Entender os aprimoramentos do Virtual Port Channel (vPC)

Contents

Introdução

Este documento descreve as melhorias comuns do Virtual Port Channel (vPC) configuradas nos switches Cisco Nexus em um domínio vPC.

Pré-requisitos

Requisitos

A Cisco recomenda que você entenda as informações básicas sobre o caso de uso, a configuração e a implementação do Virtual Port Channel (vPC). Para obter mais informações sobre esse recurso, consulte um destes documentos aplicáveis:

- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.3(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.2(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.1(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 9.3(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 9.2(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 7.x
- Guia de configuração das interfaces Cisco Nexus 7000 Series NX 8.x
- Guia de configuração das interfaces Cisco Nexus 7000 Series NX 7.x
- Guia de design e configuração: Práticas recomendadas para Virtual Port Channels (vPC) nos switches Cisco Nexus 7000 Series

Componentes Utilizados

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

Informações de Apoio

Desde a concepção do Cisco NX-OS nos switches de data center Cisco Nexus, o recurso Virtual Port Channel (vPC) recebeu vários aprimoramentos que reforçam a confiabilidade dos

dispositivos conectados por vPC durante cenários de falha e otimizam o comportamento de encaminhamento de ambos os vPC Peer Switches. Compreender a finalidade de cada aprimoramento, a mudança de comportamento que o aprimoramento proporciona e os cenários de falha que o aprimoramento resolve pode ajudar a entender por que e quando um aprimoramento deve ser configurado em um domínio de vPC, para ajudar a melhor atender às necessidades e exigências da empresa.

Hardware aplicável

O procedimento abordado neste documento é aplicável a todos os switches de data center Cisco Nexus compatíveis com vPC.

vPC Peer Switch

Esta seção descreve o aprimoramento vPC Peer Switch, que é ativado com o comando de configuração de domínio de vPC peer-switch.

Overview

Em muitos ambientes, um par de switches Nexus em um domínio de vPC são switches de agregação ou de núcleo que atuam como o limite entre os domínios de Ethernet comutados da camada 2 e os domínios roteados da camada 3. Ambos os switches são configurados com várias VLANs e são responsáveis pelo roteamento do tráfego leste-oeste entre VLANs, bem como pelo tráfego norte-sul. Nesses ambientes, os switches Nexus também atuam geralmente como pontes de origem da perspectiva do Spanning Tree Protocol.

Normalmente, um par de vPC é configurado como ponte de origem do Spanning Tree ao definir a prioridade de Spanning Tree com um valor baixo, como 0. O outro par de vPC é configurado com uma prioridade de Spanning Tree um pouco maior, como 4096, o que permite que ele assuma a função de ponte de origem no Spanning Tree, em caso de falha do par de vPC que atua como ponte de origem. Com essa configuração, o par de vPC que atua como ponte de origem origina as unidades de dados do protocolo de ponte (BPDUs) do Spanning Tree com uma ID da ponte que contém o endereço MAC do sistema.

No entanto, se o peer do vPC que atua como a bridge raiz falhar e fizer com que o outro peer do vPC assuma o controle como a bridge raiz da árvore de abrangência, o outro peer do vPC originará BPDUs da árvore de abrangência com um ID de bridge que contenha seu endereço MAC do sistema, que é diferente do endereço MAC do sistema da bridge raiz original. Dependendo de como as bridges downstream estão conectadas, o impacto dessa alteração varia e é descrito nas subseções a seguir.

Pontes não conectadas por vPC de maneira redundante

As bridges não conectadas por vPC que estão conectadas ao peer do vPC com links redundantes (de modo que um link esteja em um estado de bloqueio de uma perspectiva do Spanning Tree Protocol) que detectam a alteração na BPDU (e, portanto, a alteração na bridge raiz) observam

uma alteração na porta raiz. Outras interfaces Designated Forwarding imediatamente fazem a transição para um estado Blocking, depois passam pela máquina de estado finito do Spanning Tree Protocol (Blocking, Learning e Forwarding) com pausas entre equivalentes ao temporizador Forward Delay do Spanning Tree Protocol configurado (15 segundos por padrão).

A alteração na porta de origem e a passagem subsequente da máquina de estado finito do Spanning Tree Protocol podem causar uma quantidade significativa de interrupção na rede. O aprimoramento vPC Peer Switch foi lançado principalmente para evitar as interrupções de rede causadas por esse problema, se um dos pares de vPC ficasse off-line. Com o aprimoramento do vPC Peer Switch, a ponte conectada não-vPC ainda tem um único link redundante que está em um estado Blocking, mas faz imediatamente a transição dessa interface para um estado Forwarding se a porta raiz existente for desativada devido a uma falha do link. O mesmo processo acontece quando o peer off-line do vPC volta a ficar on-line - a interface com o menor custo para a bridge raiz assume a função de porta raiz e o link redundante passa imediatamente para um estado de bloqueio. O único impacto do plano de dados observado é a perda inevitável de pacotes em trânsito que estavam atravessando o peer do vPC quando ele ficou off-line.

Pontes conectadas por vPC

As bridges conectadas por vPC no domínio Spanning Tree detectam a alteração na BPDU (e, portanto, a alteração na bridge raiz) e eliminam dinamicamente os endereços MAC aprendidos de suas tabelas de endereços MAC locais. Esse comportamento é ineficiente e desnecessário em topologias com dispositivos conectados a vPC que não dependem do Spanning Tree Protocol para uma topologia sem loops. Os vPCs são vistos como uma única interface lógica a partir de uma perspectiva do Spanning Tree Protocol, assim como os canais de porta normais, de modo que a perda de um par de vPC é semelhante à perda de um único link dentro de um membro do canal de porta. Em ambos os cenários, o Spanning Tree não muda. Desse modo, a liberação de endereços MAC aprendidos dinamicamente nas pontes do domínio do Spanning Tree (cuja finalidade é permitir que o comportamento de inundação e aprendizagem da Ethernet aprenda novamente os endereços MAC nas interfaces recém-encaminhadas do Spanning Tree) não é necessária.

Além disso, a liberação de endereços MAC aprendidos dinamicamente pode causar interrupções. Considere um cenário em que dois hosts tenham um fluxo baseado em UDP amplamente unidirecional (como um cliente TFTP que envia dados para um servidor TFTP). Nesse fluxo, os dados fluem principalmente do cliente TFTP para o servidor TFTP. Raramente, o servidor TFTP envia um pacote de volta para o cliente TFTP. Como resultado, após uma liberação de endereços MAC aprendidos dinamicamente no domínio Spanning Tree, o MAC do servidor TFTP não é aprendido por algum tempo. Isso significa que os dados do cliente TFTP enviados para o servidor TFTP são inundados através da VLAN, pois o tráfego é unicast desconhecido. Isso pode fazer com que grandes fluxos de dados se desloquem para locais não desejados na rede e pode causar problemas de desempenho, caso esses fluxos atravessem seções da rede com excesso de assinaturas.

O aprimoramento vPC Peer Switch foi lançado para evitar que esse comportamento ineficiente e desnecessário ocorra, caso o par de vPC que atua como ponte de origem do Spanning Tree para uma ou mais VLANs seja recarregado ou desligado.

Para habilitar o aprimoramento do vPC Peer Switch, os dois pares de vPC devem ter a configuração idêntica do Spanning Tree Protocol (incluindo valores de prioridade do Spanning Tree para todas as VLANs do vPC) e ser o Root Bridge para todas as VLANs do vPC. Quando esses pré-requisitos forem atendidos, o comando de configuração de domínio de vPC peer-switch deverá ser configurado para ativar o aprimoramento vPC Peer Switch.

Observação: o aprimoramento do vPC Peer Switch só é suportado em um domínio vPC que contenha a raiz para todas as VLANs.

Quando o aprimoramento do vPC Peer Switch estiver habilitado, os dois pares de vPC começarão a originar BPDUs de Árvore Geradora idênticas com um ID de Bridge contendo o endereço MAC do sistema vPC que é compartilhado por ambos os pares de vPC. Se um peer do vPC for recarregado, o BPDU do Spanning Tree originado pelo peer do vPC restante não será alterado, de modo que outras bridges no domínio do Spanning Tree não visualizarão nenhuma alteração na bridge raiz e não reagirão de forma não ideal à alteração na rede.

Caveats

O aprimoramento vPC Peer Switch tem alguns avisos que você deve observar, antes de configurá-lo em um ambiente de produção.

Os valores de prioridade da Spanning Tree devem corresponder entre os pares de vPC

Antes de ativar o aprimoramento vPC Peer Switch, a configuração de prioridade de Spanning Tree para todas as VLANs de vPC deve ser modificada para que seja idêntica entre os dois pares de vPC.

Considere esta configuração, em que o N9K-1 está configurado para ser a ponte de origem do Spanning Tree para as VLANs 1, 10 e 20 com uma prioridade de 0. O N9K-2 é a ponte de origem secundária do Spanning Tree para as VLANs 1, 10 e 20 com uma prioridade de 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
```

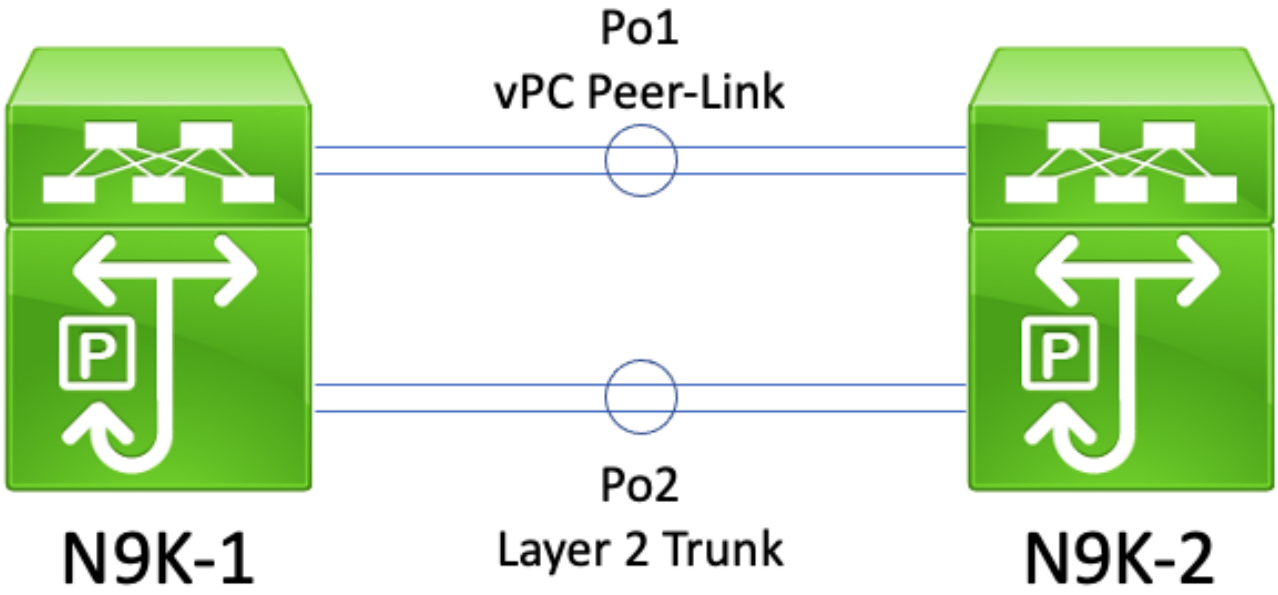
Antes de ativar o aprimoramento vPC Peer Switch, você deve modificar a configuração de prioridade de Spanning Tree das VLANs 1, 10 e 20 no N9K-2 para corresponder à configuração de prioridade de Spanning Tree das mesmas VLANs no N9K-1. Veja abaixo um exemplo dessa modificação.

<#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
```

O vPC Peer Switch afeta as VLANs não vPC

Considere esta topologia:



Nessa topologia, dois pares de vPC (N9K-1 e N9K-2) têm dois troncos da camada 2 entre eles – Po1 e Po2. Po1 é o vPC Peer-Link que transporta as VLANs de vPC, enquanto Po2 é um tronco da camada 2 que transporta todas as VLANs não vPC. Se os valores de prioridade do Spanning Tree para VLANs não vPC transportadas através de Po2 forem idênticos em N9K-1 e N9K-2, cada par de vPC originará quadros de BPDU do Spanning Tree originados do endereço MAC do sistema vPC, que é idêntico em ambos os switches. Como resultado, o N9K-1 parece receber seu próprio BPDU de Árvore de Abrangência em Po2 para cada VLAN não vPC, mesmo que o N9K-2 seja o switch que originou o BPDU de Árvore de Abrangência. De uma perspectiva do Spanning Tree, o N9K-1 coloca a Po2 em um estado de Bloqueio para todas as VLANs não-vPC.

Este é um comportamento esperado. Para evitar que esse comportamento ocorra ou contornar esse problema, os dois pares de vPC devem ser configurados com valores de prioridade de Spanning Tree diferentes em todas as VLANs não vPC. Isso permite que um peer do vPC se torne a bridge raiz da VLAN não-vPC e faça a transição do tronco da camada 2 entre os peers do vPC para um estado Designated Forwarding. Da mesma forma, o peer vPC remoto faz a transição do tronco de Camada 2 entre os peers vPC para um estado de Raiz Designada. Isso permite que o tráfego em VLANs não-vPC flua por ambos os pares vPC através do tronco de Camada 2.

Configuração

Um exemplo de como configurar o recurso vPC Peer Switch pode ser encontrado aqui.

Neste exemplo, o N9K-1 está configurado para ser a ponte de origem do Spanning Tree para as VLANs 1, 10 e 20 com uma prioridade de 0. O N9K-2 é a ponte de origem secundária do Spanning Tree para as VLANs 1, 10 e 20 com uma prioridade de 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
```

Primeiro, a configuração de prioridade de Spanning Tree do N9K-2 deve ser alterada para ser idêntica à do N9K-1. Esse requisito é estabelecido para que o recurso vPC Peer Switch funcione conforme o esperado. Se o endereço MAC do sistema de N9K-2 for inferior ao endereço MAC do sistema de N9K-1, então o N9K-2 usurpará a função de bridge raiz para o domínio Spanning Tree, o que faz com que outras bridges no domínio Spanning Tree liberem suas tabelas de endereços MAC locais para todas as VLANs afetadas. Veja abaixo um exemplo desse fenômeno.

<#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
```

Em seguida, podemos ativar o recurso vPC Peer Switch por meio do comando de configuração de domínio de vPC peer-switch. Isso altera o ID da ponte dentro das BPDUs da árvore de abrangência originadas por ambos os pares de vPC, o que faz com que outras pontes no domínio da árvore de abrangência liberem suas tabelas de endereços MAC locais para todas as VLANs afetadas.

```
<#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#
```

Você pode verificar se o recurso vPC Peer Switch está funcionando conforme esperado, validando ambos os pares de vPC que afirmam ser a ponte de origem das VLANs de vPC usando o comando show spanning-tree summary. Essa saída também deve indicar que o recurso vPC Peer Switch está ativado e operacional.

```
<#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
```

Use o comando show spanning-tree vlan {x} para exibir informações mais detalhadas sobre uma VLAN específica. O switch que tem a função de vPC Principal ou Principal Operacional tem todas as suas interfaces em um estado de Encaminhamento Designado. O switch com a função vPC secundária ou secundária operacional tem todas as suas interfaces em um estado de encaminhamento designado, exceto o vPC Peer-Link, que está em um estado de encaminhamento de raiz. Observe que o endereço MAC do sistema de vPC exibido na saída de show vpc role é idêntico à ID da ponte de origem e à ID da ponte de cada par de vPC.

<#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
```

Por fim, podemos usar o [utilitário de captura de pacotes do plano de controle do EthAnalyzer] em um dos pares de vPC para confirmar se os dois pares de vPC estão originando as BPDUs do Spanning Tree com uma ID da ponte e uma ID da ponte de origem, as quais contêm o endereço MAC do sistema de vPC compartilhado entre os dois pares de vPC.

<#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
```

Impacto

O impacto da ativação do aprimoramento do vPC Peer Switch varia dependendo se outras pontes no domínio Spanning Tree estão conectadas a ambos os pares vPC através de um vPC ou se estão conectadas de forma redundante a ambos os pares vPC sem um vPC.

Pontes não conectadas por vPC de maneira redundante

Se uma ponte não conectada por vPC com links redundantes para os dois pares de vPC (de modo que um link esteja em um estado de bloqueio da perspectiva do Spanning Tree Protocol) detectar uma alteração na ponte de origem do Spanning Tree anunciada nas BPDUs do Spanning Tree, a porta de origem da ponte poderá mudar entre as duas interfaces redundantes. Por sua vez, isso pode fazer com que outras interfaces de encaminhamento designadas passem imediatamente para um estado de bloqueio e atravessem a máquina de estado finito do Spanning Tree Protocol (bloqueio, aprendizado e encaminhamento) com pausas intermediárias

equivalentes ao temporizador de atraso de encaminhamento do Spanning Tree Protocol configurado (15 segundos por padrão). A alteração na porta de origem e a passagem subsequente da máquina de estado finito do Spanning Tree Protocol podem causar uma quantidade significativa de interrupção na rede.

Vale mencionar que esse impacto ocorre sempre que o peer do vPC que atualmente é a bridge raiz para o domínio Spanning Tree fica off-line (como no caso de falta de energia, falha de hardware ou uma recarga). Esse comportamento não é específico para o aprimoramento vPC Peer Switch – ativar o aprimoramento vPC Peer Switch simplesmente causa um comportamento semelhante ao de um par de vPC que fica off-line da perspectiva do Spanning Tree.

Pontes conectadas por vPC

Se uma bridge conectada a vPC detectar uma alteração na bridge raiz de Spanning Tree anunciada em BPDUs de Spanning Tree, a bridge libera endereços MAC aprendidos dinamicamente de sua tabela de endereços MAC. Ao configurar o recurso vPC Peer Switch, você pode observar esse comportamento nos dois cenários a seguir:

1. Quando os valores de prioridade de Spanning Tree são configurados para corresponder entre os dois pares de vPC, a ponte de origem do Spanning Tree pode mudar de um par de vPC para outro, se o par de vPC que não era anteriormente a ponte de origem tiver um endereço MAC do sistema inferior ao do par de vPC que era anteriormente a ponte de origem. Um exemplo desse cenário é mostrado na seção Configuração do vPC Peer Switch deste documento.
2. Quando o recurso vPC Peer Switch é habilitado através do comando de configuração de domínio peer-switch vPC, ambos os pares de vPC começam a operar como bridges raiz do domínio Spanning Tree. Ambos os pares de vPC começam a originar BPDUs de Árvore de Abrangência idênticas que se autodeclaram como a bridge raiz do domínio de Árvore de Abrangência.

Na maioria dos cenários e das topologias, nenhum impacto no plano de dados é observado como resultado de qualquer um desses dois cenários. No entanto, por um curto período de tempo, o tráfego do plano de dados é inundado dentro de uma VLAN devido à inundação unicast desconhecida, pois o endereço MAC destino dos quadros não é aprendido em nenhuma porta de switch como resultado direto da descarga de endereços MAC aprendidos dinamicamente. Em algumas topologias, isso pode causar breves períodos de problemas de desempenho ou perda de pacotes, se o tráfego do plano de dados for inundado para dispositivos de rede com excesso de assinaturas na VLAN. Isso também pode causar problemas com fluxos de tráfego unidirecional com uso intensivo de largura de banda ou hosts silenciosos (hosts que recebem principalmente pacotes e raramente enviam pacotes), pois esse tráfego é inundado dentro da VLAN por um longo período de tempo em vez de ser comutado diretamente para o host de destino como de costume.

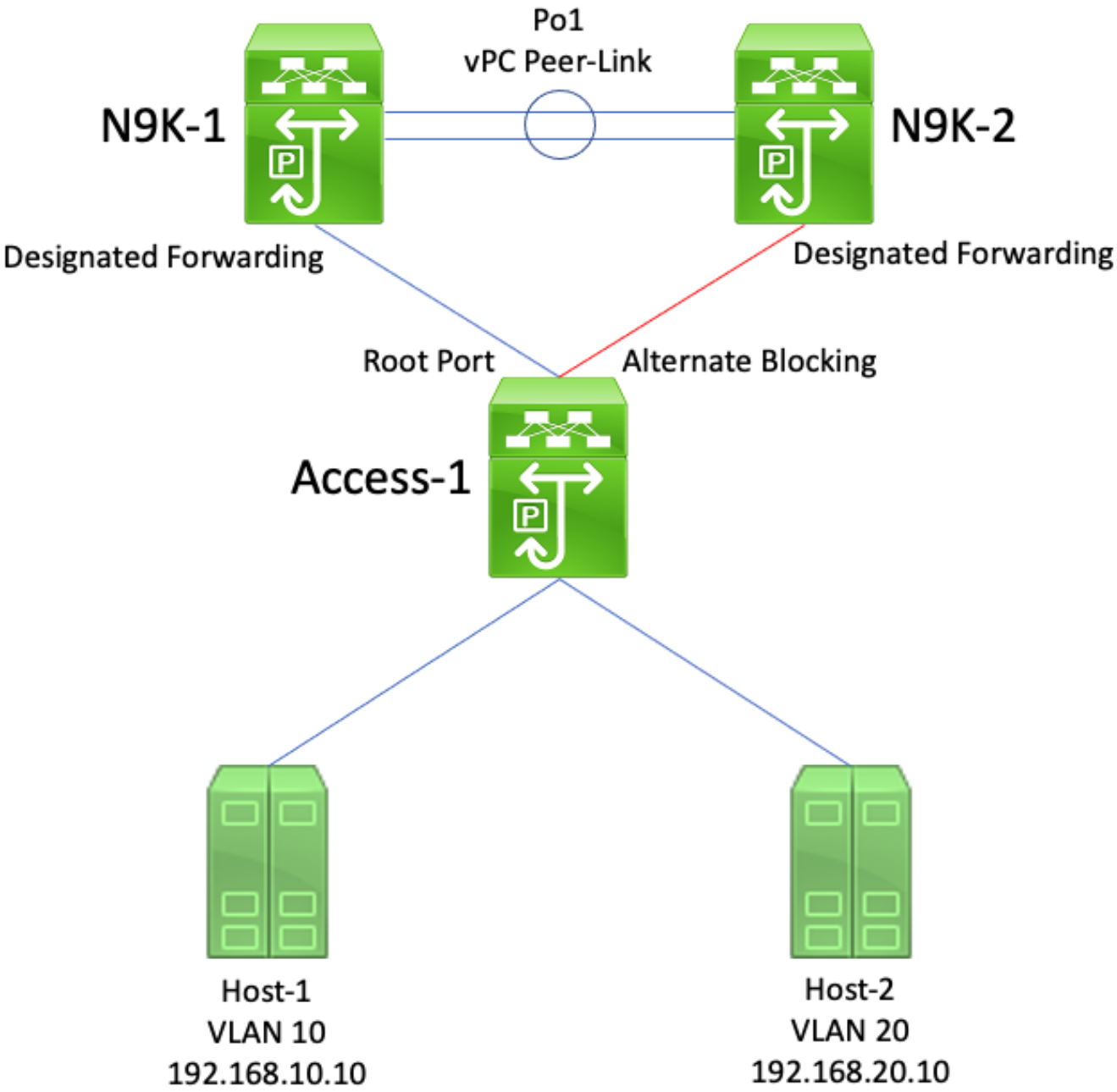
Vale mencionar que esse impacto está relacionado à liberação de endereços MAC aprendidos dinamicamente da tabela de endereços MAC das bridges dentro da VLAN afetada. Esse comportamento não é específico para o aprimoramento vPC Peer Switch ou uma alteração na ponte de origem – ele também pode ser causado por uma notificação de alteração de topologia

gerada devido a uma porta non-edge que surge na VLAN.

Cenários de falha de exemplo

Pontes não conectadas por vPC de maneira redundante que reiniciam a máquina de estado finito

Considere esta topologia:



Nesta topologia, o N9K-1 e o N9K-2 são pares de vPC em um domínio de vPC. O N9K-1 é configurado com um valor de prioridade de Spanning Tree de 0 para todas as VLANs, tornando o N9K-1 a ponte de origem para todas as VLANs. O N9K-2 é configurado com um valor de prioridade de Spanning Tree de 4096 para todas as VLANs, tornando o N9K-2 a ponte de origem secundária para todas as VLANs. O Access-1 é um switch conectado de forma redundante ao N9K-1 e ao N9K-2 por meio das portas do switch da camada 2. Essas portas do switch não são

agrupadas em um port-channel, portanto, o Spanning Tree Protocol coloca o link conectado ao N9K-1 em um estado de origem designada e o link conectado ao N9K-2 em um estado de bloqueio alternativo.

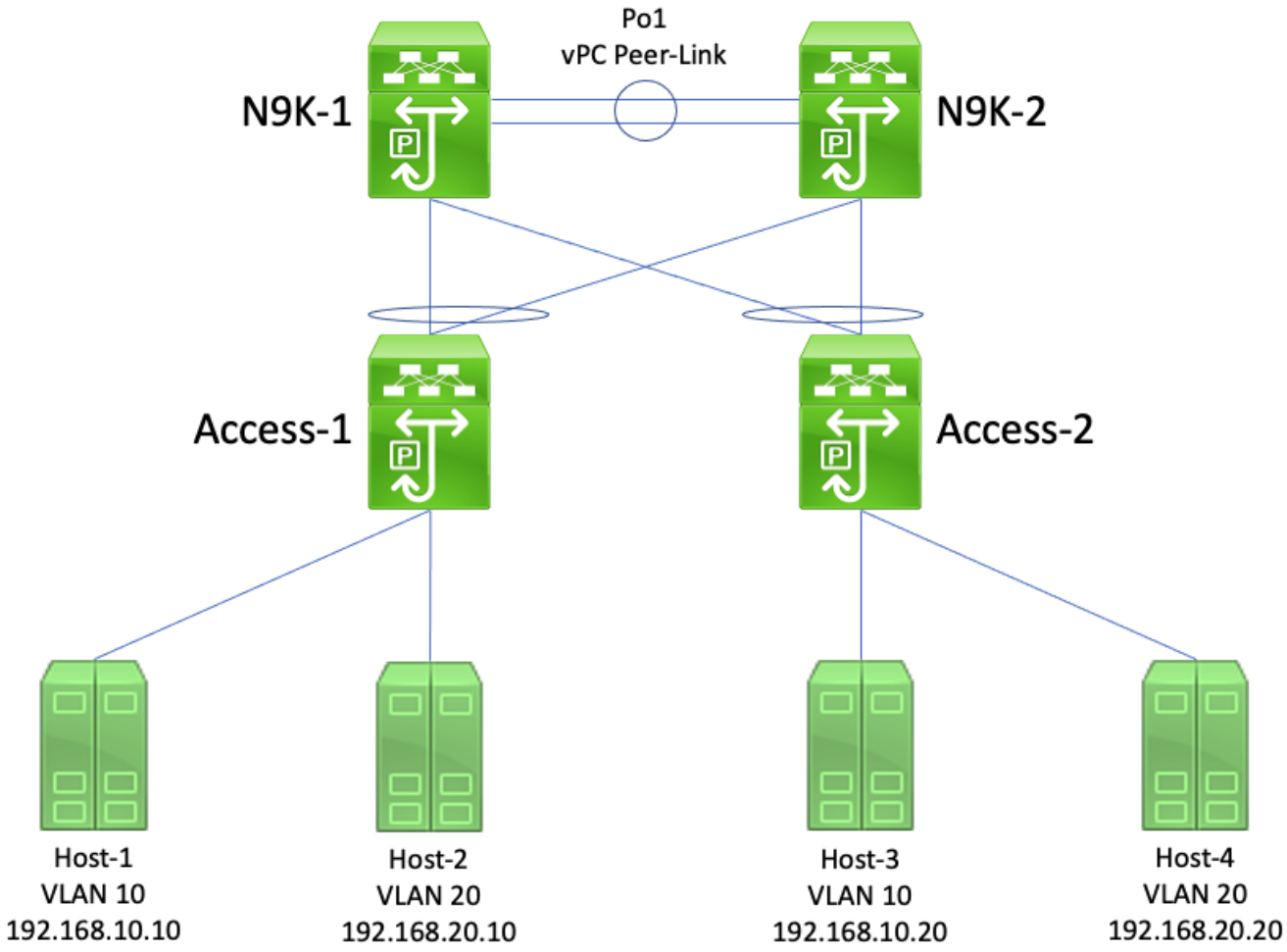
Considere um cenário de falha em que o N9K-1 fica off-line devido a uma falha de hardware, uma falta de energia ou um recarregamento do switch. O N9K-2 se declara como a bridge raiz para todas as VLANs anunciando BPDUs de árvore de abrangência usando seu endereço MAC do sistema como o ID da bridge. O Access-1 vê uma alteração no ID da bridge raiz. Além disso, sua porta de raiz designada faz a transição para um estado down/down, o que significa que a nova porta de raiz designada é o link que estava em um estado de bloqueio alternativo voltado para N9K-2.

Essa alteração nas portas da raiz designada faz com que todas as portas Spanning Tree não borda passem pela máquina de estado finito do Spanning Tree Protocol (Blocking, Learning e Forwarding) com pausas entre as equivalentes ao temporizador de retardo de encaminhamento do Spanning Tree Protocol configurado (15 segundos por padrão). Esse processo pode ser extremamente prejudicial para a rede.

No mesmo cenário de falha com o aprimoramento do vPC Peer Switch habilitado, N9K-1 e N9K-2 transmitem BPDUs de Spanning Tree idênticas usando o endereço MAC do sistema vPC compartilhado como ID da bridge. Se N9k-1 falhar, N9K-2 continuará transmitindo esse mesmo BPDU de Spanning Tree. Como resultado, o Access-1 faz imediatamente a transição do link de bloqueio alternativo em direção a N9K-2 para um estado de raiz designada e começa a encaminhar o tráfego pelo link. Além disso, como a ID da ponte de origem do Spanning Tree não muda, isso impede que as portas non-edge passem pela máquina de estado finito do protocolo Spanning Tree, o que reduz a quantidade de interrupção observada na rede.

Pontes conectadas por vPC que liberam endereços MAC aprendidos dinamicamente

Considere esta topologia:



Nesta topologia, o N9K-1 e o N9K-2 são pares de vPC em um domínio de vPC que executam o roteamento entre VLANs entre a VLAN 10 e a VLAN 20. O N9K-1 é configurado com um valor de prioridade de Spanning Tree de 0 para a VLAN 10 e a VLAN 20, tornando o N9K-1 a ponte de origem para ambas as VLANs. O N9K-2 é configurado com um valor de prioridade de Spanning Tree de 4096 para a VLAN 10 e a VLAN 20, tornando o N9K-2 a ponte de origem secundária para ambas as VLANs. Host-1, Host-2, Host-3 e Host-4 estão se comunicando continuamente.

Considere um cenário de falha em que o N9K-1 fica off-line devido a uma falha de hardware, uma falta de energia ou um recarregamento do switch. O N9K-2 se declara como a bridge raiz para a VLAN 10 e a VLAN 20 anunciando BPDUs de Spanning Tree usando seu endereço MAC de sistema como o ID da bridge. Access-1 e Access-2 veem uma alteração no ID da bridge raiz e, embora a spanning tree permaneça a mesma (o que significa que o vPC voltado para N9K-1 e N9K-2 permanece como uma porta de raiz designada), tanto Access-1 quanto Access-2 liberam seu endereço MAC de todos os endereços MAC aprendidos dinamicamente na VLAN 10 e na VLAN 20.

Na maioria dos ambientes, a liberação de endereços MAC aprendidos dinamicamente causa um impacto mínimo. Nenhum pacote é perdido (exceto os que se perderam porque foram transmitidos para N9K-1 durante a falha), mas o tráfego é temporariamente inundado em cada domínio de transmissão como tráfego unicast desconhecido, enquanto todos os switches no domínio de transmissão reaprendem os endereços MAC dinâmicos.

No mesmo cenário de falha com o aprimoramento vPC Peer Switch ativado, o N9K-1 e o N9K-2

transmitiriam as BPDUs do Spanning Tree idênticas usando o endereço MAC do sistema de vPC compartilhado como a ID da ponte. Se N9k-1 falhar, N9K-2 continuará transmitindo esse mesmo BPDU de Spanning Tree. Como resultado, Access-1 e Access-2 não estão cientes de que qualquer alteração na topologia Spanning Tree ocorreu - da perspectiva deles, as BPDUs Spanning Tree da bridge raiz são idênticas, portanto não há necessidade de descarregar endereços MAC aprendidos dinamicamente das VLANs relevantes. Isso evita a inundação de tráfego unicast desconhecido em cada domínio de transmissão nesse cenário de falha.

vPC Peer Gateway

Esta seção descreve o aprimoramento vPC Peer Gateway, que é ativado com o comando de configuração de domínio de vPC peer-gateway.

Overview

Os switches Nexus configurados em um domínio de vPC executam o encaminhamento dual-active do First Hop Redundancy Protocol (FHRP) por padrão. Isso significa que se um par vPC receber um pacote com um endereço MAC de destino que pertença a um grupo de Protocolo de Roteamento de Hot Standby (HSRP) ou Protocolo de Redundância de Roteador Virtual (VRRP) configurado no switch, o switch roteará o pacote de acordo com sua tabela de roteamento local, independentemente de seu estado de plano de controle HSRP ou VRRP. Em outras palavras, o comportamento esperado de um par de vPC em um estado de standby do HSRP ou de backup do VRRP é encaminhar os pacotes destinados ao endereço MAC virtual do HSRP ou do VRRP.

Quando um peer vPC roteia um pacote destinado a um endereço MAC virtual de FHRP, ele regrava o pacote com um novo endereço MAC origem e destino. O endereço MAC origem é o endereço MAC da interface virtual comutada (SVI) do peer do vPC dentro da VLAN para a qual o pacote é roteado. O endereço MAC destino é o endereço MAC associado ao endereço IP do próximo salto para o endereço IP destino do pacote de acordo com a tabela de roteamento local do peer do vPC. Em cenários de roteamento entre VLANs, o endereço MAC destino do pacote após a reescrita do pacote é o endereço MAC do host ao qual o pacote está destinado.

Alguns hosts não entendem o comportamento de encaminhamento padrão como recurso de otimização. Com esse comportamento, o host não executa uma tabela de roteamento e/ou pesquisa em cache ARP ao responder a um pacote de entrada. Em vez disso, o host inverte os endereços MAC de origem e de destino do pacote de entrada para o pacote de resposta. Em outras palavras, o endereço MAC de origem do pacote de entrada se torna o endereço MAC de destino do pacote de resposta, e o endereço MAC de destino do pacote de entrada se torna o endereço MAC de origem do pacote de resposta. Esse comportamento é diferente de um host que entende o comportamento de encaminhamento padrão, o qual executaria uma tabela de roteamento local e/ou pesquisa de cache ARP e definiria o endereço MAC de destino do pacote de resposta como o endereço MAC virtual do FHRP.

Esse comportamento de host não padrão poderá violar a regra de prevenção de loop de vPC se o pacote de resposta gerado pelo host for endereçado a um par de vPC, mas sair do vPC para o outro par de vPC. O outro peer do vPC recebe o pacote destinado a um endereço MAC de

propriedade de seu peer do vPC e encaminha o pacote para fora do Peer-Link do vPC para o peer do vPC que possui o endereço MAC presente no campo de endereço MAC de destino do pacote. O peer do vPC que possui o endereço MAC tenta rotear o pacote localmente. Se o pacote precisar sair de um vPC, o peer do vPC descartará esse pacote por violar a regra de Prevenção de Loop do vPC. Como resultado, você pode observar problemas de conectividade ou perda de pacotes para alguns fluxos provenientes de ou destinados a um host que apresenta esse comportamento fora do padrão.

O aprimoramento vPC Peer Gateway foi lançado para eliminar a perda de pacotes causada por hosts que apresentam esse comportamento fora do padrão. Isso é feito permitindo que um par de vPC encaminhe localmente os pacotes destinados ao endereço MAC do outro par de vPC, de modo que os pacotes destinados ao par de vPC remoto não precisem sair do vPC Peer-Link para que sejam encaminhados. Em outras palavras, o aprimoramento vPC Peer Gateway permite que um par de vPC encaminhe pacotes "em nome" do par de vPC remoto. O aprimoramento vPC Peer Gateway pode ser ativado com o comando de configuração de domínio de vPC peer-gateway.

Caveats

Oscilação das adjacências do protocolo de roteamento unicast por vPCs ou VLANs de vPC

Se as adjacências do protocolo de roteamento unicast dinâmico forem formadas entre dois pares de vPC e um roteador conectado por vPC ou um roteador conectado por meio de uma porta órfã de vPC, as adjacências do protocolo de roteamento poderão começar a flutuar continuamente, depois de ativar o aprimoramento vPC Peer Gateway, caso o aprimoramento Routing/Layer 3 over vPC não seja configurado imediatamente depois. Esses cenários de falha são descritos detalhadamente nas seções Cenário de falha de exemplo para adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway e Adjacências do protocolo de roteamento unicast por um VLAN com vPC Peer Gateway deste documento.

Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer3 peer-router imediatamente após ativar o aprimoramento vPC Peer Gateway com o comando de configuração de domínio de vPC peer-gateway.

Desativação automática dos redirecionamentos de ICMP e ICMPv6

Quando o aprimoramento do vPC Peer Gateway está habilitado, a geração de pacotes de redirecionamento de ICMP e ICMPv6 é desabilitada automaticamente em todas as SVIs de VLAN vPC (ou seja, qualquer SVI associado a uma VLAN que esteja em tronco no vPC Peer-Link). O switch faz isso configurando no ip redirects e no ipv6 redirects em todas as SVIs de VLAN de vPC. Isso evita que um switch gere pacotes de redirecionamento ICMP em resposta a pacotes que entram no switch, mas têm um endereço MAC e um endereço IP de destino do par de vPC do switch.

Se os pacotes de redirecionamento ICMP ou ICMPv6 forem necessários em seu ambiente dentro de uma VLAN específica, você precisará excluir essa VLAN para tirar proveito do aprimoramento

do vPC Peer Gateway usando o comando de configuração de domínio peer-gateway exclude-vlan <vlan-id> vPC.

Observação: o comando peer-gateway exclude-vlan <vlan-id> de configuração de domínio vPC não é suportado nos switches Nexus 9000 Series.

Configuração

Um exemplo de como configurar o recurso vPC Peer Gateway pode ser encontrado aqui.

Neste exemplo, o N9K-1 e o N9K-2 são pares de vPC em um domínio de vPC. Ambos os pares de vPC têm um grupo HSRP configurado para a VLAN 10. O N9K-1 é o roteador HSRP ativo com uma prioridade de 150, enquanto o N9K-2 é o roteador HSRP standby com a prioridade padrão de 100.

<#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)
```

A SVI da VLAN 10 do N9K-1 tem o endereço MAC 00ee.ab67.db47 e a SVI da VLAN 10 do N9K-2 tem o endereço MAC 00ee.abd8.747f. O endereço MAC virtual do HSRP para a VLAN 10 é 0000.0c07.ac0a. Nesse estado, o endereço MAC da SVI da VLAN 10 de cada switch e o endereço MAC virtual do HSRP estão presentes na tabela de endereços MAC de cada switch. O endereço MAC SVI da VLAN 10 de cada switch e o endereço MAC virtual do HSRP têm o flag Gateway (G) presente, que indica que o switch roteia localmente pacotes destinados a esse endereço MAC.

Observe que a tabela de endereços MAC do N9K-1 não apresenta o flag Gateway para o endereço MAC da SVI da VLAN 10 do N9K-2. Da mesma forma, a tabela de endereços MAC do N9K-2 não apresenta o flag Gateway para o endereço MAC da SVI da VLAN 10 do N9K-1.

<#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)
```

Podemos ativar o aprimoramento vPC Peer Gateway por meio do comando de configuração de domínio de vPC peer-gateway. Isso permite que o switch roteie localmente os pacotes recebidos com um endereço MAC de destino pertencente ao endereço MAC do peer do vPC aprendido no Peer-Link do vPC. Isso é feito configurando o flag Gateway no endereço MAC do par de vPC na tabela de endereços MAC do switch.

<#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#
```

Você pode verificar se o aprimoramento vPC Peer Gateway está funcionando conforme o

esperado, validando se o flag Gateway está presente na tabela de endereços MAC para o MAC do par de vPC.

<#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)
```

Impacto

O impacto da ativação do aprimoramento do vPC Peer Gateway pode variar dependendo da topologia ao redor e do comportamento dos hosts conectados, conforme descrito nas subseções a seguir. Se nenhuma das subseções a seguir se aplicar ao seu ambiente, a habilitação do aprimoramento do vPC Peer Gateway não causará interrupções e não terá impacto no seu ambiente.

Oscilação das adjacências do protocolo de roteamento unicast por vPCs ou VLANs de vPC

Se as adjacências do protocolo de roteamento unicast dinâmico forem formadas entre dois pares de vPC e um roteador conectado por vPC ou um roteador conectado por meio de uma porta órfã de vPC, as adjacências do protocolo de roteamento poderão começar a flutuar continuamente, depois de ativar o aprimoramento vPC Peer Gateway, caso o aprimoramento Routing/Layer 3 over vPC não seja configurado imediatamente depois. Esses cenários de falha são descritos detalhadamente nas seções Cenário de falha de exemplo para adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway e Adjacências do protocolo de roteamento unicast por um VLAN com vPC Peer Gateway deste documento.

Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer3 peer-router imediatamente após ativar o aprimoramento vPC Peer Gateway com o comando de configuração de domínio de vPC peer-gateway.

Desativação automática dos redirecionamentos de ICMP e ICMPv6

Quando o aprimoramento do vPC Peer Gateway está habilitado, a geração de pacotes de redirecionamento de ICMP e ICMPv6 é desabilitada automaticamente em todas as SVIs de VLAN vPC (ou seja, qualquer SVI associado a uma VLAN que esteja em tronco no vPC Peer-Link). O switch faz isso configurando no ip redirects e no ipv6 redirects em todas as SVIs de VLAN de vPC. Isso evita que um switch gere pacotes de redirecionamento ICMP em resposta a pacotes que entram no switch, mas têm um endereço MAC e um endereço IP de destino do par de vPC do switch.

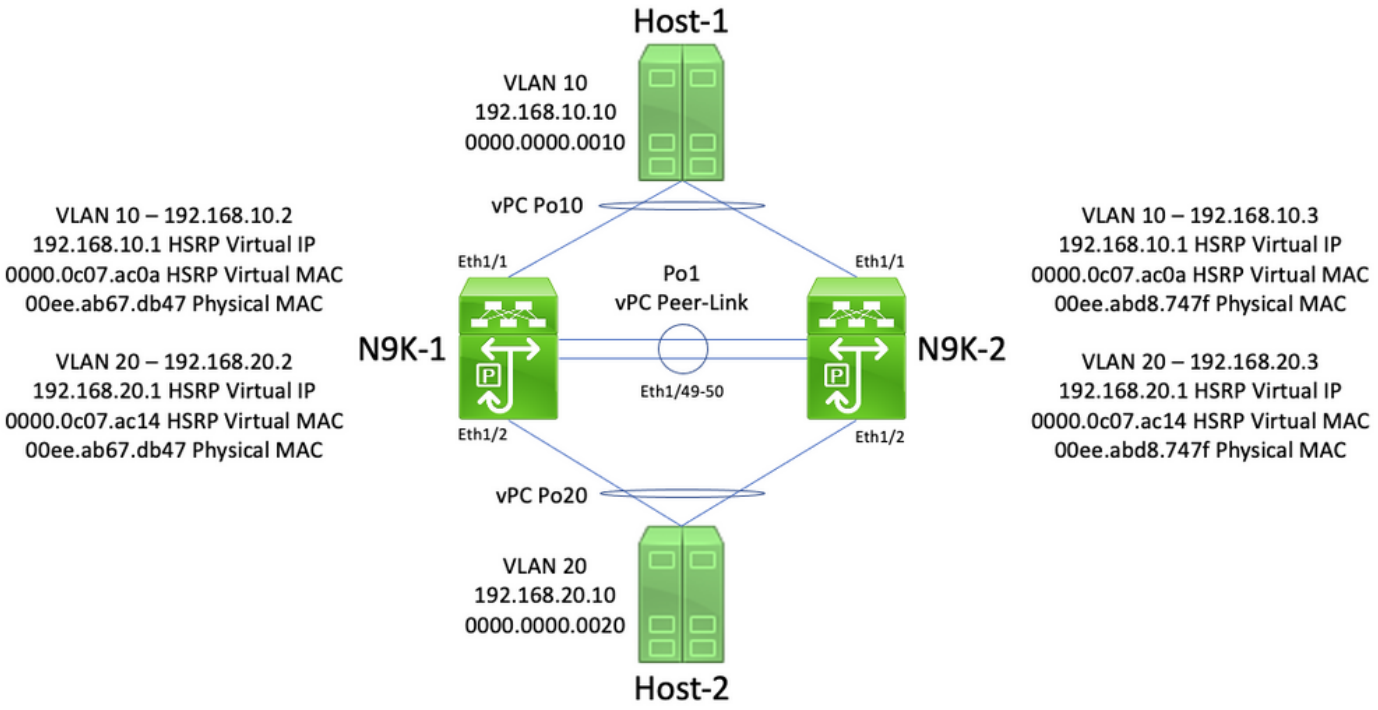
Se os pacotes de redirecionamento ICMP ou ICMPv6 forem necessários em seu ambiente dentro de uma VLAN específica, você precisará excluir essa VLAN para tirar proveito do aprimoramento do vPC Peer Gateway usando o comando de configuração de domínio peer-gateway exclude-vlan <vlan-id> vPC.

Observação: o comando peer-gateway exclude-vlan <vlan-id> de configuração de domínio vPC não é suportado nos switches Nexus 9000 Series.

Cenários de falha de exemplo

Hosts conectados por vPC com comportamento de encaminhamento não padrão

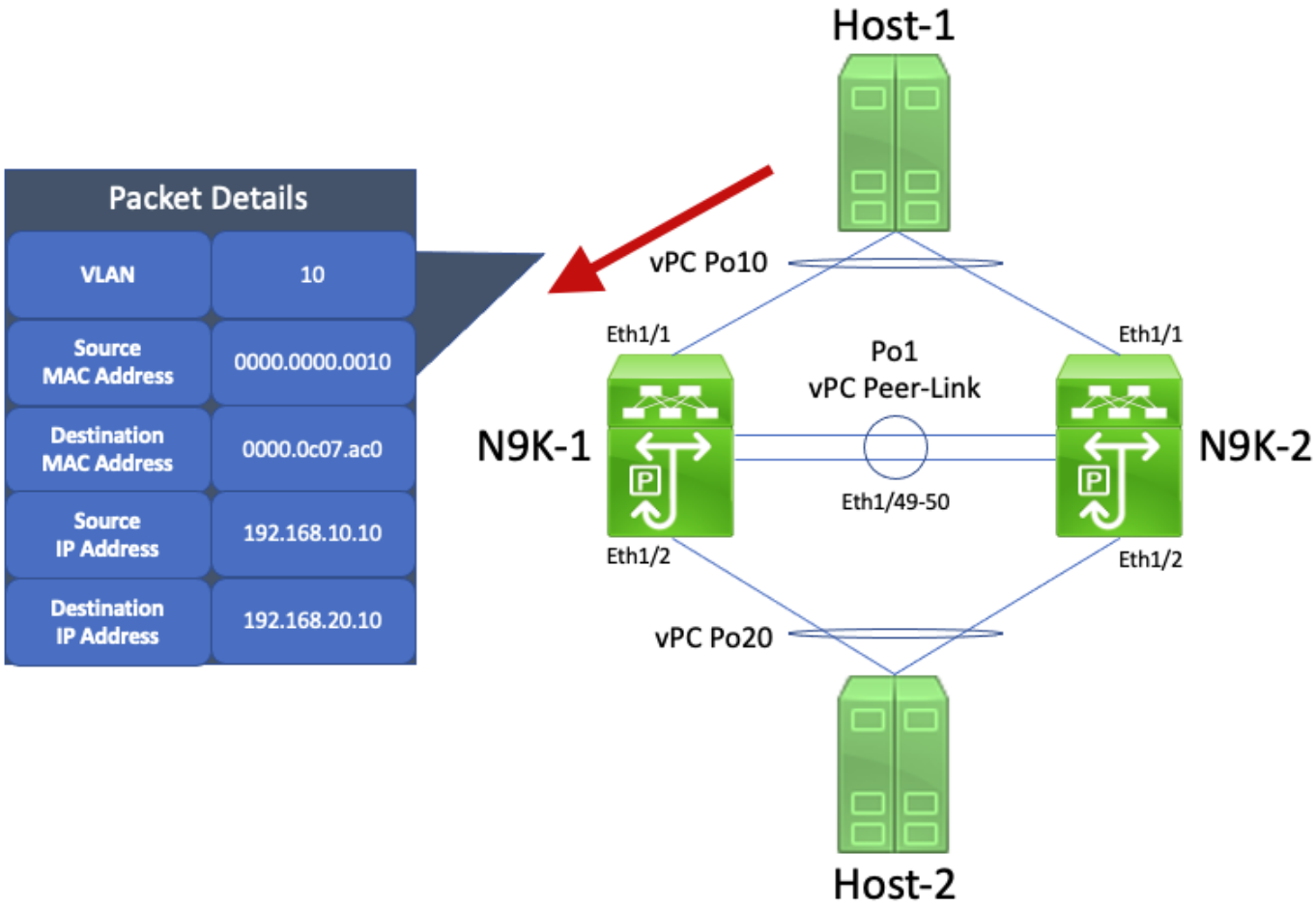
Considere esta topologia:



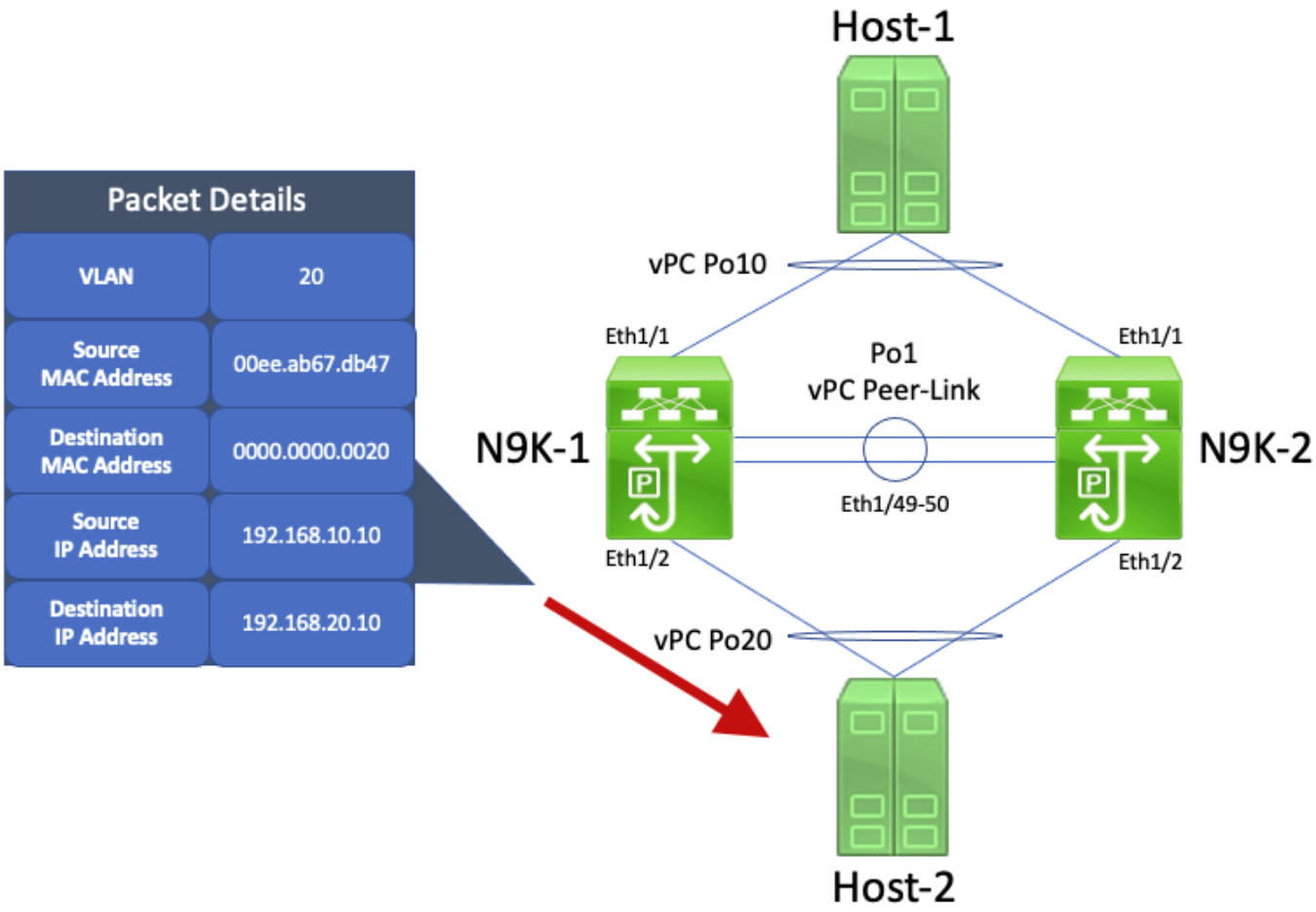
Nesta topologia, o N9K-1 e o N9K-2 são pares de vPC em um domínio de vPC que executam o roteamento entre VLANs entre a VLAN 10 e a VLAN 20. A interface Po1 é o vPC Peer-Link. Um host chamado Host-1 está conectado por meio do Po10 de vPC ao N9K-1 e ao N9K-2 na VLAN 10. O Host-1 possui o endereço IP 192.168.10.10 com o endereço MAC 0000.0000.0010. Um host chamado Host-2 está conectado por meio do Po20 de vPC ao N9K-1 e ao N9K-2 na VLAN 20. O Host-2 possui o endereço IP 192.168.20.10 com o endereço MAC 0000.0000.0020.

O N9K-1 e o N9K-2 têm SVIs na VLAN 10 e na VLAN 20 com o HSRP ativado em cada SVI. A interface da VLAN 10 do N9K-1 tem o endereço IP 192.168.10.2 e a interface da VLAN 20 do N9K-1 tem o endereço IP 192.168.20.2. As duas SVIs do N9K-1 têm o endereço MAC físico 00ee.ab67.db47. A interface da VLAN 10 do N9K-2 tem o endereço IP 1921.68.10.3 e a interface da VLAN 20 do N9K-2 tem o endereço IP 192.168.20.3. As duas SVIs do N9K-2 têm o endereço MAC físico 00ee.abd8.747f. O endereço IP virtual do HSRP para a VLAN 10 é 192.168.10.1 e o endereço MAC virtual do HSRP é 0000.0c07.ac0a. O endereço IP virtual do HSRP para a VLAN 20 é 192.168.20.1 e o endereço MAC virtual do HSRP é 0000.0c07.ac14.

Considere um cenário em que o Host-1 envia um pacote de solicitação de eco ICMP para o Host-2. Depois que o Host-1 resolve o ARP para o gateway padrão (o endereço IP virtual do HSRP), o Host-1 entende o comportamento de encaminhamento padrão e gera um pacote de solicitação de eco ICMP com o endereço IP de origem 192.168.10.10, o endereço IP de destino 192.168.20.10, o endereço MAC de origem 0000.0000.0010 e o endereço MAC de destino 0000.0c07.ac0a. Esse pacote sai para o N9K-1. Veja abaixo um exemplo visual disso.

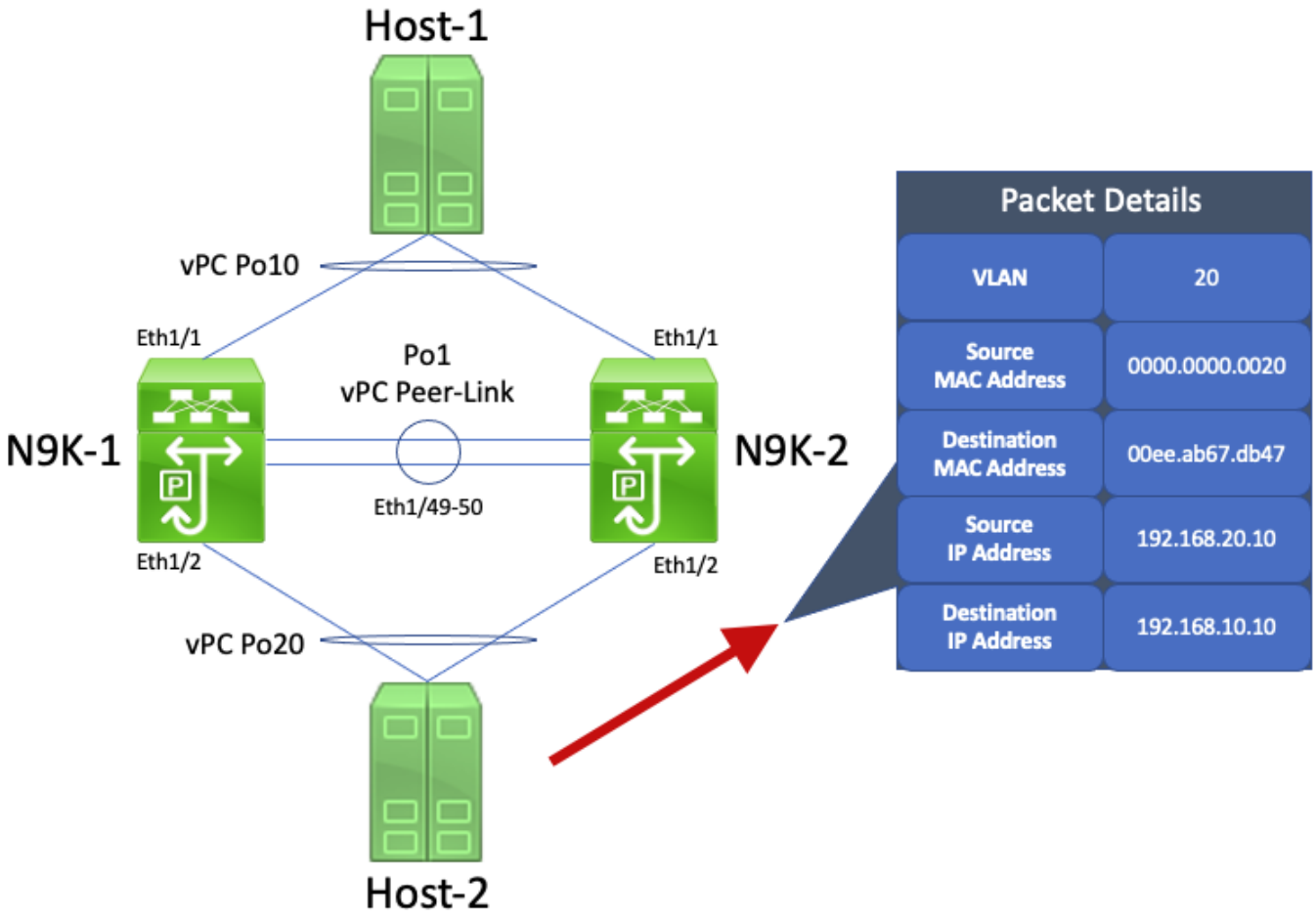


O N9K-1 recebe esse pacote. Como esse pacote é destinado ao endereço MAC virtual do HSRP, o N9K-1 pode encaminhar esse pacote de acordo com a tabela de roteamento local, independentemente do estado do plano de controle do HSRP. Esse pacote é roteado da VLAN 10 para a VLAN 20. Como parte do roteamento do pacote, o N9K-1 executa a regravação do pacote, endereçando novamente os campos de endereço MAC origem e destino do pacote. O novo endereço MAC origem do pacote é o endereço MAC físico associado à VLAN 20 SVI (00ee.ab67.db47) de N9K-1 e o novo endereço MAC destino é o endereço MAC associado ao Host-2 (0000.0000.0020). Veja abaixo um exemplo visual disso.

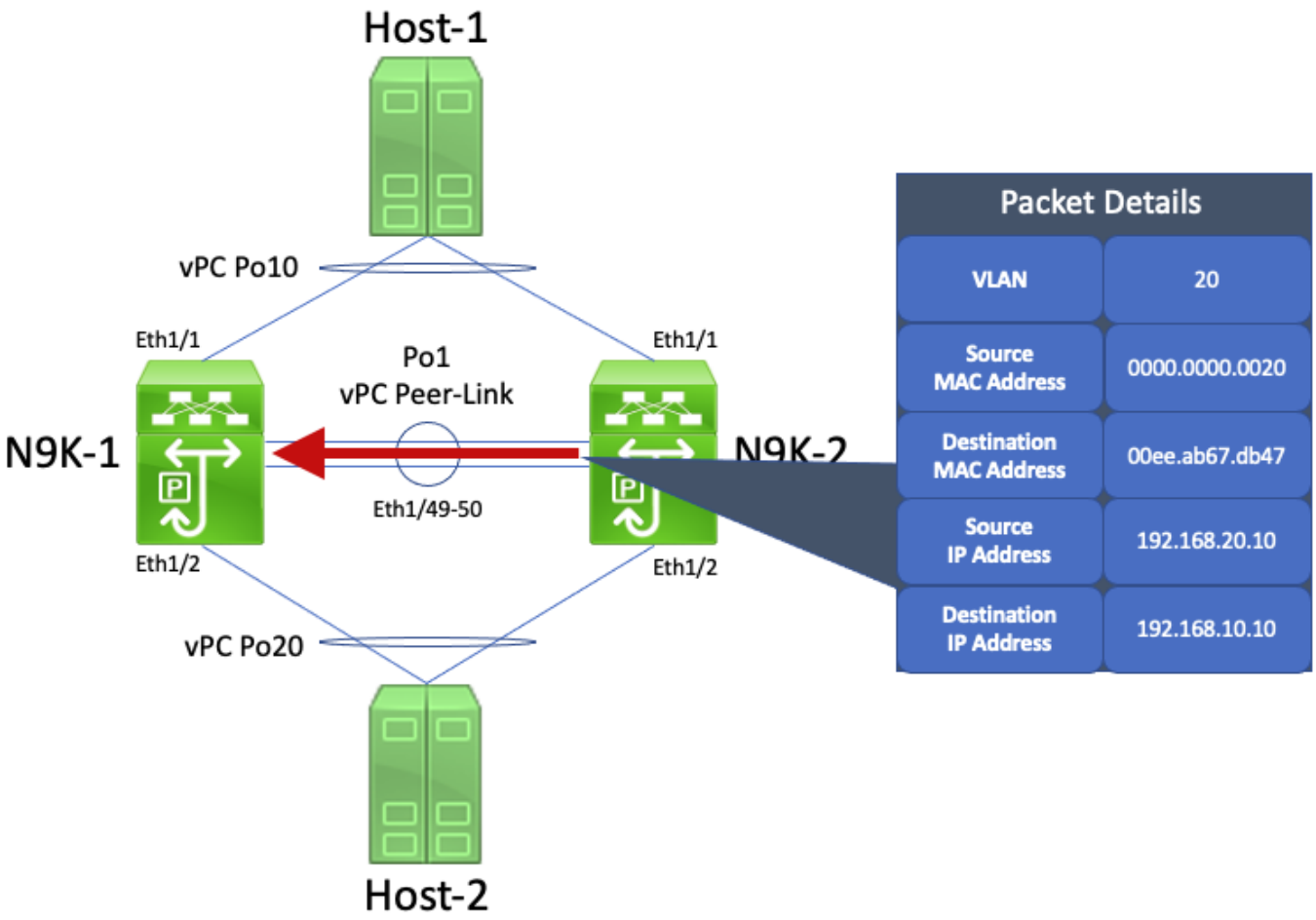


O Host-2 recebe esse pacote e gera um pacote de resposta de eco ICMP em resposta ao pacote de solicitação de eco ICMP do Host-1. No entanto, quando o Host-2 não entende o comportamento de encaminhamento padrão. Para otimizar o encaminhamento, o Host-2 não executa uma tabela de roteamento nem uma pesquisa de cache ARP para o endereço IP do Host-1 (192.168.10.10) – em vez disso, inverte os campos de endereço MAC de origem e de endereço MAC de destino do pacote de solicitação de eco ICMP que o Host -2 recebeu originalmente. Como resultado, o pacote de Resposta de Eco ICMP gerado pelo Host-2 tem um endereço IP origem de 192.168.20.10, um endereço IP destino de 192.168.10.10, um endereço MAC origem de 0000.0000.0020 e um endereço MAC destino de 00ee.ab67.db47.

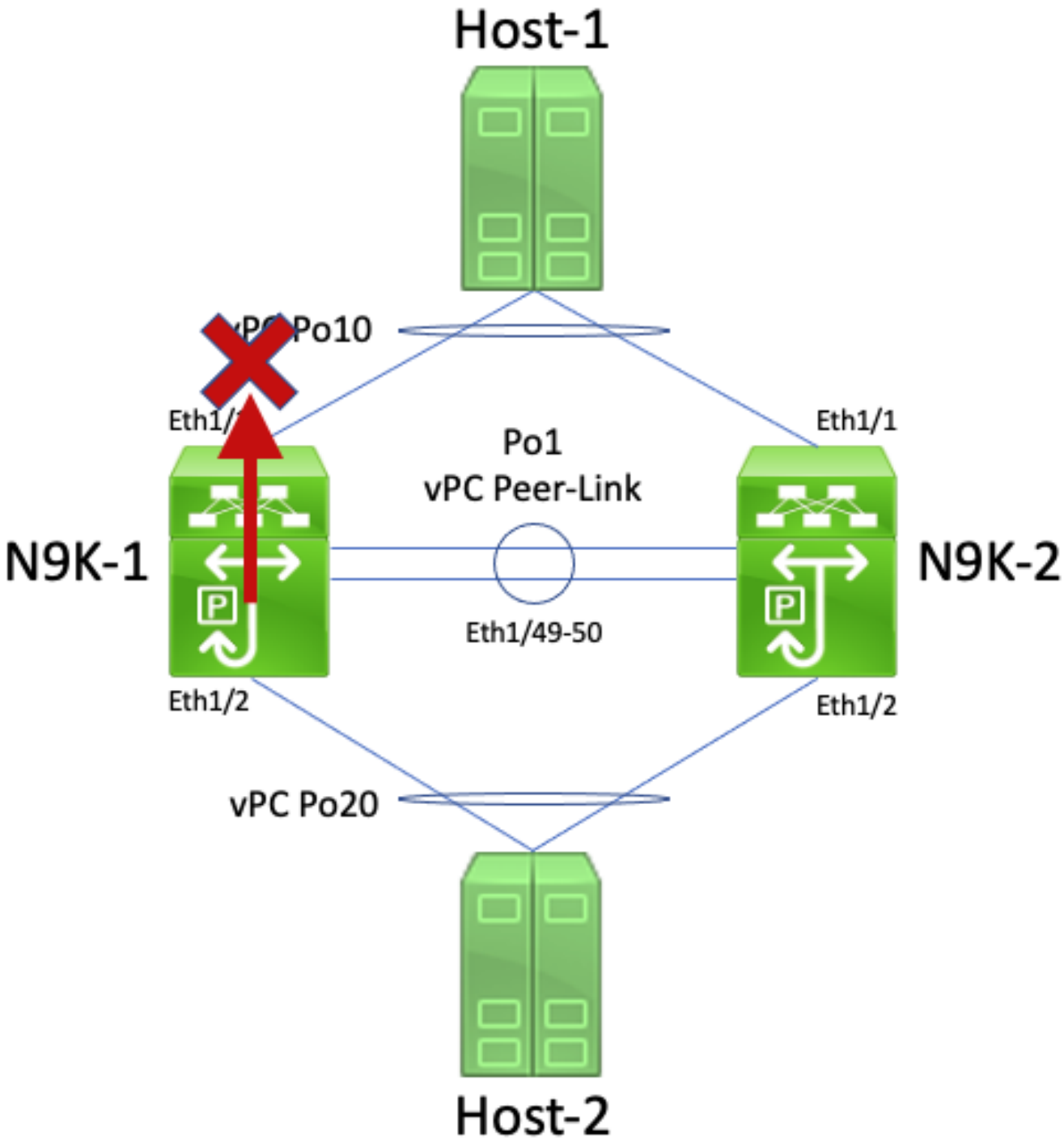
Se esse pacote de Resposta de Eco ICMP sair em direção a N9K-1, ele será encaminhado em direção ao Host-1 sem problemas. No entanto, considere um cenário em que esse pacote de resposta de eco ICMP sai para o N9K-2, conforme mostrado aqui.



O N9K-2 recebe esse pacote. Como esse pacote é destinado ao endereço MAC físico da VLAN 20 SVI da N9K-1, a N9K-2 encaminha esse pacote pelo vPC Peer-Link em direção à N9K-1, pois a N9K-2 não pode rotear esse pacote em nome da N9K-1. Veja abaixo um exemplo visual disso.



O N9K-1 recebe esse pacote. Como esse pacote é destinado ao endereço MAC físico da SVI da VLAN 20 do N9K-1, o N9K-1 pode encaminhar esse pacote de acordo com a tabela de roteamento local, independentemente do estado do plano de controle do HSRP. Esse pacote é roteado da VLAN 20 para a VLAN 10. No entanto, a interface de saída para essa rota resolve para vPC Po10, que está ativado em N9K-2. Isso é uma violação da regra de Prevenção de Loop vPC - se N9K-1 receber um pacote através do vPC Peer-Link, N9K-1 não poderá encaminhar esse pacote para fora de uma interface vPC se a mesma interface vPC estiver ativada em N9K-2. O N9K-1 descarta esse pacote como resultado dessa violação. Veja abaixo um exemplo visual disso.



Você pode resolver esse problema ativando o aprimoramento vPC Peer Gateway com o comando de configuração de domínio de vPC peer-gateway. Isso permite que o N9K-2 encaminhe o pacote de resposta de ICMP (e outros pacotes endereçados de forma semelhante) em nome do N9K-1, mesmo que o endereço MAC de destino do pacote pertença ao N9K-1 e não ao N9K-2. Como resultado, o N9K-2 pode encaminhar esse pacote da interface do Po10 de vPC, em vez de encaminhá-lo pelo vPC Peer-Link.

Roteamento/camada 3 por vPC (layer3 peer-router)

Esta seção descreve o aprimoramento de roteamento/camada 3 por vPC, que é ativado com o comando de configuração de domínio de vPC layer3 peer-router.

Observação: a formação de adjacências de protocolo de roteamento multicast (ou seja, adjacências de Protocol Independent Multicast [PIM]) sobre um vPC não é suportada com o aprimoramento de Roteamento/Camada 3 sobre vPC habilitado.

Overview

Em alguns ambientes, os clientes gostariam de conectar um roteador a um par de switches Nexus por meio do vPC e formar adjacências do protocolo de roteamento unicast pelo vPC com os dois pares de vPC. Como alternativa, os clientes podem querer conectar um roteador a um único par de vPC por meio de uma VLAN de vPC e formar adjacências do protocolo de roteamento unicast com ambos os pares de vPC pela VLAN de vPC. Como resultado, o roteador conectado por vPC teria o Equal-Cost Multi-Path (ECMP) para prefixos anunciados por ambos os switches Nexus. Isso pode ser preferível ao uso de links de roteamento dedicados entre o roteador conectado por vPC e os dois pares de vPC para preservar a utilização do endereço IP (são necessários 3 endereços IP, em vez de 4) ou reduzir a complexidade da configuração (interfaces encaminhadas em conjunto com as SVIs, especialmente em ambientes de VRF-Lite que exigiriam subinterfaces).

Historicamente, a formação de adjacências do protocolo de roteamento unicast por um vPC não era compatível com as plataformas Cisco Nexus. No entanto, os clientes podem ter implementado uma topologia em que as adjacências do protocolo de roteamento unicast se formam por um vPC sem problemas, mesmo que não sejam compatíveis. Após algumas alterações na rede (como um upgrade de software do roteador conectado por vPC ou os próprios pares de vPC, um failover de firewall e assim por diante), as adjacências do protocolo de roteamento unicast por um vPC param de funcionar, resultando na perda de pacotes para tráfego do plano de dados ou em adjacências do protocolo de roteamento unicast que não criam um ou ambos os pares de vPC. Os detalhes técnicos do motivo pelo qual esses cenários falham e não são compatíveis são abordados na seção Cenários de falha de exemplo deste documento.

O aprimoramento Routing/Layer 3 over vPC foi lançado para adicionar suporte à formação de adjacências do protocolo de roteamento unicast por um vPC. Isso é feito permitindo que os pacotes do protocolo de roteamento unicast com um TTL de 1 sejam encaminhados pelo vPC Peer-Link, sem diminuir o TTL do pacote. Como resultado, as adjacências do protocolo de roteamento unicast podem ser formadas por um vPC ou uma VLAN de vPC, sem problemas. O aprimoramento Routing/Layer 3 over vPC pode ser ativado com o comando de configuração de domínio de vPC layer3 peer-router após o aprimoramento vPC Peer Gateway ter sido ativado com o comando de configuração de domínio de vPC peer-gateway.

As versões de software NX-OS que lançaram o suporte para o aprimoramento Routing/Layer 3 over vPC para cada plataforma Cisco Nexus estão documentadas na Tabela 2 ("Suporte a adjacências dos protocolos de roteamento por VLANs de vPC") no documento Topologias compatíveis com o roteamento por Virtual Port Channel nas plataformas Nexus.

Caveats

Syslogs ocasionais VPC-2-L3_VPC_UNEQUAL_WEIGHT

Depois que o aprimoramento de Roteamento/Camada 3 sobre vPC estiver habilitado, os dois pares de vPC começarão a gerar syslogs semelhantes a um dos seguintes itens a cada hora:

```
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
```

Nenhum desses syslogs indica problemas com o switch. Esses syslogs são avisos para o administrador de que a configuração, o custo e a ponderação de roteamento devem ser idênticos em ambos os pares de vPC, quando o aprimoramento Routing/Layer 3 over vPC é ativado, para garantir que ambos os pares de vPC possam encaminhar o tráfego de forma idêntica. Isso não indica necessariamente que haja incompatibilidade na configuração, no custo ou na ponderação de roteamento em qualquer um dos pares de vPC.

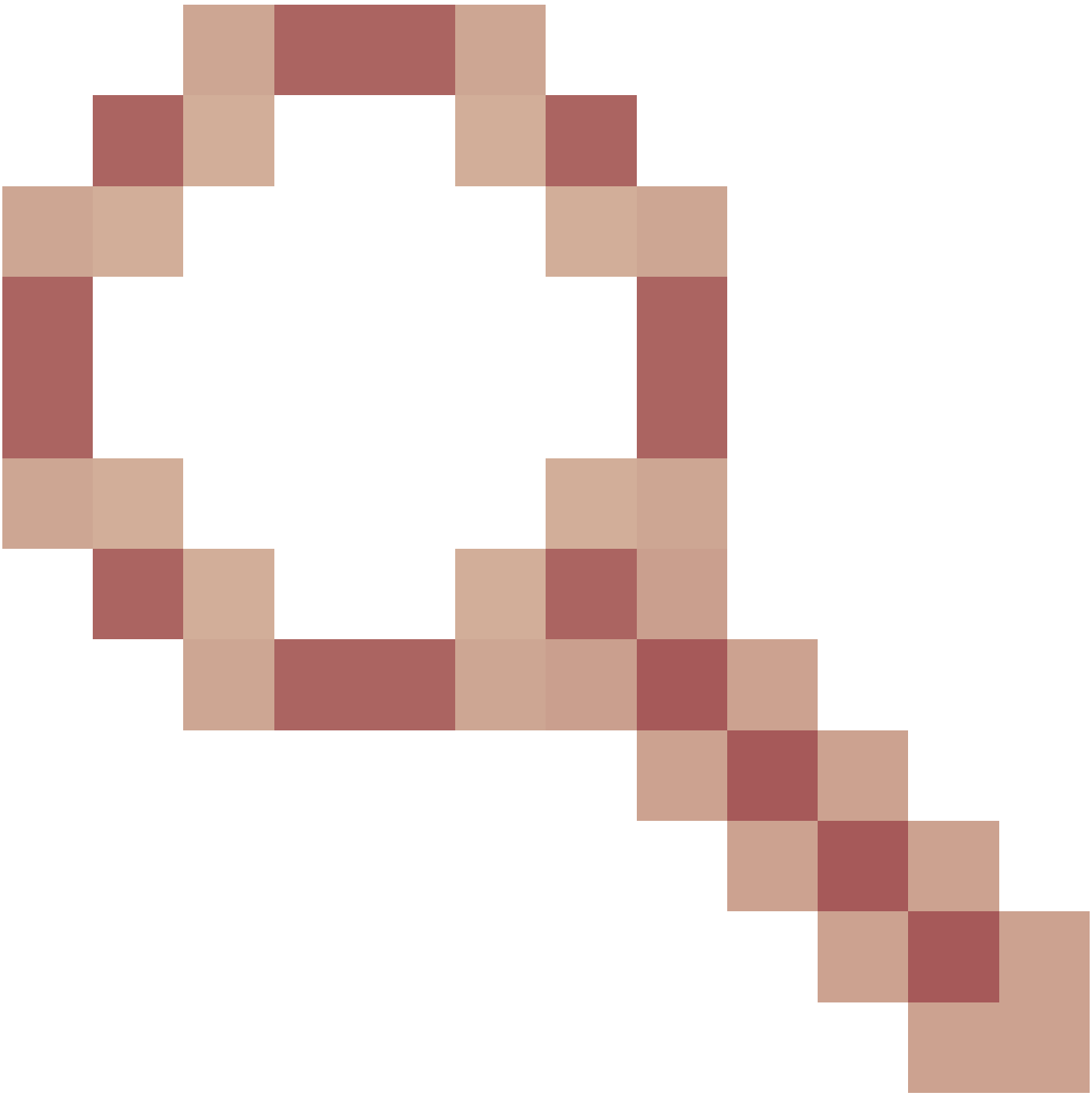
Esses syslogs podem ser desativados por meio da configuração mostrada aqui.

<#root>

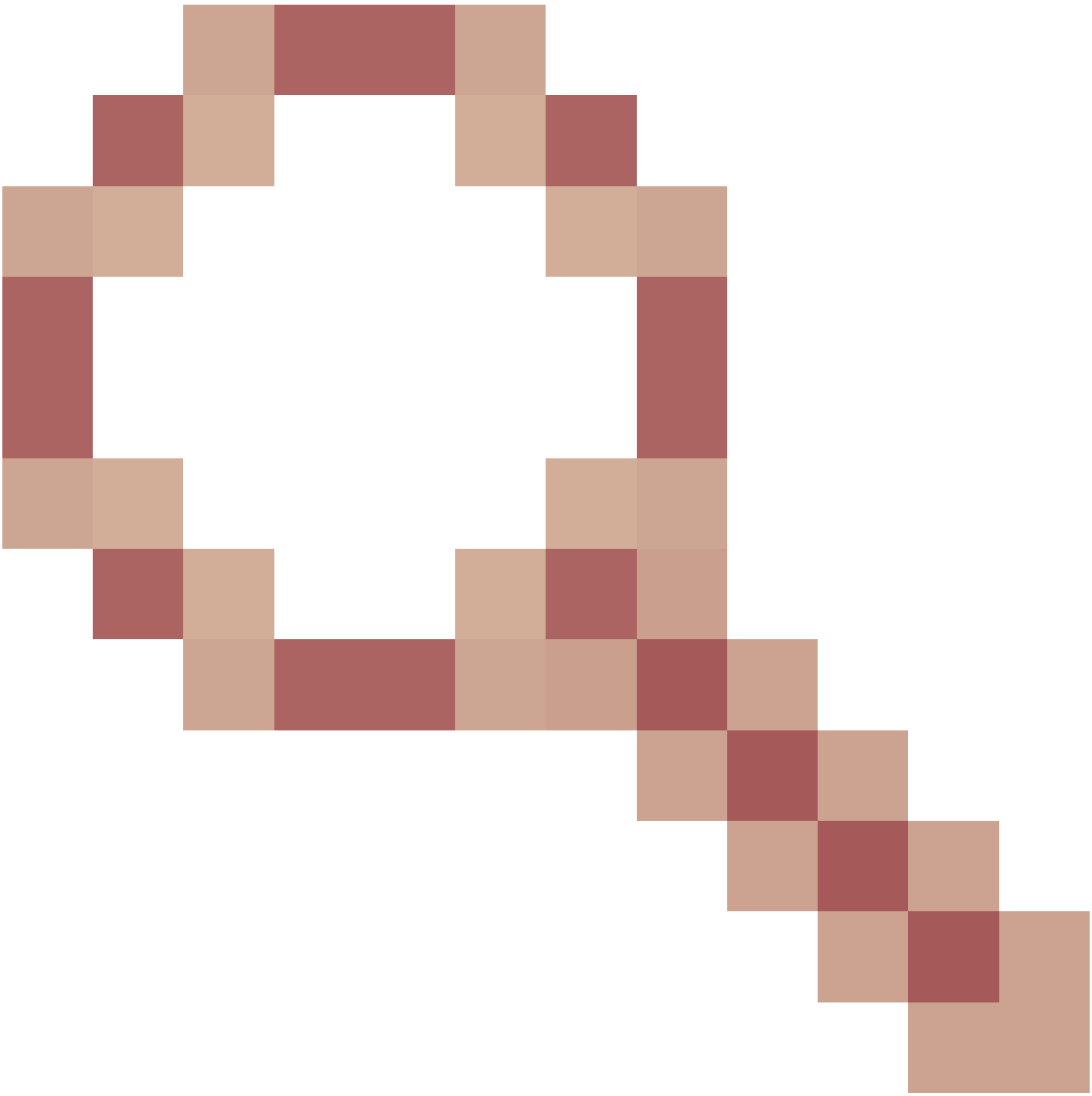
```
switch#
configure terminal
switch(config)#
vpc domain 1
switch(config-vpc-domain)#
no layer3 peer-router syslog
switch(config-vpc-domain)#
end
switch#
```

Essa configuração precisa ser executada em ambos os pares do vPC para desabilitar o syslog em ambos os pares do vPC.

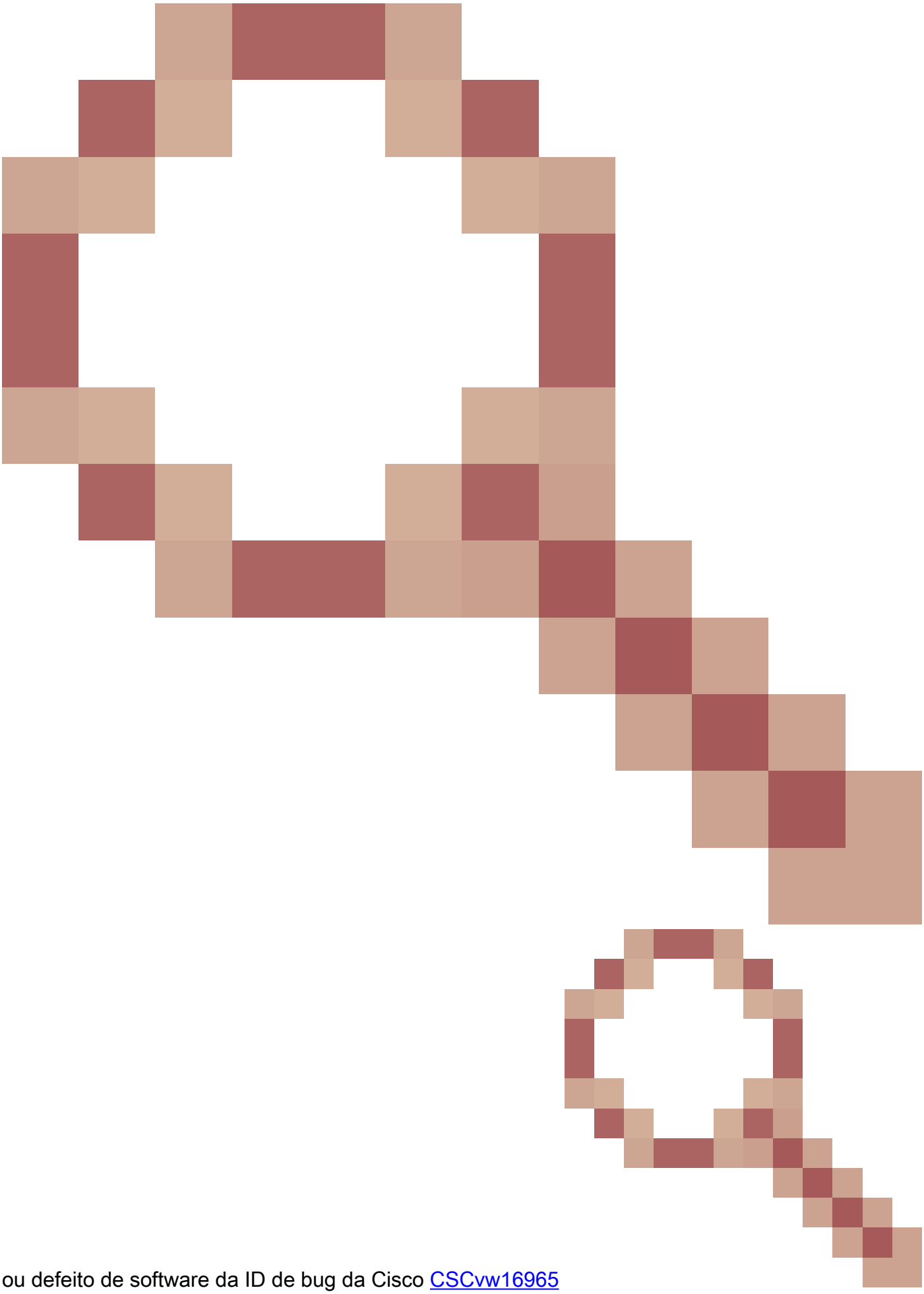
Tráfego plano de dados com TTL de 1 software encaminhado devido ao bug da Cisco ID CSCvs82183



e ID de bug da Cisco CSCvw16965



Quando o aprimoramento de Roteamento/Camada 3 sobre vPC é habilitado nos switches Nexus 9000 Series equipados com um ASIC de escala de nuvem executando uma versão do software NX-OS anterior à versão 9.3(6) do software NX-OS, o tráfego do plano de dados que não está associado a um protocolo de roteamento unicast que tem um TTL de 1 é apontado para o supervisor e encaminhado em software em vez de hardware. Dependendo se o switch Nexus é um switch de chassi fixo (também chamado de "Top of Rack") ou um switch de chassi modular (também chamado de "End of Row"), bem como a versão atual do software NX-OS do switch, a causa raiz desse problema pode ser atribuída a qualquer defeito de software da ID de bug CSCvs82183 da Cisco



ou defeito de software da ID de bug da Cisco CSCvw16965

. Ambos os defeitos de software afetam apenas os switches Nexus 9000 Series equipados com um ASIC em escala de nuvem – as outras plataformas de hardware Cisco Nexus não são afetadas por nenhum desses problemas. Para obter mais detalhes, consulte as informações em cada defeito de software individual.

Para evitar esses defeitos de software, a Cisco recomenda fazer upgrade do NX-OS versão de software 9.3(6) ou posterior. Como recomendação geral, a Cisco indica o upgrade regular para a versão de software do NX-OS atual recomendada para o switch Nexus 9000 Series referenciada pelo documento Recomendações de versões do Cisco NX-OS para switches Cisco Nexus 9000 Series.

Configuração

Veja abaixo um exemplo de como configurar o aprimoramento Routing/Layer 3 over vPC.

Neste exemplo, o N9K-1 e o N9K-2 são pares de vPC em um domínio de vPC. Ambos os pares de vPC já têm o aprimoramento vPC Peer Gateway ativado, o que é necessário para ativar o aprimoramento Routing/Layer 3 over vPC. Ambos os pares de vPC têm uma SVI na VLAN 10 que é ativada no processo 1 do OSPF. N9K-1 e N9K-3 estão presos em um estado OSPF EXSTART/EXCHANGE com um roteador OSPF conectado por vPC com um endereço IP e a ID do vizinho 192.168.10.3.

<#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
```

Podemos ativar o aprimoramento de roteamento/camada 3 por vPC através do comando de configuração de domínio de vPC layer3 peer-router. Isso evita que um peer do vPC diminua o TTL dos pacotes do protocolo de roteamento unicast roteados como resultado do aprimoramento do vPC Peer Gateway estar habilitado.

```
<#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#
```

Você pode verificar se o aprimoramento Routing/Layer 3 over vPC está funcionando conforme o esperado, validando se a adjacência do OSPF com o vizinho do OSPF conectado por vPC para o estado FULL, logo após ativar o aprimoramento Routing/Layer 3 over vPC.

```
<#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
```

Impacto

A ativação do aprimoramento Routing/Layer 3 over vPC não causa impactos inerentes ao domínio de vPC. Isso significa que quando você habilita o aprimoramento de Roteamento/Camada 3 sobre vPC, nenhum par de vPC suspende vPCs, nem nenhum tráfego de plano de dados é inerentemente afetado pela habilitação desse aprimoramento.

No entanto, se as adjacências do protocolo de roteamento dinâmico, que estavam anteriormente inativas como resultado de não ter o aprimoramento Routing/Layer 3 over vPC ativado, surgirem de repente como resultado da ativação desse aprimoramento (dependendo da função das adjacências afetadas do protocolo de roteamento, dos prefixos específicos anunciados por essas adjacências e do estado atual da tabela de roteamento unicast), algumas interrupções poderão ser observadas ao ativar o aprimoramento Routing/Layer 3 over vPC.

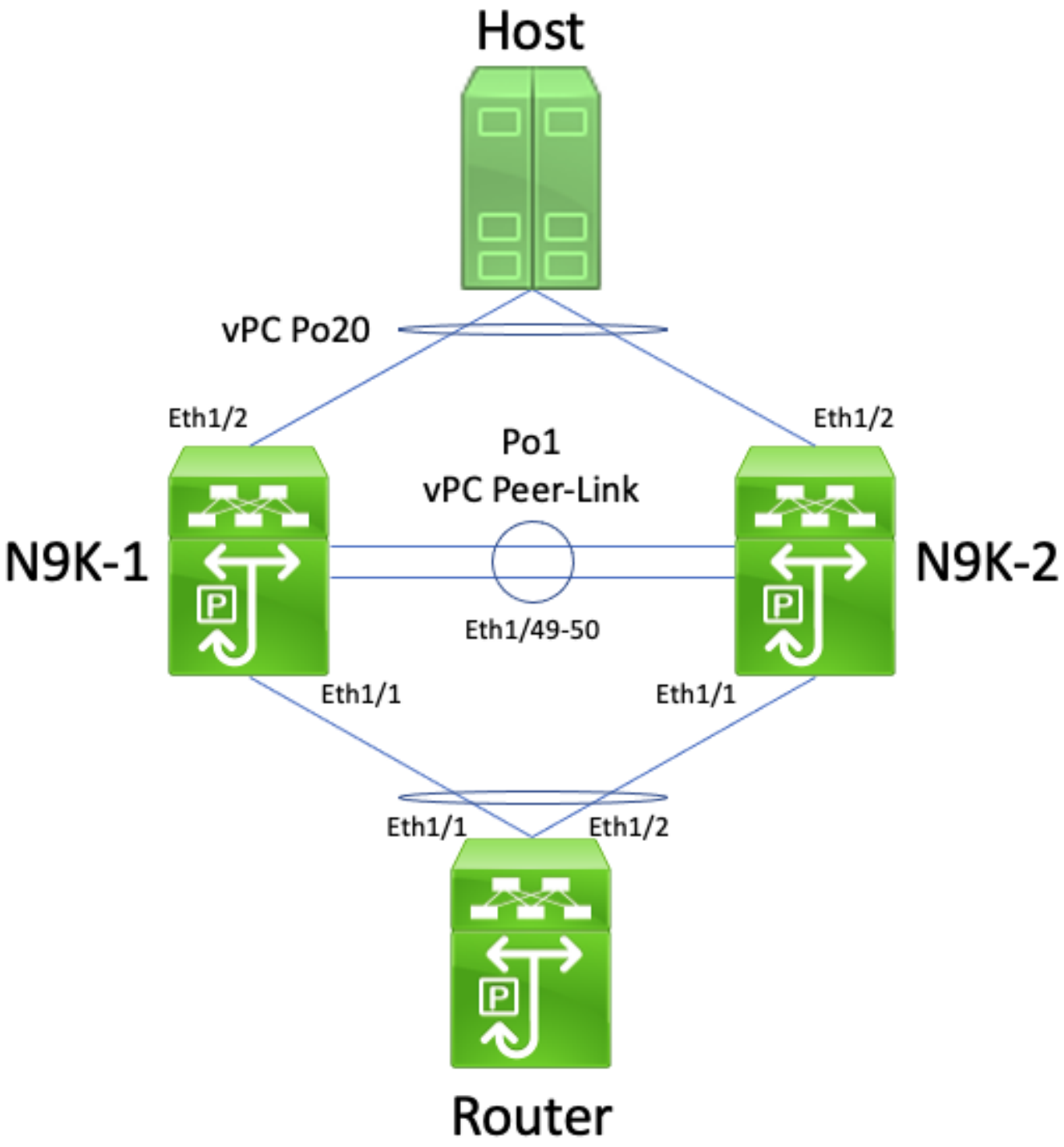
Por esse motivo, a Cisco aconselha que os clientes habilitem essa melhoria durante uma janela de manutenção com a expectativa de que possa haver interrupção no plano de controle e no plano de dados, a menos que os clientes estejam extremamente confiantes de que as adjacências do protocolo de roteamento afetadas não afetam significativamente a operação da rede.

A Cisco também recomenda analisar atentamente a seção Avisos deste documento para detectar todos os defeitos de software que afetam a versão de software NX-OS, os quais possam fazer com que o tráfego natural do plano de dados com um TTL de 1 seja processado no software, em vez do hardware.

Cenários de falha de exemplo

Adjacências do protocolo de roteamento unicast por um vPC sem vPC Peer Gateway

Considere a topologia mostrada aqui:



Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway não foi ativado. A interface Po1 é o vPC Peer-Link. Um roteador com um nome de host de Router está conectado por meio do Po10 de vPC ao N9K-1 e ao N9K-2. Um host está conectado ao N9K-1 e ao N9K-2 por meio do Po20 de vPC. A interface Po10 do roteador é um port-channel roteado que é ativado de acordo com um protocolo de roteamento unicast. O N9K-1 e o N9K-2 têm interfaces SVI ativadas no mesmo protocolo de roteamento unicast e estão no mesmo domínio de transmissão que o roteador.

As adjacências do protocolo de roteamento unicast por um vPC sem o aprimoramento vPC Peer Gateway ativado não são compatíveis, pois a decisão de hash do ECMP do roteador conectado por vPC e a decisão de hash do port-channel da Camada 2 podem ser diferentes. Nessa

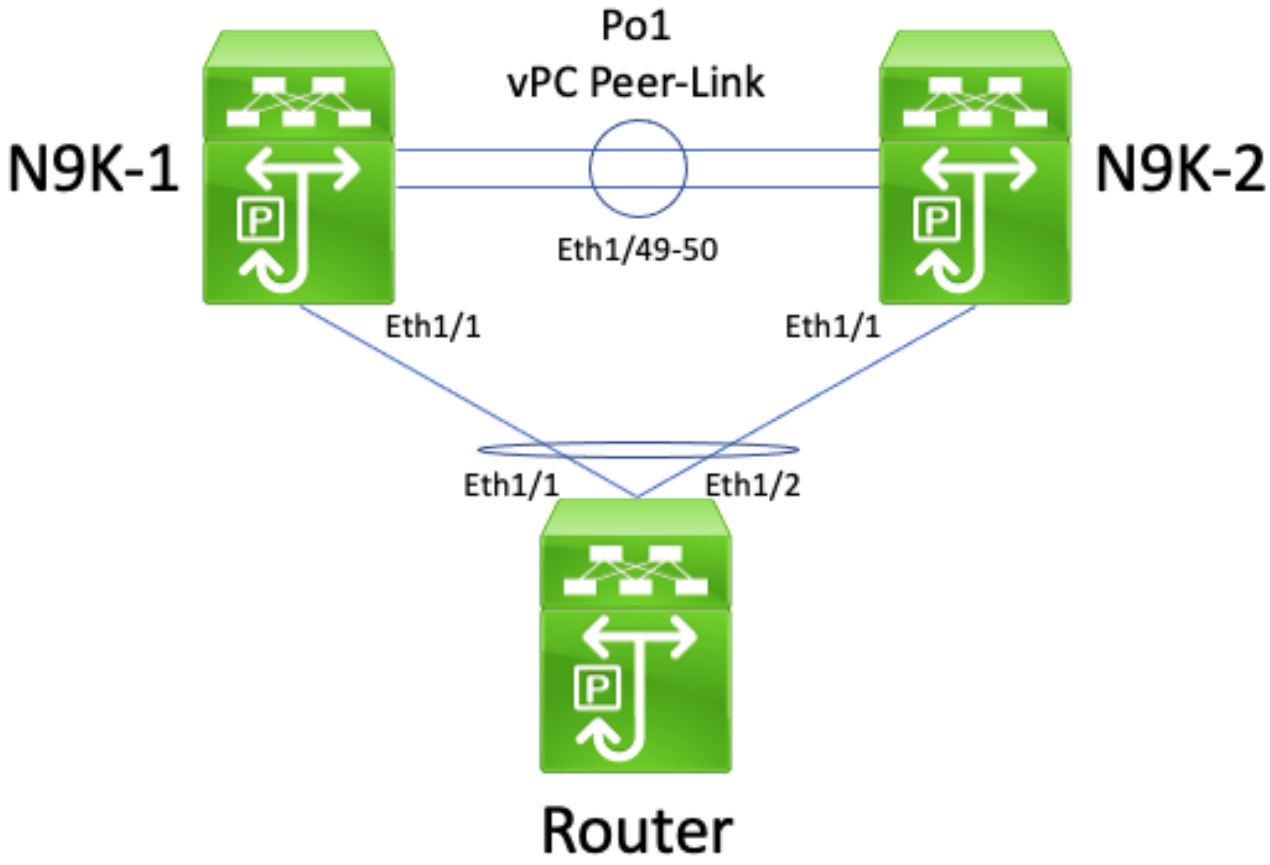
topologia, adjacências de protocolos de roteamento se formariam com êxito entre o Roteador, N9K-1 e N9K-2. Considere o fluxo de tráfego entre o roteador e o host. O tráfego do plano de dados que atravessa o roteador destinado ao host pode ser regravado com um endereço MAC de destino que pertence ao endereço MAC da SVI do N9K-1 (devido à decisão de hash do ECMP tomada pelo roteador), mas sai da interface Ethernet1/2 (devido à decisão de hash do port-channel da camada 2 tomada pelo roteador).

O N9K-2 recebe esse pacote e o encaminha pelo vPC Peer-Link, já que o endereço MAC de destino pertence ao N9K-1 e o aprimoramento do vPC Peer Gateway (que permite ao N9K-2 rotear o pacote em nome do N9K-1) não está habilitado. O N9K-1 recebe esse pacote no vPC Peer-Link e reconhece que precisaria encaminhar o pacote do Ethernet1/2 no Po20 de vPC. Isso viola a regra de prevenção de loop do vPC, portanto, o N9K-1 descarta o pacote no hardware. Como resultado, você pode observar problemas de conectividade ou perda de pacotes para alguns fluxos que atravessam o domínio de vPC nessa topologia.

Para resolver esse problema, ative o aprimoramento vPC Peer Gateway com o comando de configuração de domínio de vPC peer-gateway e ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer3 peer-router. Para minimizar a interrupção, você deve ativar ambos os aprimoramentos de vPC em sucessão rápida, para que o cenário de falha descrito nas Adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway não tenha tempo para ocorrer.

Adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway

Considere a topologia mostrada aqui:



Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. A interface Po1 é o vPC Peer-Link. Um roteador com um nome de host de Router está conectado por meio do Po10 de vPC ao N9K-1 e ao N9K-2. A interface Po10 do roteador é um port-channel roteado que é ativado de acordo com um protocolo de roteamento unicast. O N9K-1 e o N9K-2 têm interfaces SVI ativadas no mesmo protocolo de roteamento unicast e estão no mesmo domínio de transmissão que o roteador.

Não há suporte para adjacências dos protocolos de roteamento unicast por um vPC com o aprimoramento vPC Peer Gateway ativado, pois o aprimoramento vPC Peer Gateway pode impedir a formação de adjacências do protocolo de roteamento unicast entre o roteador conectado por vPC e os dois pares de vPC. Nessa topologia, uma adjacência de protocolo de roteamento entre o Roteador e N9K-1 ou N9K-2 pode falhar ao ser ativada como esperado, dependendo de como os pacotes de protocolo de roteamento unicast originados pelo Roteador para hash N9K-1 ou N9K-2 através do vPC Po10.

Todos os roteadores podem enviar e receber pacotes do protocolo de roteamento multicast de link-local (normalmente chamados de pacotes "Hello") sem problemas, pois esses pacotes são enviados para a VLAN de vPC com êxito. No entanto, considere um cenário em que um pacote do protocolo de roteamento unicast proveniente do roteador destinado ao N9K-1 sai do Ethernet1/2 para o N9K-2, devido à decisão de hash do port-channel da camada 2 do roteador. Esse pacote é destinado ao endereço MAC SVI de N9K-1, mas ingressa na interface Ethernet1/1 de N9K-2. O N9K-2 vê que o pacote é destinado ao endereço MAC SVI do N9K-1, que é instalado na tabela de endereços MAC do N9K-2 com o flag "G" ou "Gateway", devido à habilitação do aprimoramento do vPC Peer Gateway. Como resultado, o N9K-2 tenta rotear localmente o pacote

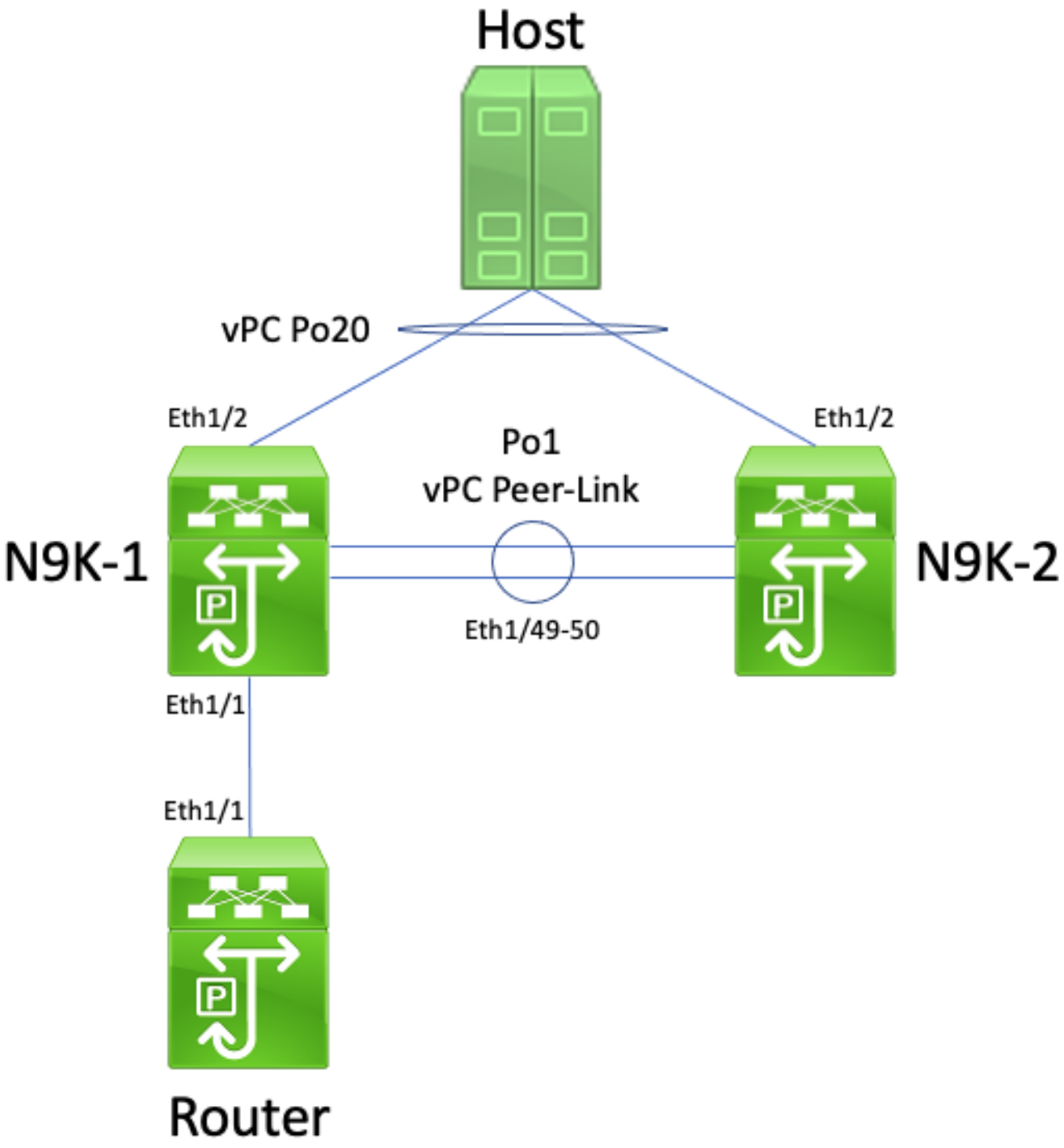
do protocolo de roteamento unicast em nome do N9K-1.

No entanto, ao rotear o pacote, o Time to Live (TTL) do pacote é diminuído e o TTL da maioria dos pacotes do protocolo de roteamento unicast é 1. Como resultado, o TTL do pacote é decrementado para 0 e descartado por N9K-2. Da perspectiva do N9K-1, o N9K-1 está recebendo pacotes do protocolo de roteamento multicast de link-local do roteador e pode enviar pacotes do protocolo de roteamento unicast ao roteador, mas não está recebendo pacotes do protocolo de roteamento unicast do roteador. Como resultado, o N9K-1 desfaz a adjacência do protocolo de roteamento com o Roteador e reinicia sua máquina de estado finito local para o protocolo de roteamento. Da mesma forma, o Roteador reinicia sua máquina de estado finito local para o protocolo de roteamento.

Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer 3 peer-router. Isso permite que os pacotes do protocolo de roteamento unicast com um TTL de 1 sejam encaminhados pelo vPC Peer-Link, sem diminuir o TTL do pacote. Como resultado, as adjacências do protocolo de roteamento unicast podem ser formadas por um vPC ou uma VLAN de vPC, sem problemas.

Adjacências do protocolo de roteamento unicast por uma VLAN de vPC sem vPC Peer Gateway

Considere a topologia mostrada aqui:



Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway não foi ativado. A interface Po1 é o vPC Peer-Link. Um roteador com um nome de host do roteador está conectado por meio do Ethernet1/1 ao Ethernet1/1 do N9K-1. A interface Ethernet1/1 do roteador é uma interface roteada que é ativado de acordo com um protocolo de roteamento unicast. O N9K-1 e o N9K-2 têm interfaces SVI ativadas no mesmo protocolo de roteamento unicast e estão no mesmo domínio de transmissão que o roteador.

As adjacências do protocolo de roteamento unicast por uma VLAN de vPC sem o aprimoramento vPC Peer Gateway ativado não são compatíveis, pois a decisão de hash do ECMP do roteador conectado por VLAN de vPC pode fazer com que o N9K-2 descarte o tráfego do plano de dados

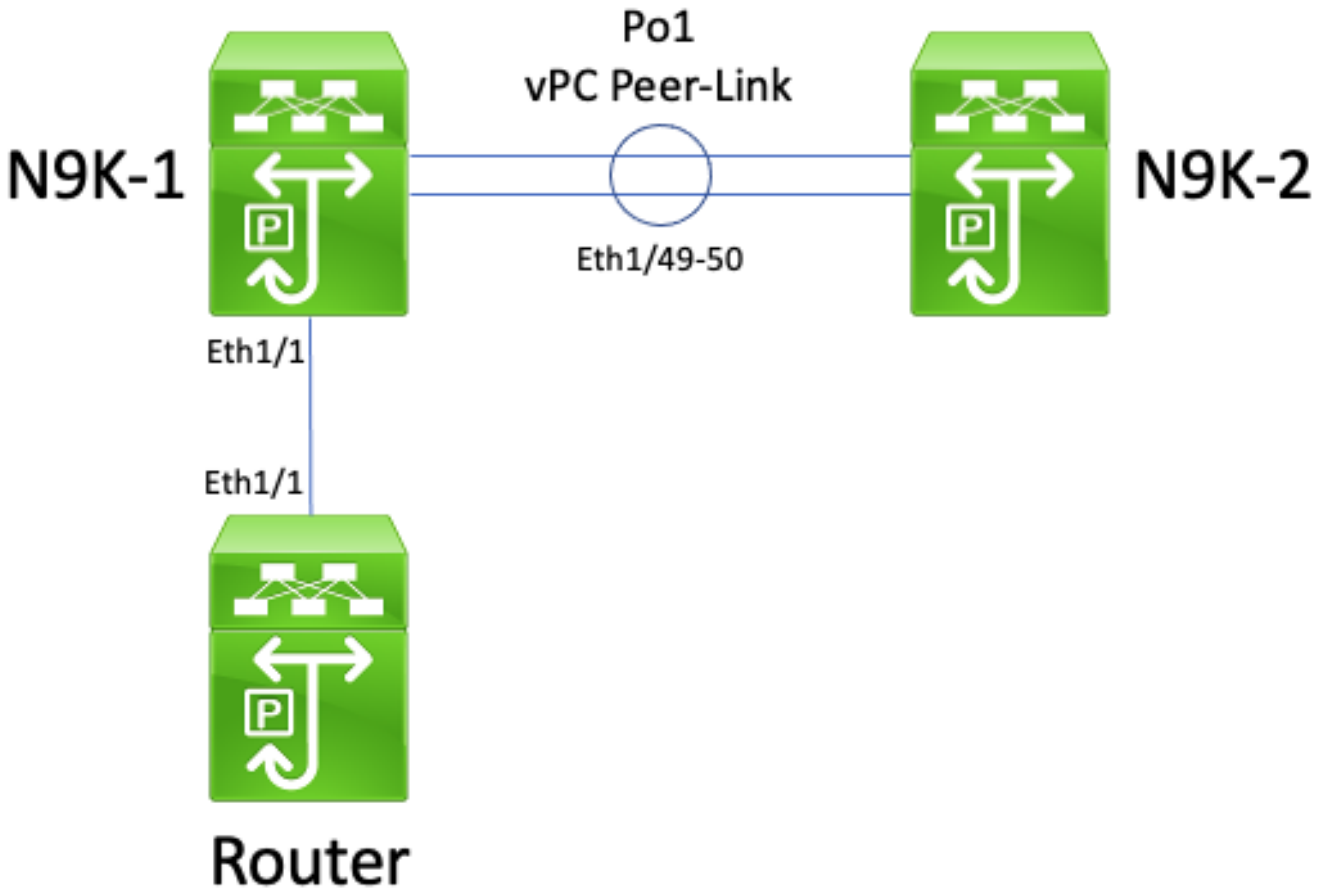
devido à violação da regra de prevenção de loop do vPC. Nessa topologia, adjacências de protocolos de roteamento se formariam com êxito entre o Roteador, N9K-1 e N9K-2. Considere o fluxo de tráfego entre o roteador e o host. O tráfego do plano de dados que atravessa o roteador destinado ao host pode ser regravado com um endereço MAC de destino que pertence ao endereço MAC da SVI do N9K-2 (devido à decisão de hash do ECMP tomada pelo roteador) e sair da interface Ethernet1/1 para o N9K-1.

O N9K-1 recebe esse pacote e o encaminha pelo vPC Peer-Link, já que o endereço MAC de destino pertence ao N9K-2 e o aprimoramento do vPC Peer Gateway (que permite ao N9K-1 rotear o pacote em nome do N9K-2) não está habilitado. O N9K-2 recebe esse pacote no vPC Peer-Link e reconhece que precisaria encaminhar o pacote do Ethernet1/2 no Po20 de vPC. Isso viola a regra de prevenção de loop do vPC, portanto, o N9K-2 descarta o pacote no hardware. Como resultado, você pode observar problemas de conectividade ou perda de pacotes para alguns fluxos que atravessam o domínio de vPC nessa topologia.

Para resolver esse problema, ative o aprimoramento vPC Peer Gateway com o comando de configuração de domínio de vPC peer-gateway e ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer3 peer-router. Para minimizar a interrupção, você deve ativar ambos os aprimoramentos de vPC em sucessão rápida, para que o cenário de falha descrito nas Adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway não tenha tempo para ocorrer.

Adjacências do protocolo de roteamento unicast por uma VLAN de vPC com vPC Peer Gateway

Considere a topologia mostrada aqui:



Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. A interface Po1 é o vPC Peer-Link. Um roteador com um nome de host do roteador está conectado por meio do Ethernet1/1 ao Ethernet1/1 do N9K-1. A interface Ethernet1/1 do roteador é uma interface roteada que é ativado de acordo com um protocolo de roteamento unicast. O N9K-1 e o N9K-2 têm interfaces SVI ativadas no mesmo protocolo de roteamento unicast e estão no mesmo domínio de transmissão que o roteador.

As adjacências do protocolo de roteamento unicast sobre uma VLAN vPC com o aprimoramento vPC Peer Gateway habilitado não são suportadas porque o aprimoramento vPC Peer Gateway impede que adjacências do protocolo de roteamento unicast sejam formadas entre o roteador conectado à VLAN vPC e o peer vPC ao qual o roteador conectado à VLAN vPC não está diretamente conectado. Nessa topologia, uma adjacência de protocolo de roteamento entre o Roteador e o N9K-2 falha em surgir como esperado como resultado de pacotes de protocolo de roteamento unicast N9K-1 destinados ao endereço MAC SVI do N9K-2 devido ao aprimoramento do vPC Peer Gateway estar ativado. Como os pacotes estão sendo encaminhados, o TTL deve ser diminuído. Normalmente, os pacotes do protocolo de roteamento unicast têm um TTL de 1, e um roteador que diminui o TTL de um pacote para 0 deve descartar esse pacote.

Todos os roteadores podem enviar e receber pacotes do protocolo de roteamento multicast de link-local (normalmente chamados de pacotes "Hello") sem problemas, pois esses pacotes são enviados para a VLAN de vPC com êxito. No entanto, considere um cenário em que um pacote do protocolo de roteamento unicast proveniente do roteador destinado ao N9K-2 sai do

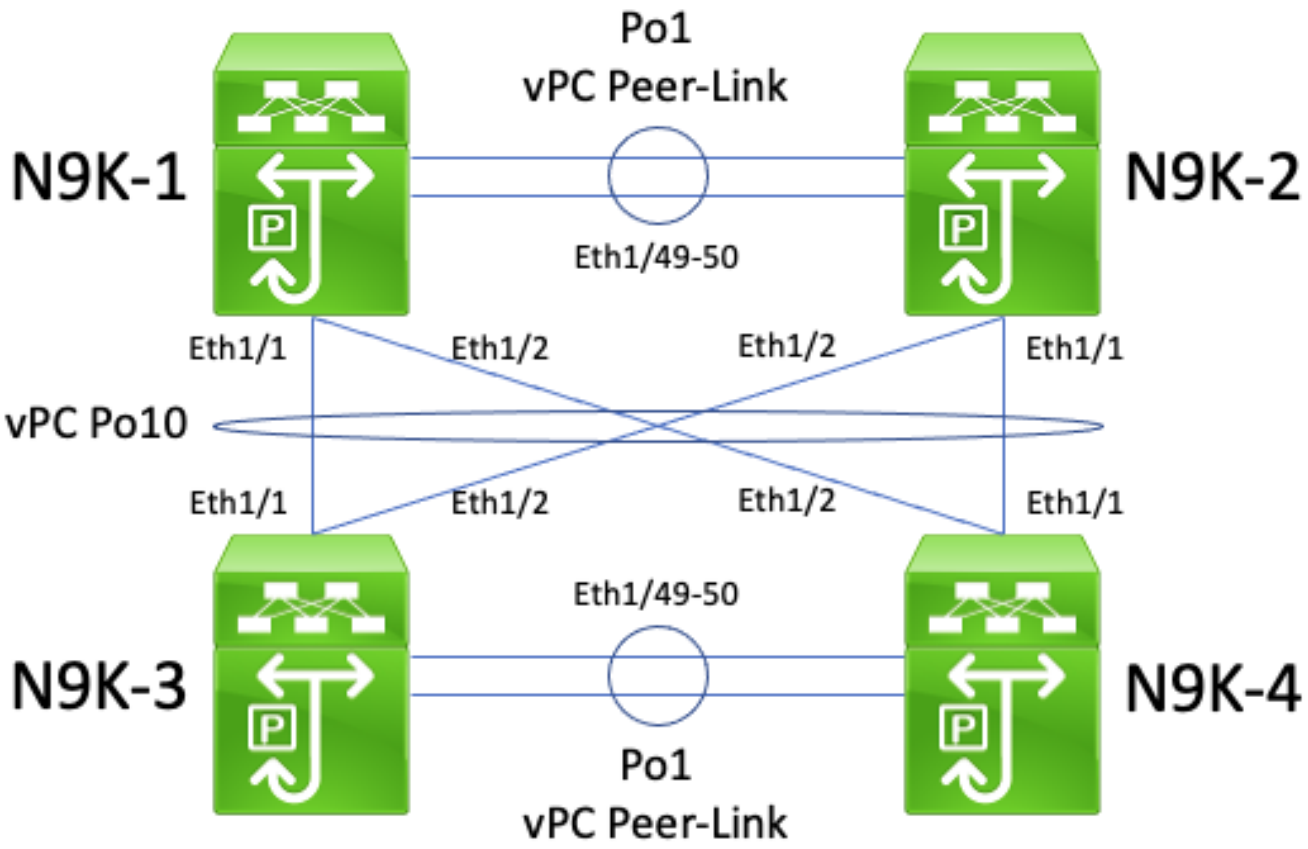
Ethernet1/1 para o N9K-1. Esse pacote é destinado ao endereço MAC SVI de N9K-2, mas ingressa na interface Ethernet1/1 de N9K-1. O N9K-1 vê que o pacote é destinado ao endereço MAC SVI do N9K-2, que é instalado na tabela de endereços MAC do N9K-1 com o flag "G" ou "Gateway", devido à habilitação do aprimoramento do vPC Peer Gateway. Como resultado, o N9K-1 tenta rotear localmente o pacote do protocolo de roteamento unicast em nome do N9K-2.

No entanto, ao rotear o pacote, o TTL do pacote é diminuído e o TTL da maioria dos pacotes do protocolo de roteamento unicast é 1. Como resultado, o TTL do pacote é decrementado para 0 e descartado por N9K-1. Da perspectiva do N9K-2, o N9K-2 está recebendo pacotes do protocolo de roteamento multicast de link-local do roteador e pode enviar pacotes do protocolo de roteamento unicast ao roteador, mas não está recebendo pacotes do protocolo de roteamento unicast do roteador. Como resultado, o N9K-2 desfaz a adjacência do protocolo de roteamento com o Roteador e reinicia sua máquina de estado finito local para o protocolo de roteamento. Da mesma forma, o Roteador reinicia sua máquina de estado finito local para o protocolo de roteamento.

Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer 3 peer-router. Isso permite que os pacotes do protocolo de roteamento unicast com um TTL de 1 sejam encaminhados pelo vPC Peer-Link, sem diminuir o TTL do pacote. Como resultado, as adjacências do protocolo de roteamento unicast podem ser formadas por um vPC ou uma VLAN de vPC, sem problemas.

Adjacências do protocolo de roteamento unicast por um back-to-back vPC com vPC Peer Gateway

Considere a topologia mostrada aqui:



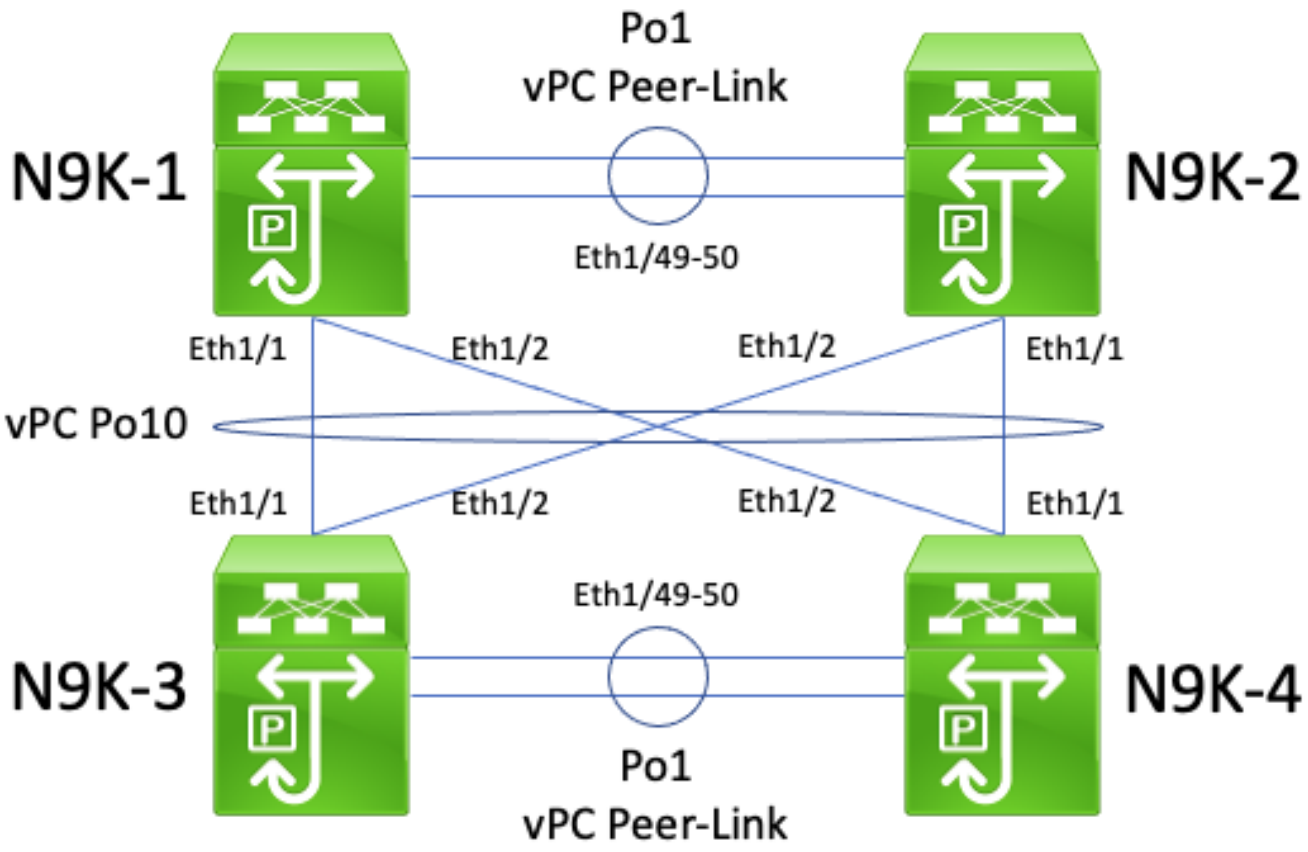
Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. Os switches Nexus N9K-3 e N9K-4 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. Ambos os domínios de vPC são conectados entre si por um Po10 de vPC back-to-back. Todos os quatro switches têm interfaces SVI ativadas em um protocolo de roteamento unicast e estão no mesmo domínio de transmissão.

Não há suporte para adjacências dos protocolos de roteamento unicast por vPCs back-to-back com o aprimoramento vPC Peer Gateway ativado, pois o aprimoramento vPC Peer Gateway pode impedir a formação de adjacências do protocolo de roteamento unicast entre um domínio de vPC e outro. Nessa topologia, uma adjacência de protocolo de roteamento entre N9K-1 e N9K-3 ou N9K-4 (ou ambos) pode falhar ao ser ativada como esperado. Da mesma forma, uma adjacência do protocolo de roteamento entre o N9K-2 e o N9K-3 ou o N9K-4 (ou ambos) pode não surgir conforme o esperado. Isso ocorre porque os pacotes do protocolo de roteamento unicast podem ser destinados a um roteador (por exemplo, o N9K-3), mas encaminhados a um outro roteador (por exemplo, o N9K-4) de acordo com a decisão de hash do port-channel da camada 2 do roteador de origem.

A causa do problema é idêntica à causa do problema descrita na seção Adjacências do protocolo de roteamento unicast por um vPC com vPC Peer Gateway deste documento. Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer 3 peer-router. Isso permite que os pacotes do protocolo de roteamento unicast com um TTL de 1 sejam encaminhados pelo vPC Peer-Link, sem diminuir o TTL do pacote. Como resultado, as adjacências do protocolo de roteamento unicast podem ser formadas por um vPC back-to-back, sem problemas.

Adjacências do OSPF por vPC com vPC Peer Gateway em que o prefixo está presente no OSPF LSDB, mas não na tabela de roteamento

Considere a topologia mostrada aqui:



Nesta topologia, os switches Nexus N9K-1 e N9K-2 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. Os switches Nexus N9K-3 e N9K-4 são pares de vPC em um domínio de vPC em que o aprimoramento vPC Peer Gateway foi ativado. Ambos os domínios de vPC são conectados entre si por um Po10 de vPC back-to-back. Todos os quatro switches têm interfaces SVI ativadas em um protocolo de roteamento unicast e estão no mesmo domínio de transmissão. O N9K-4 é o roteador designado (DR) do OSPF do domínio de transmissão, enquanto o N9K-3 é o roteador designado de backup (BDR) do OSPF do domínio de transmissão.

Nesse cenário, uma adjacência do OSPF entre o N9K-1 e o N9K-3 passa para um estado FULL devido aos pacotes do OSPF unicast que saem da Ethernet1/1 de ambos os switches. Da mesma forma, uma adjacência do OSPF entre o N9K-2 e o N9K-3 passa para um estado FULL devido aos pacotes do OSPF unicast que saem da Ethernet1/2 de ambos os switches.

Entretanto, uma adjacência do OSPF entre o N9K-1 e o N9K-4 fica presa em um estado EXSTART ou EXCHANGE, devido aos pacotes do OSPF unicast que saem da Ethernet1/1 de ambos os switches e são descartados pelo N9K-2 e pelo N9K-4, conforme descrito na seção Adjacências do protocolo roteamento unicast por vPC back-to-back com vPC Peer Gateway deste documento. Da mesma forma, uma adjacência do OSPF entre o N9K-2 e o N9K-4 fica presa em um estado EXSTART ou EXCHANGE, devido aos pacotes do OSPF unicast que saem da

Ethernet1/2 de ambos os switches e são descartados pelo N9K-1 e pelo N9K-3, conforme descrito na seção Adjacências do protocolo roteamento unicast por vPC back-to-back com vPC Peer Gateway deste documento.

Como resultado, o N9K-1 e o N9K-2 estão em um estado FULL com o BDR do domínio de transmissão, mas estão em um estado EXSTART ou EXCHANGE com o DR do domínio de transmissão. O DR e o BDR de um domínio de transmissão mantêm uma cópia completa do OSPF Link State Data Base (LSDB), mas os roteadores OSPF DROTHER devem estar em um estado FULL com o DR do domínio de transmissão para instalar os prefixos aprendidos por meio do OSPF no DR ou no BDR. Como resultado, tanto N9K-1 como N9K-2 parecem ter prefixos aprendidos de N9K-3 e N9K-4 presentes no LSDB do OSPF, mas esses prefixos não são instalados na tabela de roteamento unicast até a transição N9K-1 e N9K-2 para um estado FULL com N9K-4 (o DR para o domínio de broadcast).

Para resolver esse problema, ative o aprimoramento Routing/Layer 3 over vPC com o comando de configuração de domínio de vPC layer 3 peer-router. Isso permite que os pacotes do protocolo de roteamento unicast com um TTL de 1 sejam encaminhados pelo vPC Peer-Link, sem diminuir o TTL do pacote. Como resultado, as adjacências do protocolo de roteamento unicast podem ser formadas por um vPC back-to-back, sem problemas. Como resultado, o N9K-1 e o N9K-2 fazem transições para um estado FULL com o N9K-4 (o DR para o domínio de broadcast) e instalam com êxito os prefixos aprendidos do N9K-3 e do N9K-4 via OSPF em suas respectivas tabelas de roteamento unicast.

Informações Relacionadas

- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.3(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.2(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 10.1(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 9.3(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 9.2(x)
- Guia de configuração das interfaces Cisco Nexus 9000 Series NX, versão 7.x
- Guia de configuração das interfaces Cisco Nexus 7000 Series NX 8.x
- Guia de configuração das interfaces Cisco Nexus 7000 Series NX 7.x
- Guia de design e configuração: Práticas recomendadas para Virtual Port Channels (vPC) nos switches Cisco Nexus 7000 Series
- Topologias compatíveis com o roteamento por Virtual Port Channel nas plataformas Nexus

Sobre esta tradução
A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês (link fornecido) seja sempre consultado.