 查找路由表或 ARP 缓存，而是翻转原先收到的 ICMP 回应请求数据包主机的源 MAC 地址和目的 MAC 地址字段。因此，Host-2生成的ICMP回应应答数据包的源IP地址为192.168.20.10，目的IP地址为192.168.10.10，源MAC地址为0000.0000.0020，目的MAC地址为00ee.ab67.db47。

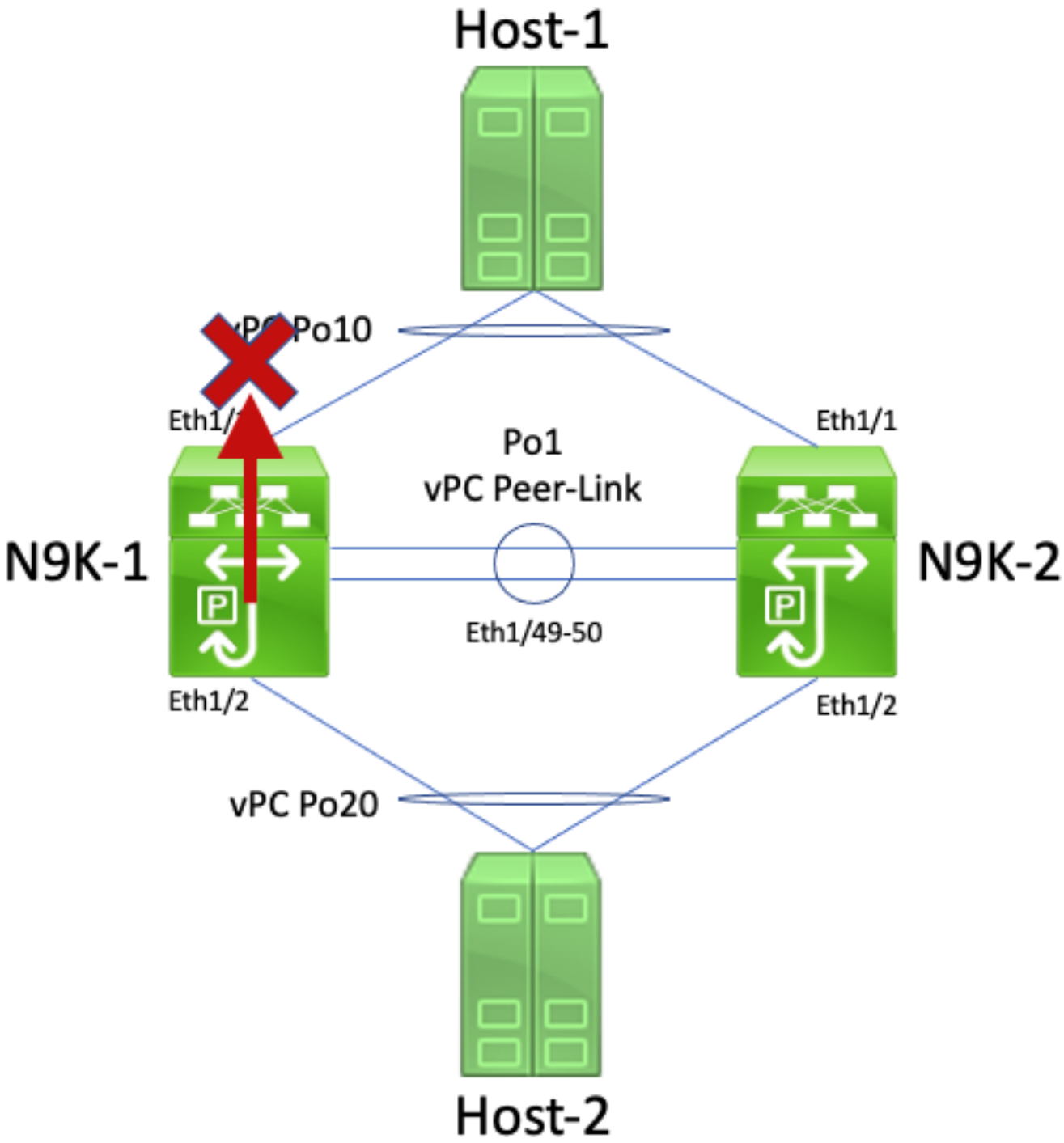
如果此ICMP回应应答数据包流向N9K-1，则将该数据包转发到Host-1而不会出现问题。但是，请考虑此 ICMP 回应应答数据包向 N9K-2 发出的情形，如下所示。



N9K-2 收到此数据包。由于此数据包的目的地是N9K-1的VLAN 20 SVI的物理MAC地址，因此N9K-2通过vPC对等链路将此数据包转发到N9K-1，因为N9K-2无法代表N9K-1路由此数据包。此处显示了一个直观示例。



N9K-1 收到此数据包。由于此数据包发往 N9K-1 的 VLAN 20 的 SVI 物理 MAC 地址，因此无论其 HSRP 控制平面状态如何，N9K-1 都能够根据本地路由表路由此数据包。此数据包从VLAN 20路由到VLAN 10。但是，此路由的出口接口解析为vPC Po10（在N9K-2上启用）。这违反了vPC环路避免规则 — 如果N9K-1通过vPC对等链路收到数据包，如果N9K-2上启用同一vPC接口，则N9K-1无法将该数据包从vPC接口转发出去。由于此违规，N9K-1丢弃此数据包。此处显示了一个直观示例。



您可以通过使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能来解决此问题。这允许N9K-2代表N9K-1路由ICMP回应应答数据包（以及以类似方式寻址的其他数据包），即使数据包的目的MAC地址由N9K-1而不是N9K-2所有。因此，N9K-2可以从其vPC Po10接口转发此数据包，而不是通过vPC对等链路转发此数据包。

vPC 上的路由/第 3 层（第 3 层对等路由器）

本部分介绍通过 layer3 peer-router vPC 域配置命令启用 vPC 路由/第 3 层增强功能。

注意：启用vPC上的路由/第3层增强后，不支持在vPC上形成组播路由协议邻接（即协议无关组播[PIM]邻接）。

概述

在某些环境中，客户希望通过 vPC 将路由器连接到一对 Nexus 交换机，并通过 vPC 与两个 vPC 对等体形成单播路由协议邻接。或者，客户可能希望通过 vPC VLAN 将路由器连接到单个 vPC 对等体，并在 vPC VLAN 上与两个 vPC 对等体形成单播路由协议邻接。因此，vPC 连接的路由器将为两台 Nexus 交换机通告的前缀提供等价多路径 (ECMP) 路由。这可能优于在 vPC 连接的路由器和两个 vPC 对等体之间使用专用路由链路，以节省 IP 地址（需要 3 个 IP 地址而不是 4 个 IP 地址），或降低配置复杂性（SVI 旁边的路由接口，尤其是在需要子接口的 VRF-Lite 环境中）。

过去，思科 Nexus 平台不支持通过 vPC 形成单播路由协议邻接。但是，即使不受支持。客户可能已经实施了某种拓扑， 让 vPC 在其中形成单播路由协议邻接而不会出现问题。在网络中发生某些变化（例如，vPC 连接的路由器或 vPC 对等体的软件升级、防火墙故障切换等）后，vPC上 的单播路由协议邻接停止工作，导致数据平面流量丢包，或单播路由协议邻接无法与一个或两个 vPC 对等体上建立起来。本文档的“故障情形示例”部分将讨论这些失败和不受支持的场景背后的技术细节。

我们引入了基于 vPC 的路由/第 3 层增强功能，以更多地支持在 vPC 上形成单播路由协议邻接。实现的方法是允许在 vPC 对等链路上转发 TTL 为 1 的单播路由协议数据包，同时不递减数据包的 TTL。因此，可以在 vPC 或 vPC VLAN 上形成单播路由协议邻接而不会出现问题。在使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能后，可以使用第 3 层 peer-router vPC 域配置命令启用路由/第 3 层 vPC 增强功能。

NX-OS 软件版本引入了对每个思科 Nexus 平台的路由/第 3 层 vPC 增强功能的支持，详情请参阅 Nexus 平台说明文档中的“通过虚拟端口通道进行路由所支持的拓扑”（详见表 2：“vPC VLAN 上的路由协议邻接支持”）。

注意事项

偶发 VPC-2-L3_VPC_UNEQUAL_WEIGHT 系统日志

启用vPC上的路由/第3层增强功能后，两台vPC对等设备每小时都会生成类似于以下其中一种的系统日志：

```
2021 May 26 19:13:47.079 switch %VPC-2-L3_VPC_UNEQUAL_WEIGHT: Layer3 peer-router is enabled. Please mak
2021 May 26 19:13:47.351 switch %VPC-2-L3_VPC_UNEQUAL_WEIGHT: Unequal weight routing is not supported i
```

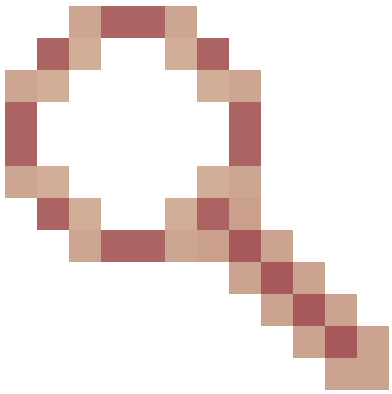
这两种系统日志均不表示交换机存在问题。这些系统日志向管理员发出警告：在启用 vPC上的路由/第 3 层增强功能时，为了确保两个 vPC 对等体能够以同样方式路由流量，两个 vPC 对等体的路由配置、开销和权重应相同。它不一定表示任一 vPC 对等体上存在不匹配的路由、配置开销或权重。

这些系统日志可通过此处所示的配置禁用。

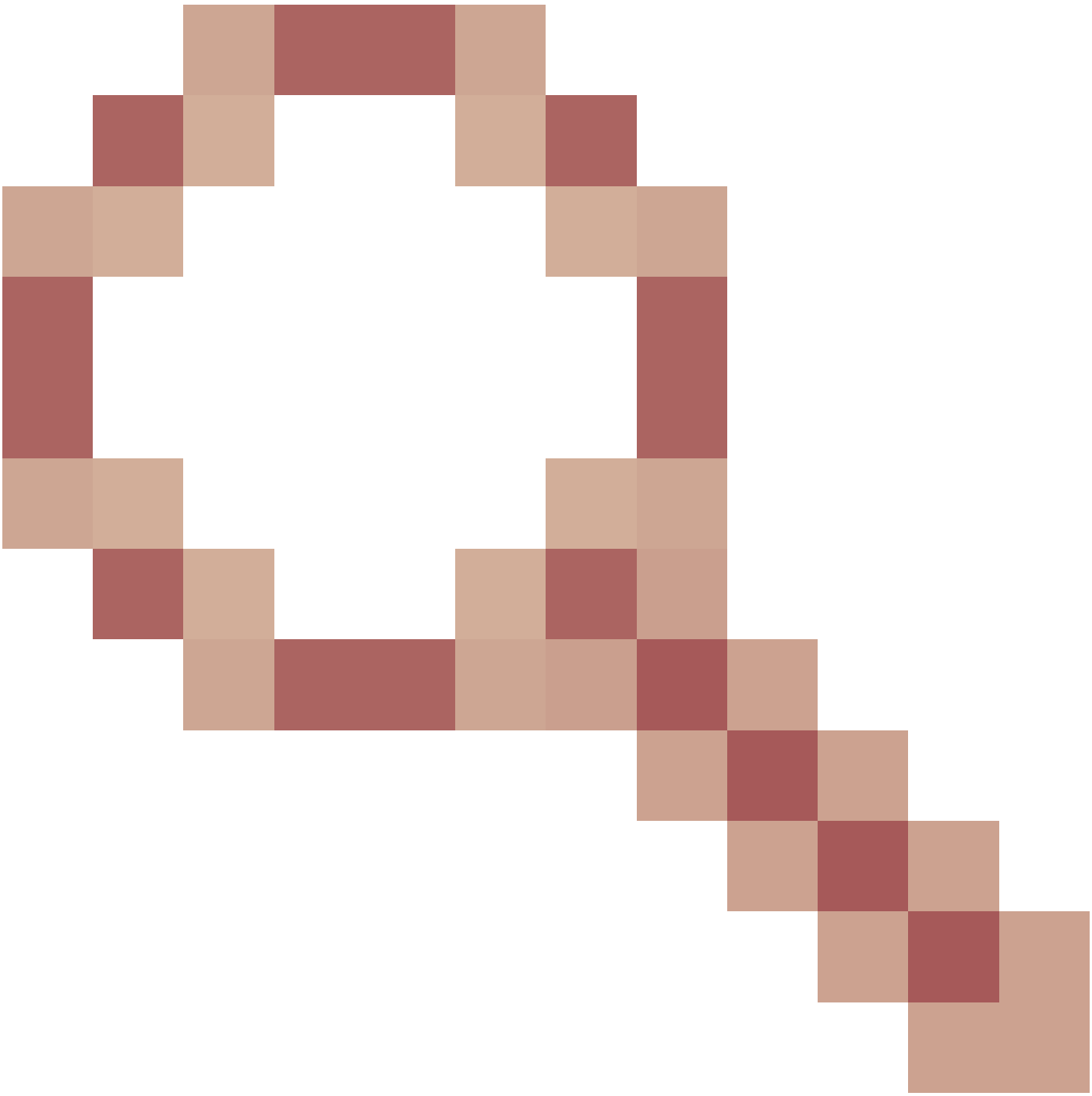
<#root>

```
switch#
configure terminal
switch(config)#
vpc domain 1
switch(config-vpc-domain)#
no layer3 peer-router syslog
switch(config-vpc-domain)#
end
switch#
```

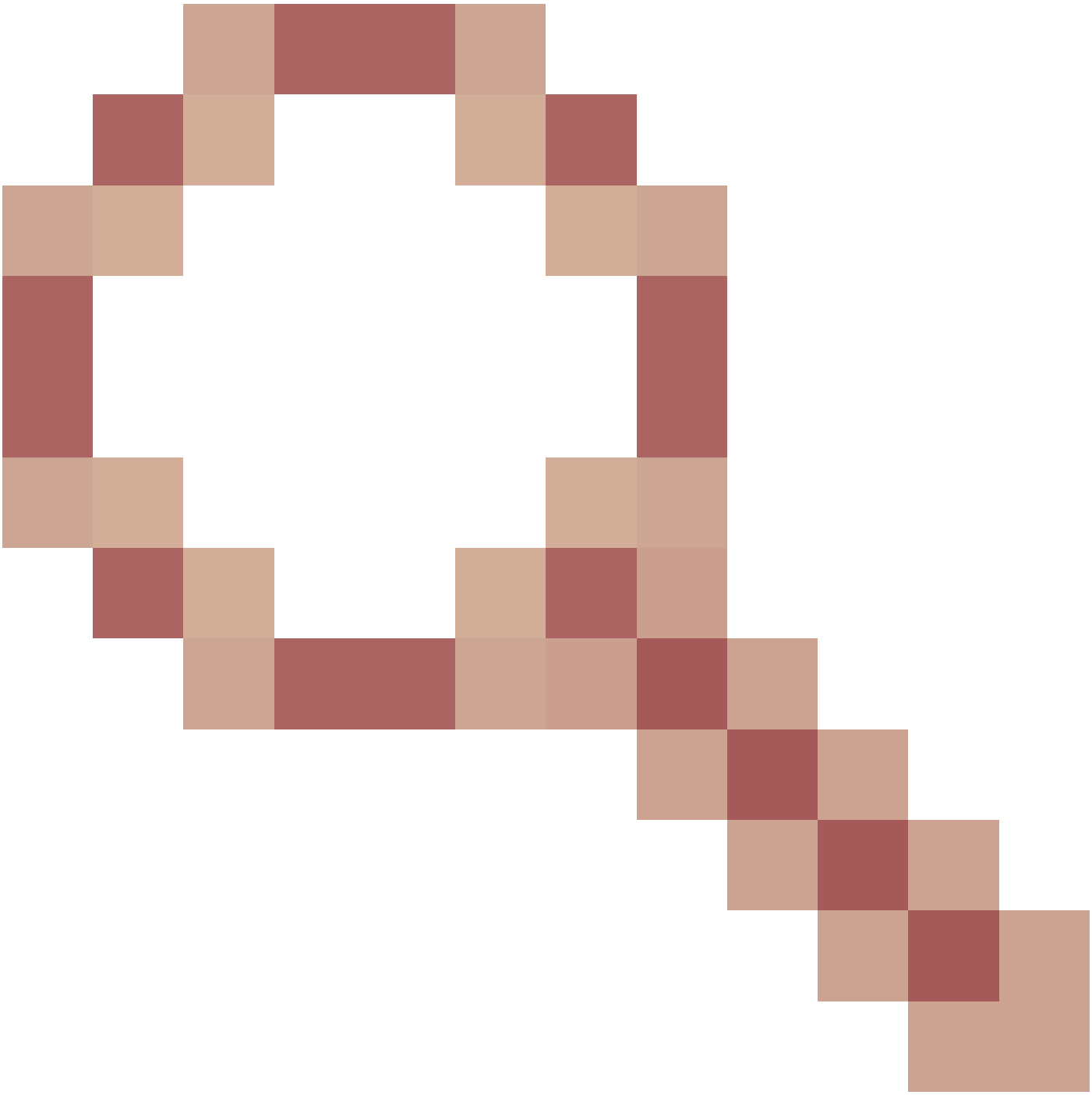
需要在两个vPC对等设备上执行此配置，以禁用两个vPC对等设备上的系统日志。



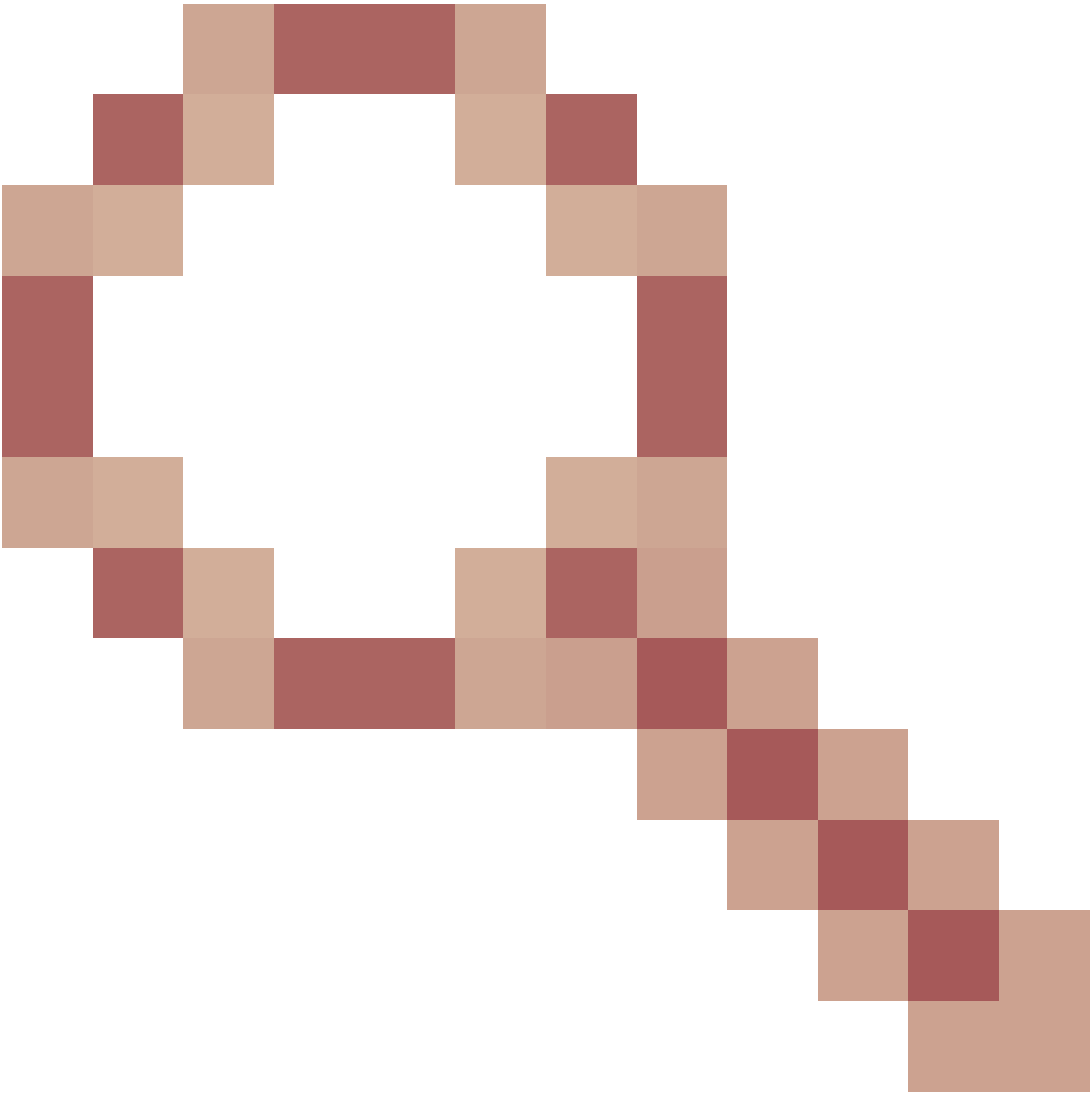
由于Cisco Bug ID CSCvs82183，已转发TTL为1的软件的数据平面流量
和Cisco Bug ID CSCvw16965



在配备云扩展ASIC且运行NX-OS软件版本9.3(6)之前的Nexus 9000系列交换机上启用vPC上的路由/第3层增强功能后，不与TTL为1的单播路由协议关联的数据平面流量将被传送给管理引擎，并在软件中而不是硬件中转发。根据Nexus交换机是固定机箱（也称为“架顶式”）交换机还是模块化机箱（也称为“行尾式”）交换机，以及交换机当前的NX-OS软件版本，此问题的根本原因可能是软件缺陷Cisco Bug ID CSCvs82183



或软件缺陷Cisco Bug ID CSCvw16965



.这两个软件缺陷仅影响配备云扩展ASIC的Nexus 9000系列交换机 — 其他Cisco Nexus硬件平台不会受到这两个问题的影响。有关详细信息，请参阅每个软件缺陷中的信息。

为避免这些软件缺陷，思科建议升级到 NX-OS 软件版本 9.3(6) 或更高版本。作为一般建议，思科建议定期升级到当前推荐的适用于 Nexus 9000 系列交换机的 NX-OS 软件版本（参见文档思科 Nexus 9000 系列交换机的思科推荐 NX-OS 版本）。

配置

请点击此处，详细了解如何通过 vPC 增强配置路由/第 3 层的示例。

在本例中，N9K-1 和 N9K-2 是 vPC 域中的 vPC 对等体。两个 vPC 对等体都已启用 vPC 对等网关增强功能，这是启用 vPC 上的路由/第 3 层增强功能必需的前提条件。两个vPC对等体在VLAN 10中有一个SVI，VLAN 10在OSPF进程1下启用。N9K-1和N9K-3与vPC连接的OSPF路由器（IP地

址和邻居ID为192.168.10.3) 处于OSPF EXSTART/EXCHANGE状态。

<#root>

```
N9K-1#
show running-config vpc

<snip>
vpc domain 1
  role priority 150
  peer-keepalive destination 10.122.190.196
  peer-gateway

interface port-channel1
  vpc peer-link

N9K-2#
show running-config vpc

<snip>
vpc domain 1
  peer-keepalive destination 10.122.190.195
  peer-gateway

interface port-channel1
  vpc peer-link

N9K-1#
show running-config interface Vlan10


interface Vlan10
  no shutdown
  no ip redirects
  ip address 192.168.10.1/24
  no ipv6 redirects
  ip router ospf 1 area 0.0.0.0

N9K-2#
show running-config interface Vlan10


interface Vlan10
  no shutdown
  no ip redirects
  ip address 192.168.10.2/24
  no ipv6 redirects
  ip router ospf 1 area 0.0.0.0

N9K-1#
show running-config ospf



feature ospf

router ospf 1
```

```
interface Vlan10
  ip router ospf 1 area 0.0.0.0

N9K-2#
```

show running-config ospf

```
feature ospf

router ospf 1

interface Vlan10
  ip router ospf 1 area 0.0.0.0

N9K-1#
```

show ip ospf neighbors

```
 OSPF Process ID 1 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 192.168.10.2      1 TWOWAY/DROTHER   00:08:10 192.168.10.2    Vlan10
 192.168.10.3      1 EXCHANGE/BDR     00:07:43 192.168.10.3    Vlan10

N9K-2#
```

show ip ospf neighbors

```
 OSPF Process ID 1 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 192.168.10.1      1 TWOWAY/DROTHER   00:08:21 192.168.10.1    Vlan10
 192.168.10.3      1 EXSTART/BDR      00:07:48 192.168.10.3    Vlan10
```

我们可以通过 layer3 peer-router vPC 域配置命令启用基于 vPC 的路由/第 3 层增强功能。这可以防止vPC对等设备降低因vPC对等网关增强功能而路由的单播路由协议数据包的TTL。

<#root>

```
N9K-1#
```

configure terminal

```
Enter configuration commands, one per line. End with CNTL/Z.
N9K-1(config)#
```

vpc domain 1

```
N9K-1(config-vpc-domain)#
```

layer3 peer-router

```
N9K-1(config-vpc-domain)#
```

end

```
N9K-1#
```

```
N9K-2#
configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
N9K-2(config)#
vpc domain 1
N9K-2(config-vpc-domain)#
layer3 peer-router
N9K-2(config-vpc-domain)#
end
N9K-2#
```

您可以验证 vPC 上的路由/第 3 层增强功能是否按预期运行，方法是在启用该功能后，验证与 vPC 连接的 OSPF 邻居的 OSPF 邻接很快过渡到“FULL”状态。

```
<#root>
N9K-1#
show ip ospf neighbors

 OSPF Process ID 1 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 192.168.10.2      1 TWOWAY/DROTHER   00:12:17 192.168.10.2    Vlan10
 192.168.10.3      1 FULL/BDR         00:00:29 192.168.10.3    Vlan10

N9K-2#
show ip ospf neighbors

 OSPF Process ID 1 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 192.168.10.1      1 TWOWAY/DROTHER   00:12:27 192.168.10.1    Vlan10
 192.168.10.3      1 FULL/BDR         00:00:19 192.168.10.3    Vlan10
```

影响

在 vPC 上启用路由/第 3 层增强功能不会对 vPC 域造成任何本质上的影响。这意味着当您启用 vPC上的路由/第3层增强功能时，vPC对等设备不会暂停任何vPC，也不会因启用此增强功能而影响任何数据平面流量。

但是，如果之前由于未启用 vPC 上的路由/第 3 层增强功能而关闭的动态路由协议邻接因启用此增强功能而突然开启，则根据受影响的路由协议邻接的角色，通过这些邻接通告的特定前缀，以及单播路由表的当前状态，在启用路由/第 3 层增强功能时可能会出现某些中断。

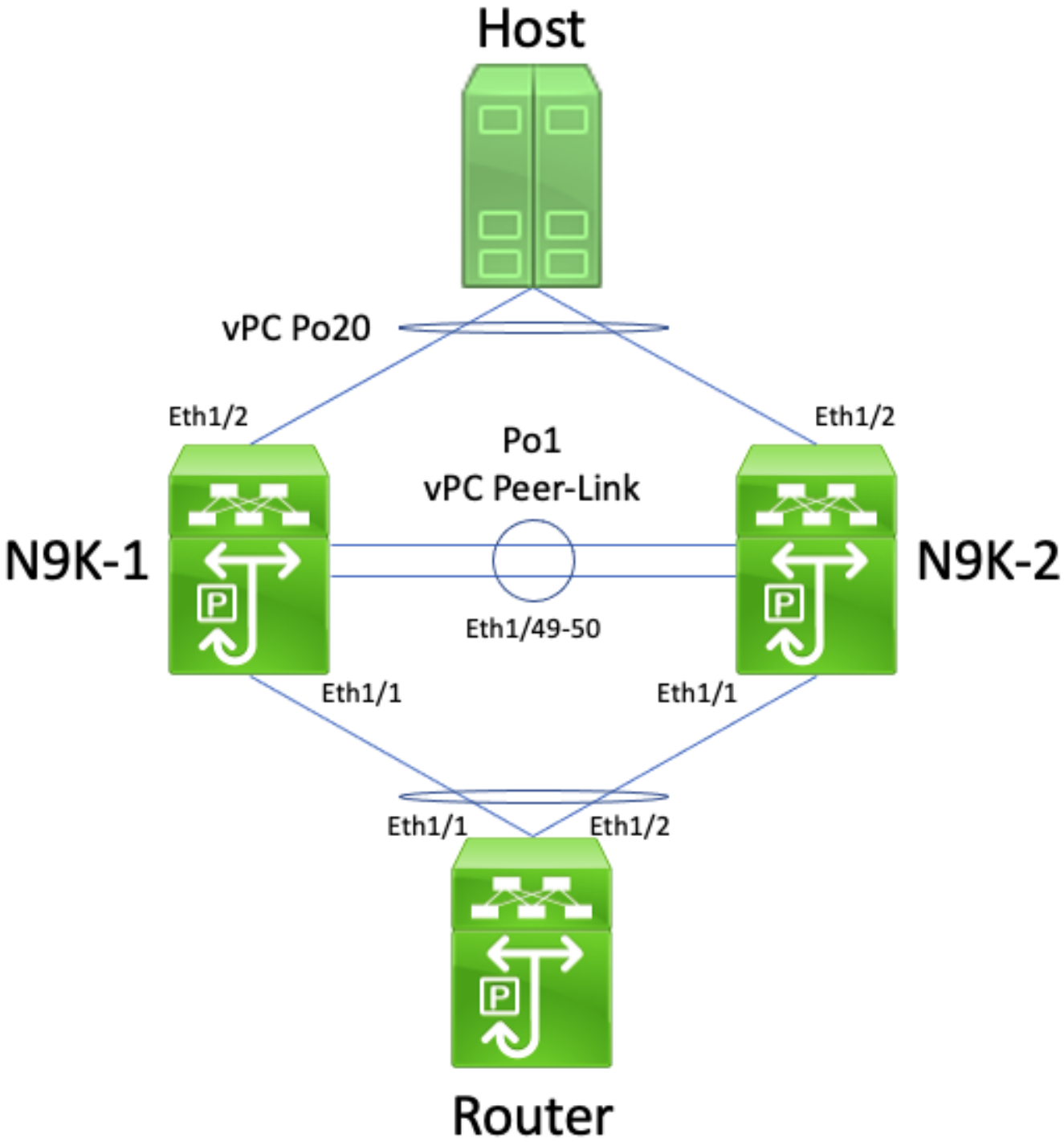
因此，思科建议客户在维护时段启用此增强功能，并预期可能会出现控制平面和数据平面中断，除非客户非常确信受影响的路由协议邻接关系不会显着影响网络运行。

思科还建议仔细查看本文档的"警告"部分，了解影响 NX-OS 软件版本的任何软件缺陷，这些缺陷可能导致 TTL 为 1 的自然数据平面流量由软件而非硬件处理。

故障情形示例

无 vPC 对等网关的 vPC 上的单播路由协议邻接

考虑图中显示的拓扑：



在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是未启用 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。接口 Po1 是 vPC 对等链路。主机名为Router的路由器通过vPC Po10连接到N9K-1和N9K-2。主机通过vPC Po20连接到N9K-1和N9K-2。路由器的Po10接口是在单播路由协议下激活的路由端口通道。N9K-1 和 N9K-2 都使用相同的单播路由协议激活 SVI 接口，并且与路由器位于同一广播域中。

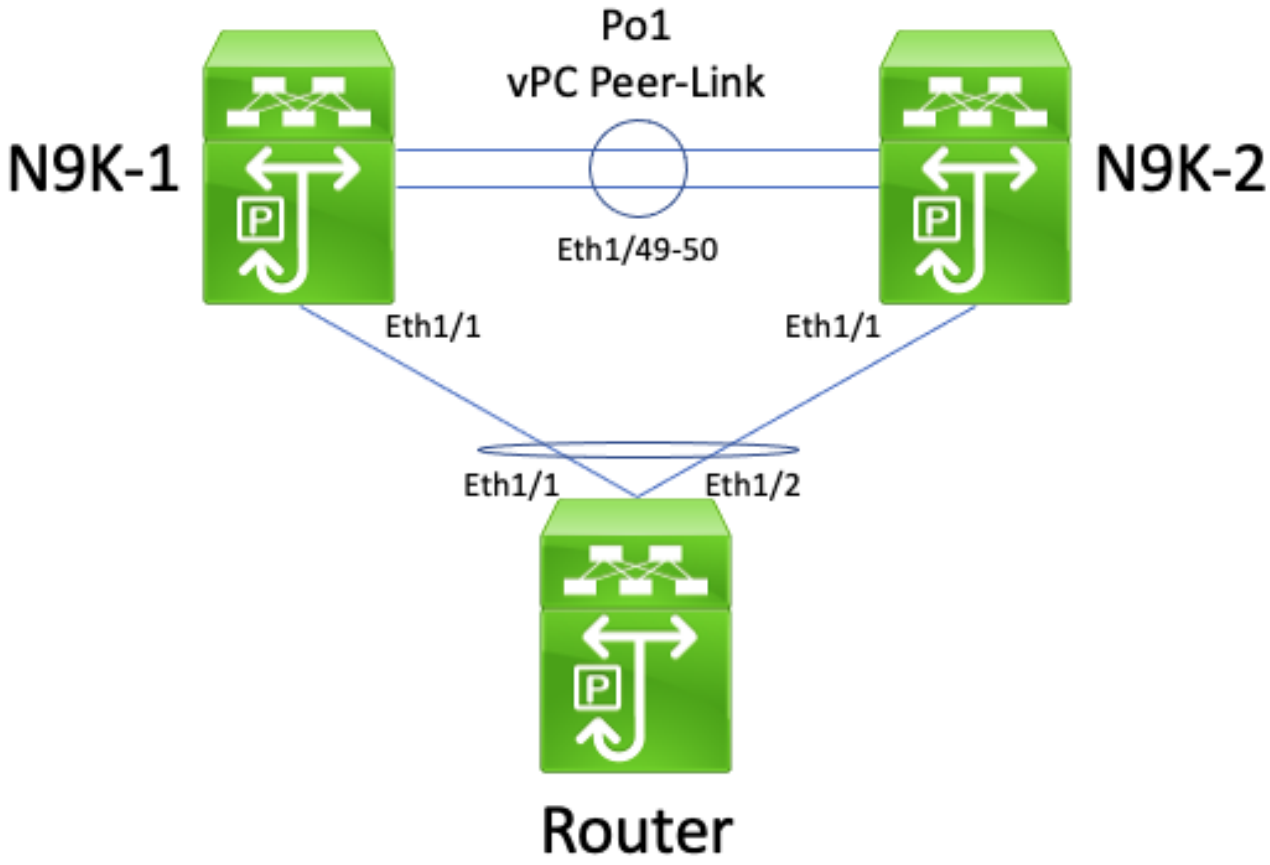
不支持未启用 vPC 对等网关增强的 vPC 上的单播路由协议邻接，因为 vPC 连接的路由器的 ECMP 散列决定和其第 2 层端口通道散列决定可能不同。在此拓扑中，路由器、N9K-1和N9K-2之间将成功形成路由协议邻接。考虑路由器和主机之间的流量传输。发往主机的数据平面流量经过路由器时，其目的 MAC 地址可能会重写为 N9K-1 的 SVI MAC 地址（路由器做出的 ECMP 散列决定），但会从接口 Ethernet1/2 传出（路由器做出的第 2 层 端口通道散列决策）。

N9K-2收到此数据包并通过vPC对等链路转发它，因为目的MAC地址属于N9K-1，并且vPC对等网关增强功能（允许N9K-2代表N9K-1路由数据包）未启用。N9K-1在vPC对等链路上接收此数据包，并识别出需要在vPC Po20中将数据包从其Ethernet1/2转发出去。这违反了vPC环路避免规则，因此N9K-1在硬件中丢弃数据包。因此，针对通过此拓扑中的 vPC 域的某些流，您可能会观察到连接问题或丢包现象。

您可以通过使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能，然后使用 layer3 peer-router vPC 域配置命令启用 vPC 上的路由/第 3 层增强功能来解决此问题。为了最大限度地减少中断，您应快速连续启用这两个 vPC 增强功能，以使“带 vPC 对等网关的 vPC 上的单播路由协议邻接”描述的故障情形不会发生。

带 vPC 对等网关的 vPC 上的单播路由协议邻接

考虑图中显示的拓扑：



在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。接口 Po1 是 vPC 对等链路。主机名为Router的路由器通过vPC Po10连接到N9K-1和N9K-2。路由器的Po10接口是在单播路由协议下激活的路由端口通道。N9K-1 和 N9K-2 都使用相同的单播路由协议激活 SVI 接口，并且与路由器位于同一广播域中。

不支持已启用 vPC 对等网关增强功能的 vPC 上的单播路由协议邻接，因为 vPC 对等网关增强功能可能会阻止 vPC 连接的路由器和两个 vPC 对等体之间形成单播路由协议邻接。在此拓扑中，路由器与N9K-1或N9K-2之间的路由协议邻接关系可能无法按预期建立，具体取决于路由器通过vPC Po10向N9K-1或N9K-2散列发起的单播路由协议数据包的方式。

所有路由器都能够发送和接收本地链路组播路由协议数据包（通常称为“Hello”数据包），因为这些数据包已成功泛洪到 vPC VLAN。但是，请考虑以下场景：由于路由器的第 2 层端口通道散列决定，从路由器发往 N9K-1 的单播路由协议数据包将通过 Ethernet1/2 接口传出到 N9K-2。此数据包的目的地为N9K-1的SVI MAC地址，但进入N9K-2的Ethernet1/1接口。N9K-2发现数据包的目的地是N9K-1的SVI MAC地址，由于启用了vPC对等网关增强功能，该地址安装在N9K-2的MAC地址表中，带有“G”或“网关”标志。因此，N9K-2尝试代表N9K-1本地路由单播路由协议数据包。

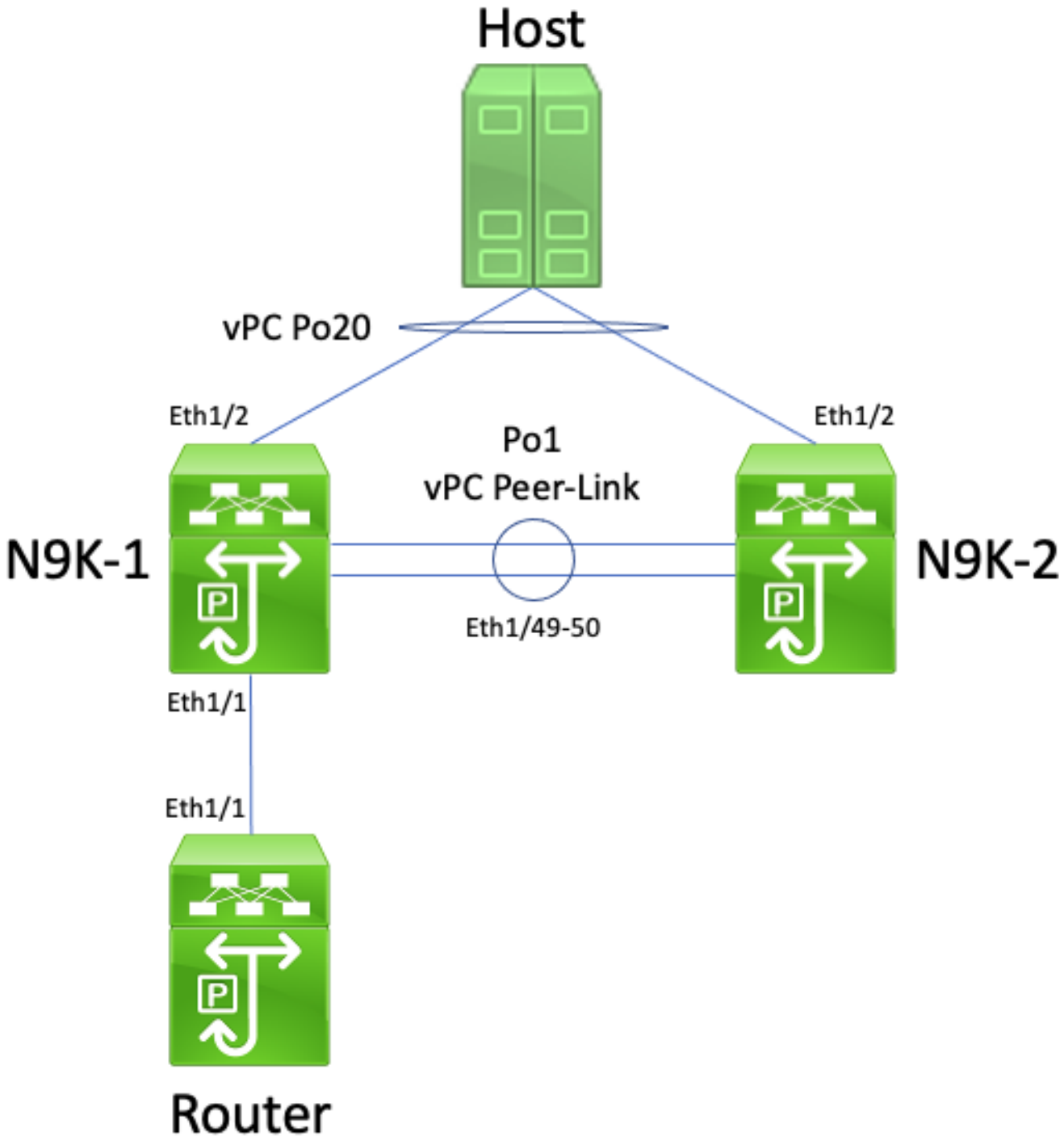
但是，通过路由数据包，数据包的生存时间(TTL)会递减，大多数单播路由协议数据包的TTL为1。因此，数据包的TTL会减为0，然后丢弃N9K-2。从N9K-1的角度来看，N9K-1正在接收来自路由器的本地链路组播路由协议数据包，能够向路由器发送单播路由协议数据包，但是没有接收来自路由器的单播路由协议数据包。因此，N9K-1断开了路由协议与路由器的邻接关系，并重新启动路由协议的本地有限状态机。同样，路由器会重新启动其用于路由协议的本地有限状态机。

我们可以通过 layer3 peer-router vPC 域配置命令启用基于 vPC 的路由/第 3 层增强功能以解决此问

题。结果是允许在 vPC 对等链路上转发 TTL 为 1 的单播路由协议数据包，而无需递减数据包的 TTL。因此，可以在 vPC 或 vPC VLAN 上形成单播路由协议邻接而不会出现问题。

无 vPC 对等网关的 vPC VLAN上 的单播路由协议邻接

考虑图中显示的拓扑：



在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是未启用 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。接口 Po1 是 vPC 对等链路。主机名为Router的路由器通过Ethernet1/1连接到N9K-1的Ethernet1/1。路由器的Ethernet1/1接口是在单播路由协议下激活的路由接口。N9K-1 和 N9K-2 都使用相同的单播路由协议激活 SVI 接口，并且与路由器位于同一广播域中。

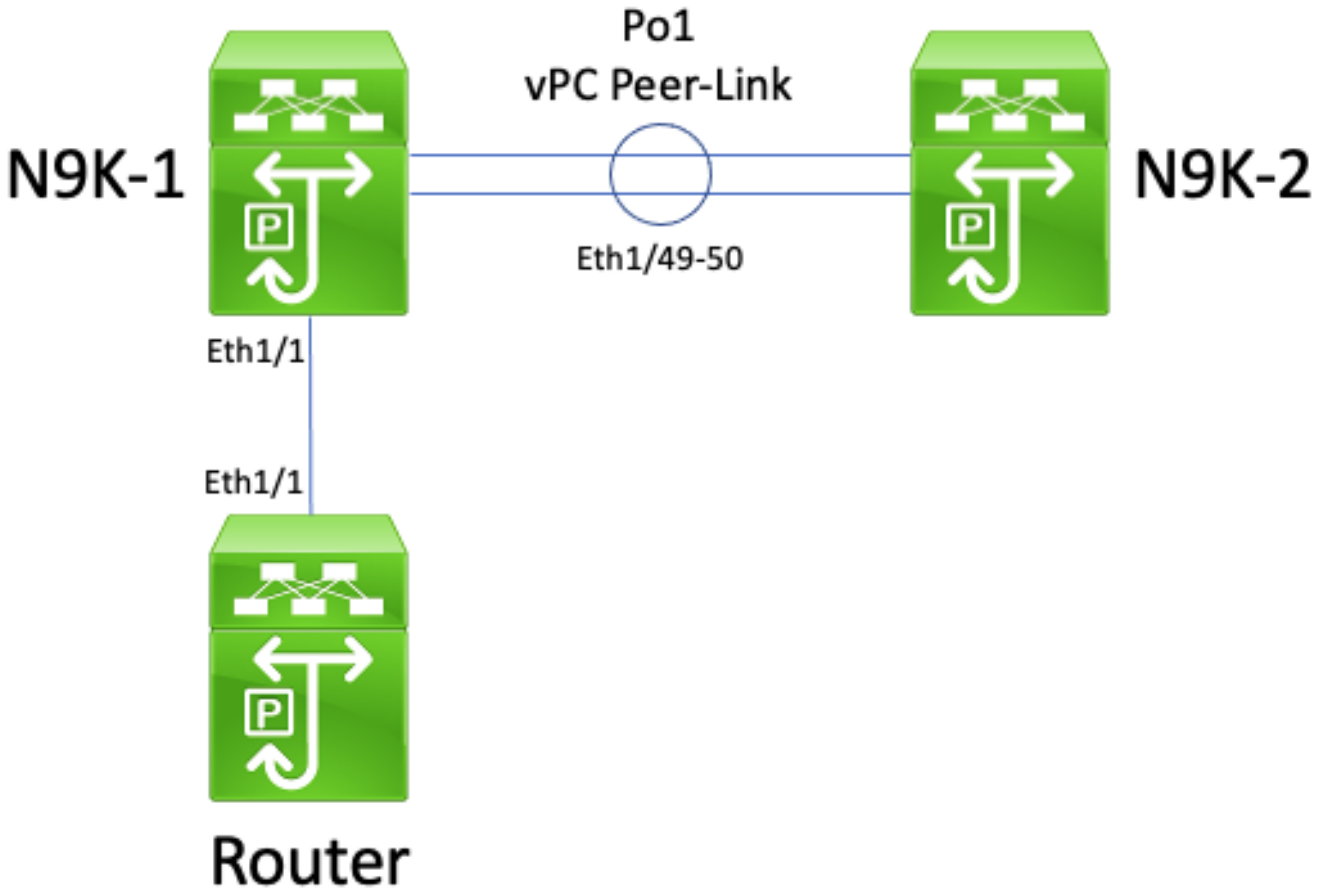
不支持未启用 vPC 对等网关增强功能的 vPC VLAN 上的单播路由协议邻接，因为连接 vPC VLAN 的路由器的 ECMP 散列决策可能导致 N9K-2 因违反 vPC 环路规避规则而丢弃数据平面流量。在此拓扑中，路由器、N9K-1和N9K-2之间将成功形成路由协议邻接。考虑路由器和主机之间的流量传输。发往主机的数据平面流量经过路由器时，其目的 MAC 地址可能会重写为 N9K-2 的 SVI MAC 地址（路由器做出的 ECMP 散列决定），并从接口 Ethernet1/1 传出到 N9K-1。

N9K-1收到此数据包，并通过vPC对等链路转发它，因为目的MAC地址属于N9K-2，并且vPC对等网关增强功能（允许N9K-1代表N9K-2路由数据包）未启用。N9K-2在vPC对等链路上收到此数据包，并识别出它需要在vPC Po20中将数据包从其Ethernet1/2转发出去。这违反了vPC环路避免规则，因此N9K-2在硬件中丢弃数据包。因此，针对通过此拓扑中的 vPC 域的某些流，您可能会观察到连接问题或丢包现象。

您可以通过使用 peer-gateway vPC 域配置命令启用 vPC 对等网关增强功能，然后使用 layer3 peer-router vPC 域配置命令启用 vPC 上的路由/第 3 层增强功能来解决此问题。为了最大限度地减少中断，您应快速连续启用这两个 vPC 增强功能，以使“带 vPC 对等网关的 vPC 上的单播路由协议邻接”描述的故障情形不会发生。

带 vPC 对等网关的 vPC VLAN 上的单播路由协议邻接

考虑图中显示的拓扑：



在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。接口 Po1 是 vPC 对等链路。主机名为Router的路由器通过Ethernet1/1连接到N9K-1的

Ethernet1/1。路由器的Ethernet1/1接口是在单播路由协议下激活的路由接口。N9K-1 和 N9K-2 都使用相同的单播路由协议激活 SVI 接口，并且与路由器位于同一广播域中。

不支持vPC VLAN上启用了vPC对等网关增强的单播路由协议邻接，因为vPC对等网关增强可防止在vPC VLAN连接的路由器和vPC VLAN连接的路由器未直接连接的vPC对等之间形成单播路由协议邻接。在此拓扑中，由于启用了vPC对等网关增强功能，N9K-1路由单播路由协议数据包发往N9K-2的SVI MAC地址，因此路由器和N9K-2之间的路由协议邻接关系无法按预期建立。由于数据包正在经路由传输，其生存时间 (TTL) 必须递减。单播路由协议数据包的 TTL 通常为 1，将数据包的 TTL 递减为 0 的路由器必须丢弃该数据包。

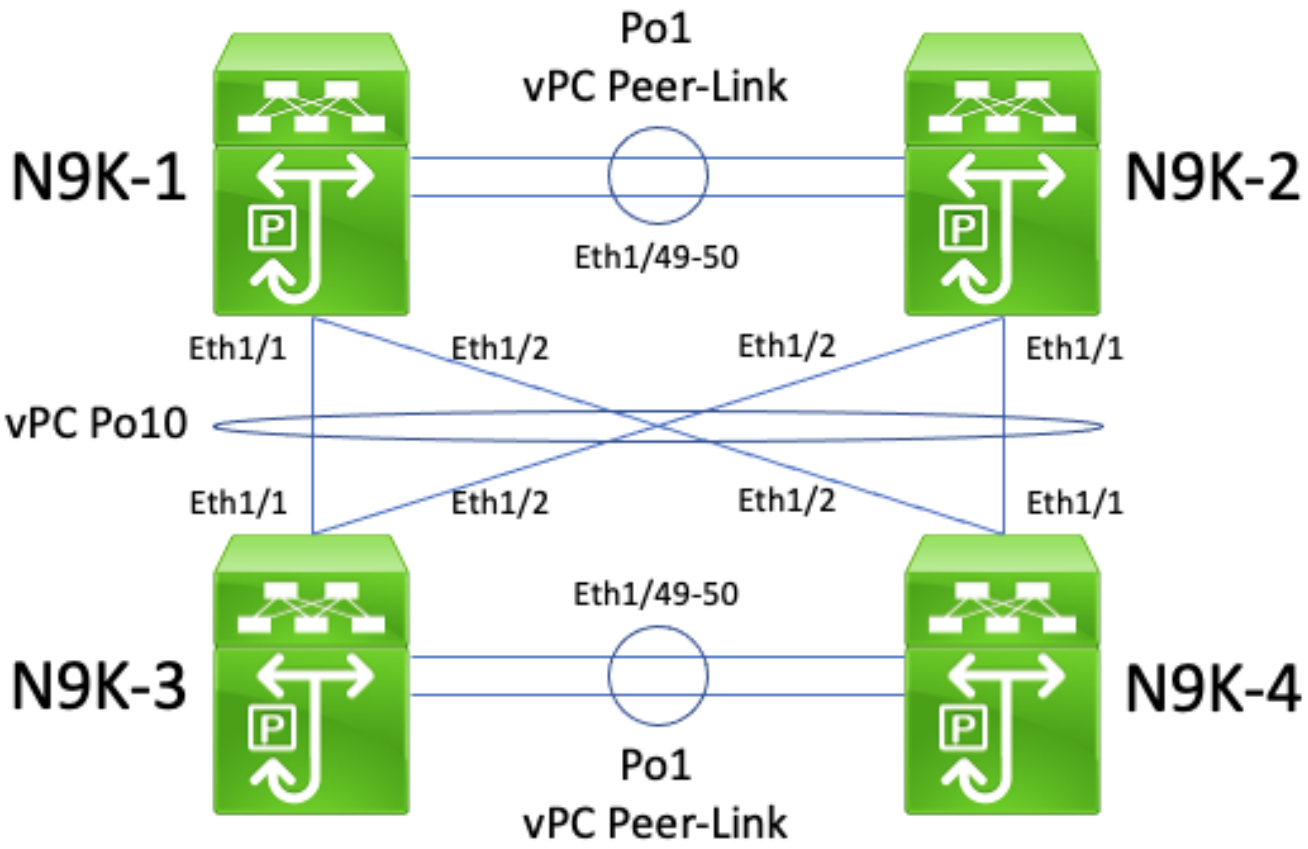
所有路由器都能够发送和接收本地链路组播路由协议数据包（通常称为“Hello”数据包），因为这些数据包已成功泛洪到 vPC VLAN。但是，请考虑以下场景：从路由器发往N9K-2的单播路由协议数据包从Ethernet1/1发往N9K-1。此数据包的目的地为N9K-2的SVI MAC地址，但进入N9K-1的Ethernet1/1接口。N9K-1发现数据包的目的地是N9K-2的SVI MAC地址，由于启用了vPC对等网关增强功能，该地址安装在N9K-1的MAC地址表中，带有“G”或“网关”标志。因此，N9K-1尝试代表N9K-2本地路由单播路由协议数据包。

但是，通过路由数据包，数据包的TTL会递减，而大多数单播路由协议数据包的TTL是1。因此，数据包的TTL会递减到0，然后丢弃N9K-1。从N9K-2的角度来看，N9K-2正在接收来自路由器的本地链路组播路由协议数据包，能够向路由器发送单播路由协议数据包，但是没有接收来自路由器的单播路由协议数据包。因此，N9K-2断开了路由协议与路由器的邻接关系，并重新启动其路由协议的本地有限状态机。同样，路由器会重新启动其用于路由协议的本地有限状态机。

我们可以通过 layer3 peer-router vPC 域配置命令启用基于 vPC 的路由/第 3 层增强功能以解决此问题。结果是允许在 vPC 对等链路上转发 TTL 为 1 的单播路由协议数据包，而无需递减数据包的 TTL。因此，可以在 vPC 或 vPC VLAN 上形成单播路由协议邻接而不会出现问题。

带 vPC 对等网关的背靠背 vPC 上的单播路由协议邻接

考虑图中显示的拓扑：



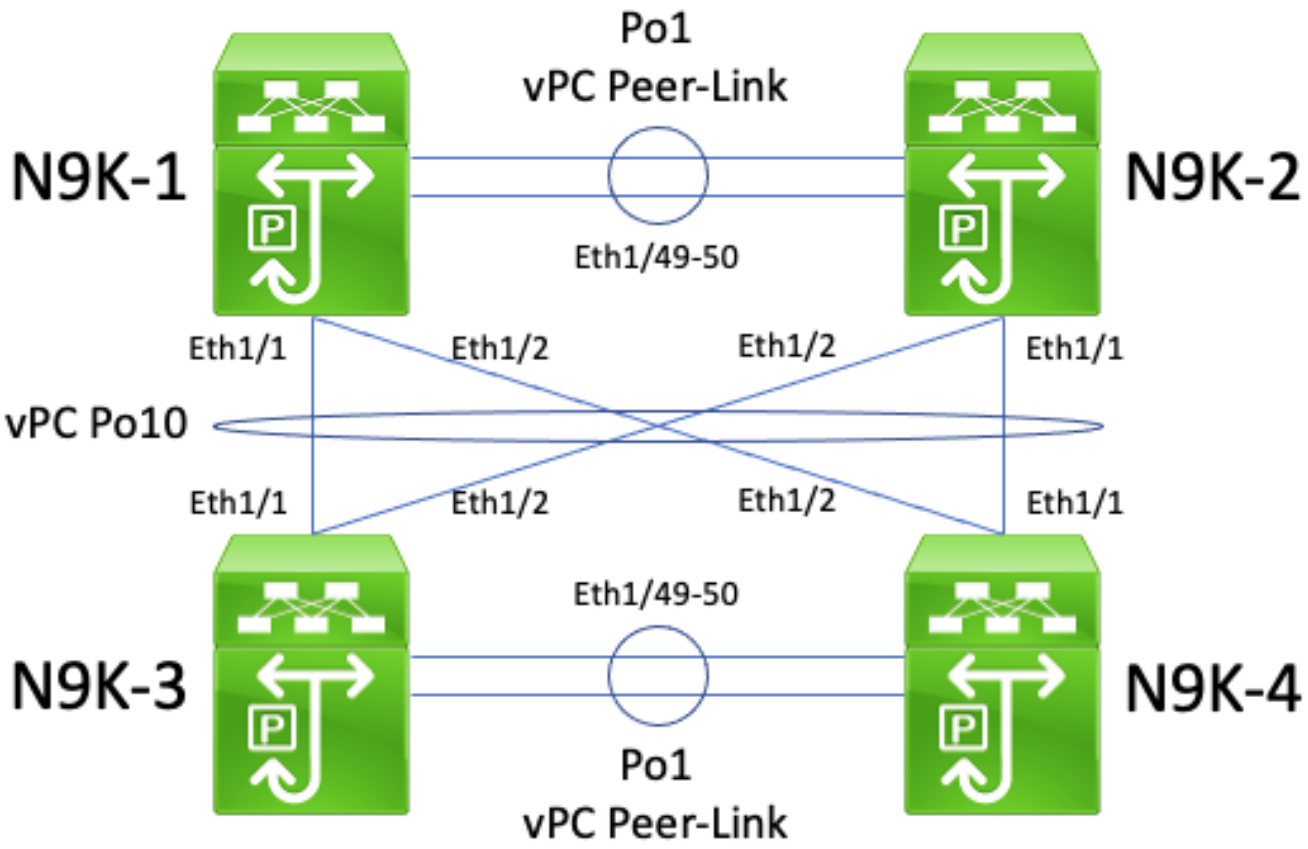
在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。Nexus 交换机 N9K-3 和 N9K-4 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。两个vPC域通过背靠背的vPC Po10相互连接。所有四台交换机均具有在单播路由协议下激活的SVI接口，并且位于同一个广播域中。

不支持已启用 vPC 对等网关增强功能的背靠背 vPC 之间的单播路由协议邻接，因为 vPC 对等网关增强功能可能会阻止两个 vPC 域之间形成单播路由协议邻接。在此拓扑中，N9K-1和N9K-3或N9K-4（或两者）之间的路由协议邻接关系可能无法正常运作。同样，N9K-2 和 N9K-3 或 N9K-4（或两者）之间的路由协议邻接可能无法按预期建立。这是因为单播路由协议数据包可能应该发往一台路由器（例如 N9K-3），但会根据源路由器的第 2 层端口通道散列决定转发到另一台路由器（例如 N9K-4）。

此问题的根本原因与本文档的“带 vPC 对等网关的 vPC 上的单播路由协议邻接”部分所述的根本原因相同。我们可以通过 layer3 peer-router vPC 域配置命令启用基于 vPC 的路由/第 3 层增强功能以解决此问题。结果是允许在 vPC 对等链路上转发 TTL 为 1 的单播路由协议数据包，而无需递减数据包的 TTL。因此，可以在背靠背 vPC 上形成单播路由协议邻接而不会出现问题。

带 vPC 对等网关的 vPC 上的 OSPF 邻接，其中前缀存在于 OSPF LSDB 中，但不在路由表中

考虑图中显示的拓扑：



在此拓扑中，Nexus 交换机 N9K-1 和 N9K-2 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。Nexus 交换机 N9K-3 和 N9K-4 是启用了 vPC 对等网关增强功能的 vPC 域内的 vPC 对等体。两个vPC域通过背靠背的vPC Po10相互连接。所有四台交换机均具有在单播路由协议下激活的SVI接口，并且位于同一个广播域中。N9K-4 是广播域的 OSPF 指定路由器 (DR)，而 N9K-3 是广播域的 OSPF 备用指定路由器 (BDR)。

在此场景中，由于单播 OSPF 数据包从两台交换机的 Ethernet1/1 接口传出，N9K-1 和 N9K-3 之间的 OSPF 邻接会转换为“FULL”状态。同样，由于单播 OSPF 数据包从两台交换机的 Ethernet1/2 传出，N9K-2 和 N9K-3 之间的 OSPF 邻接会转换为“FULL”状态。

但是，由于单播 OSPF 数据包从两台交换机的 Ethernet1/1 传出并被 N9K-2 和 N9K-4 丢弃（如本文档的“带有 vPC对等网关的背靠背 vPC 上的单播路由协议邻接”部分所述），N9K-1 和 N9K-4 之间的 OSPF 邻接卡在“EXSTART“或“EXCHANGE”状态。同样，由于单播 OSPF 数据包从两台交换机的 Ethernet1/2 传出并被 N9K-1 和 N9K-3 丢弃（如本文档的“带有 vPC对等网关的背靠背 vPC 上的单播路由协议邻接”部分所述），N9K-2和 N9K-4 之间的 OSPF 邻接卡在“EXSTART”或“EXCHANGE”状态。

因此，N9K-1 和N9K-2 在广播域的 BDR 处于“FULL”状态，但在广播域的 DR 处于“EXSTART”或“EXCHANGE”状态。广播域的 DR 和 BDR 都保留 OSPF 链路状态数据库 (LSDB) 的完整副本，但 OSPF DROTHER 路由器必须针对广播域中的 DR 处于“FULL”状态，以便安装通过 DR 或 BDR 获取的前缀。因此，N9K-1和N9K-2看起来都有从N9K-3和N9K-4获知的前缀存在于OSPF LSDB中，但是在N9K-1和N9K-2转换到N9K-4（广播域的DR）的FULL状态之前，这些前缀不会安装到单播路由表中。

我们可以通过 layer3 peer-router vPC 域配置命令启用基于 vPC 的路由/第 3 层增强功能以解决此问

题。结果是允许在 vPC 对等链路上转发 TTL 为 1 的单播路由协议数据包，而无需递减数据包的 TTL。因此，可以在背靠背 vPC 上形成单播路由协议邻接而不会出现问题。因此，N9K-1和N9K-2使用N9K-4（广播域的DR）转换到FULL状态，并将通过OSPF从N9K-3和N9K-4学习的前缀成功安装到各自的单播路由表中。

相关信息

- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.3(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.2(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 10.1(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 9.3(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 9.2(x) 版
- Cisco Nexus 9000 系列 NX-OS 接口配置指南 7.x 版
- Cisco Nexus 7000 系列 NX-OS 接口配置指南 8.x 版
- Cisco Nexus 7000 系列 NX-OS 接口配置指南 7.x 版
- 设计和配置指南：Cisco Nexus 7000系列交换机虚拟端口通道(vPC)的最佳实践
- Nexus 平台上通过虚拟端口通道进行路由所支持的拓扑

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。