

为NXOS配置VXLAN vPC交换矩阵对等并对其进行故障排除

目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[网络图](#)

[配置](#)

[TCAM配置](#)

[TCAM切分](#)

[vPC配置](#)

[VPC域](#)

[Keep-alive](#)

[虚拟对等链路的第3层接口](#)

[VPC对等链路](#)

[上行链路](#)

[主干配置](#)

[采用入口复制封装的广播、未知单播和组播流量](#)

[采用入口复制解除封装的广播、未知单播和组播流量](#)

[使用组播封装的广播、未知单播和组播流量](#)

[使用组播解封的广播、未知单播和组播流量](#)

[验证](#)

[相关信息](#)

简介

本文档介绍如何为NXOS和BUM流量配置和验证vPC交换矩阵对等。

先决条件

要求

建议掌握下列主题的相关知识：

- vPC (虚拟端口通道)
- 虚拟可扩展局域网(VXLAN)

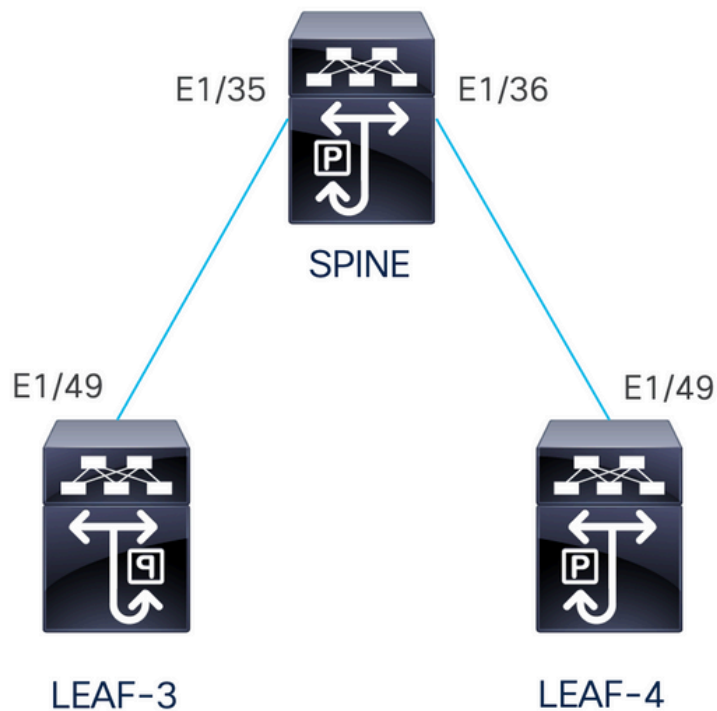
使用的组件

本文档中的信息基于以下软件和硬件版本：

- 适用于枝叶交换机的N9K-C93240YC-FX2版本：10.3(3)
- 主干交换机的N9K-C9336C-FX2版本：10.3(3)

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。

网络图



vPC交换矩阵对等提供增强的双宿主接入解决方案，而不会为vPC对等链路浪费物理端口。此功能保留了传统vPC的所有特征。

在此部署中，枝叶3和枝叶4配置为具有交换矩阵对等的vPC。

配置

TCAM配置

在配置之前，需要检查TCAM内存：

```

LEAF-4(config-if)# sh hardware access-list tcam region
          NAT ACL[nat] size = 0
        Ingress PACL [ing-ifac] size = 0
          VACL [vac] size = 0
        Ingress RAACL [ing-racl] size = 2304
        Ingress L2 QOS [ing-l2-qos] size = 256
        Ingress L3/VLAN QOS [ing-l3-vlan-qos] size = 512
          Ingress SUP [ing-sup] size = 512
  
```

```

Ingress L2 SPAN filter [ing-l2-span-filter] size = 256
Ingress L3 SPAN filter [ing-l3-span-filter] size = 256
    Ingress FSTAT [ing-fstat] size = 0
        span [span] size = 512
    Egress RAcl [egr-racl] size = 1792
    Egress SUP [egr-sup] size = 256
    Ingress Redirect [ing-redirect] size = 0
    Egress L2 QOS [egr-l2-qos] size = 0
    Egress L3/VLAN QOS [egr-l3-vlan-qos] size = 0
    Ingress Netflow/Analytics [ing-netflow] size = 512
        Ingress NBM [ing-nbm] size = 0
        TCP NAT ACL[tcp-nat] size = 0
    Egress sup control plane[egr-copp] size = 0
    Ingress Flow Redirect [ing-flow-redirect] size = 0 <<<<<<<<<
Ingress PAcl IPv4 Lite [ing-ifacl-ipv4-lite] size = 0
Ingress PAcl IPv6 Lite [ing-ifacl-ipv6-lite] size = 0
    Ingress CNTACL [ing-cntacl] size = 0
    Egress CNTACL [egr-cntacl] size = 0
    MCAST NAT ACL[mcast-nat] size = 0
    Ingress DAcl [ing-dacl] size = 0
    Ingress PAcl Super Bridge [ing-pacl-sb] size = 0
    Ingress Storm Control [ing-storm-control] size = 0
    Ingress VAcl redirect [ing-vacl-nh] size = 0
    Egress PAcl [egr-ifacl] size = 0
    Egress Netflow [egr-netflow] size = 0

```

vPC交换矩阵对等需要应用TCAM区域划分流重定向。TCAM雕刻要求先保存配置并重新加载交换机，然后才能使用该功能。

TCAM上的此空间是双宽度，因此我们可以分配的最小值为512。

TCAM切分

在此场景中，Ing-racl有足够的空间占用512并将这些512分配给Ing-flow-redirect。

```

LEAF-4(config-if)# hardware access-list tcam region ing-racl 1792
Please save config and reload the system for the configuration to take effect

```

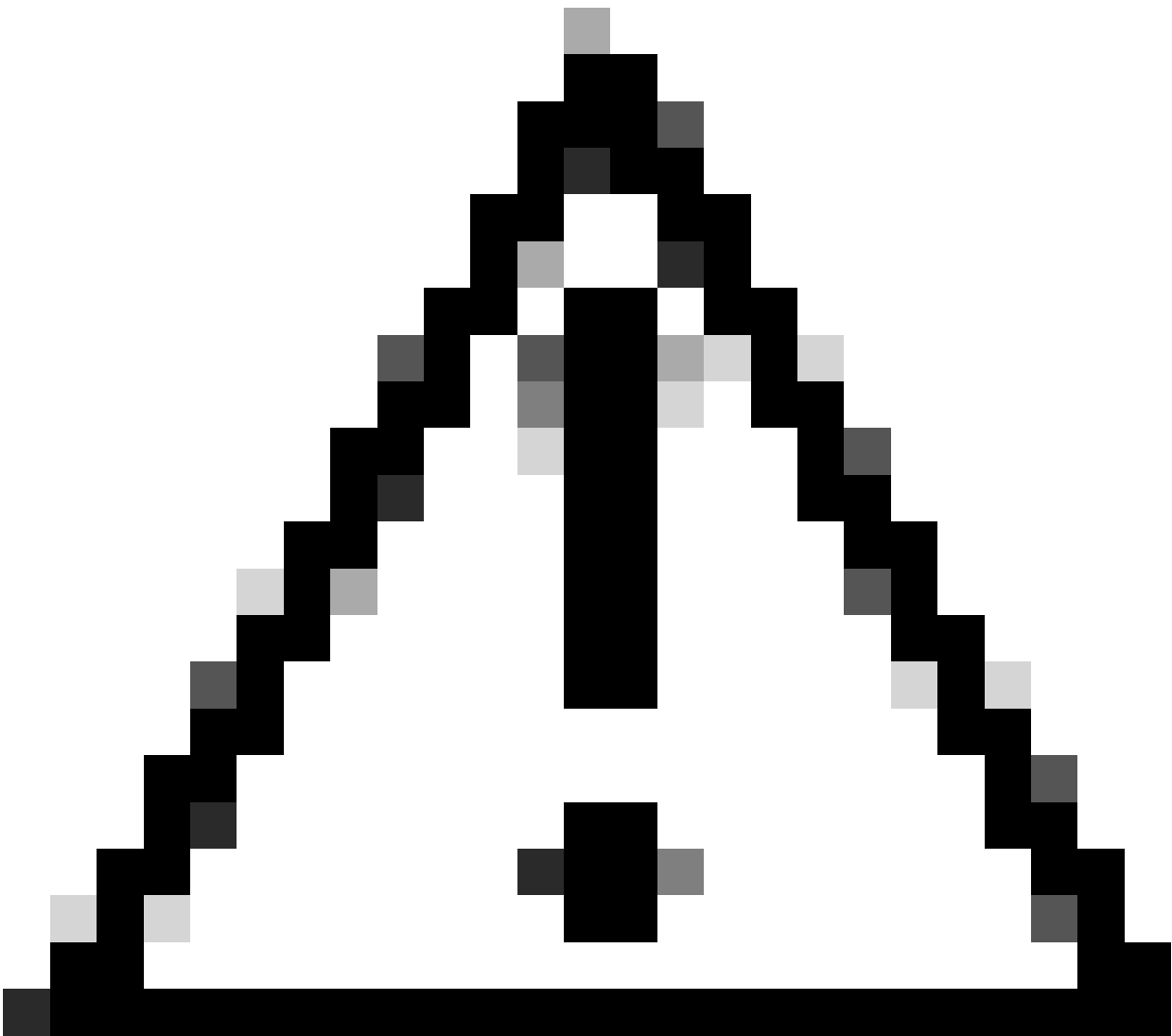
```

LEAF-4(config)# hardware access-list tcam region ing-flow-redirect 512
Please save config and reload the system for the configuration to take effect

```



```
Egress L3/VLAN QOS [egr-l3-vlan-qos] size = 0
Ingress Netflow/Analytics [ing-netflow] size = 512 <<<<<
    Ingress NBM [ing-nbm] size = 0
    TCP NAT ACL [tcp-nat] size = 0
Egress sup control plane [egr-copp] size = 0
Ingress Flow Redirect [ing-flow-redirect] size = 0
Ingress PACL IPv4 Lite [ing-ifacl-ipv4-lite] size = 0
Ingress PACL IPv6 Lite [ing-ifacl-ipv6-lite] size = 0
    Ingress CNTACL [ing-cntacl] size = 0
    Egress CNTACL [egr-cntacl] size = 0
    MCAST NAT ACL [mcast-nat] size = 0
    Ingress DAACL [ing-dacl] size = 0
Ingress PACL Super Bridge [ing-pacl-sb] size = 0
Ingress Storm Control [ing-storm-control] size = 0
    Ingress VACL redirect [ing-vacl-nh] size = 0
    Egress PACL [egr-ifacl] size = 0
```



注意：请确保在TCAM上进行更改后重新加载设备，否则，由于TCAM上没有应用更改，VPC不会出现。

vPC配置

VPC域

在VPC域中的枝叶3和枝叶4上，配置是为保持连接和虚拟对等链路指定IP地址

```
vpc domain 1
  peer-keepalive destination 192.168.1.1 source 192.168.1.2 vrf management
  virtual peer-link destination 10.10.10.2 source 10.10.10.1 dscp 56

interface port-channel1
  vpc peer-link
```

Keep-alive

vPC对等设备之间的任何直接第3层链路必须仅用于对等保持连接。它必须位于仅用于保持连接的单独VRF中。在此场景中，我们使用的是交换机的接口管理。

```
LEAF-3
interface mgmt0
  vrf member management
  ip address 192.168.1.1/24
```

```
LEAF-4
interface mgmt0
  vrf member management
  ip address 192.168.1.2/24
```

虚拟对等链路的第3层接口

用于虚拟对等链路的第3层接口不能与我们用于保持连接的接口相同，您可以使用用于底层接口的相同环回，也可以是Nexus上的专用环回

在这里，loopback0用于底层，loopback2是用于虚拟对等链路的专用环回，而loopback1是与接口NVE关联的接口。

```
LEAF-3
interface loopback0
  ip address 10.1.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback1
  ip address 172.16.1.2/32
  ip address 172.16.1.1/32 secondary
  ip router ospf 1 area 0.0.0.0
```

```
ip pim sparse-mode

interface loopback2
ip address 10.10.10.2/32
ip router ospf 1 area 0.0.0.0
```

LEAF-4

```
interface loopback0
ip address 10.1.1.2/32
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode

interface loopback1
ip address 172.16.1.3/32
ip address 172.16.1.1/32 secondary
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode

interface loopback2
ip address 10.10.10.1/32
ip router ospf 1 area 0.0.0.0
```

VPC对等链路

即使我们不打算将物理接口分配给port-channel，对等链路也需要分配port-channel。

```
LEAF-3(config-if)# sh run interface port-channel 1 membership
```

```
interface port-channel1
switchport
switchport mode trunk
spanning-tree port type network
vpc peer-link
```

上行链路

配置的最后一部分是使用port-type fabric命令配置指向SPINE的两个枝叶上的链路。

```
interface Ethernet1/49
port-type fabric <<<<<<<<<
medium p2p
ip unnumbered loopback0
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode
no shutdown
```



注意：如果不配置端口类型交换矩阵，您将无法看到Nexus正在生成的保活

主干配置

建议在脊柱上设置QoS，以匹配VPC域上配置的DSCP值，因为vPC交换矩阵对等对等链路是通过传输网络建立的。

用于同步端口状态信息、VLAN信息、VLAN到VNI映射、主机MAC地址和IGMP监听组的控制平面信息CFS消息通过交换矩阵传输。CFS消息使用相应的DSCP值进行标记，必须在传输网络中保护该值。

```
class-map type qos match-all CFS
  match dscp 56
```

```
policy-map type qos CFS
  class CFS
    Set qos-group 7 <<< Depending on the platform it can be 4
```



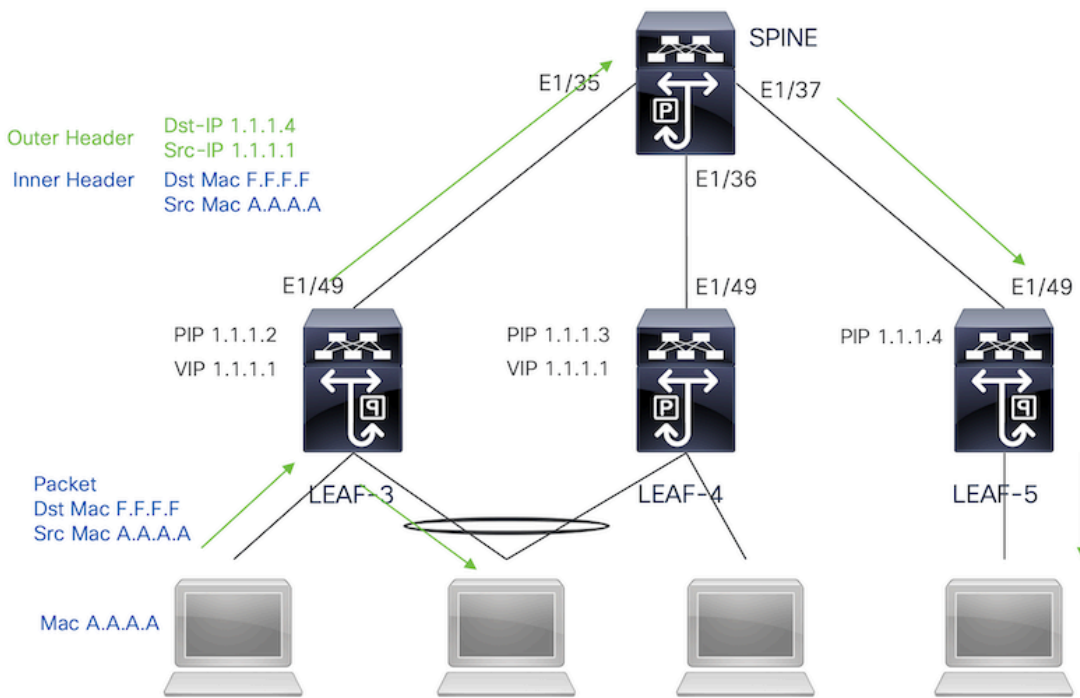
```
interface Ethernet 1/35-36
 service-policy type qos input CFS
```

采用入口复制封装的广播、未知单播和组播流量

当nexus收到需要广播的数据包时，会生成该数据包的2个副本。

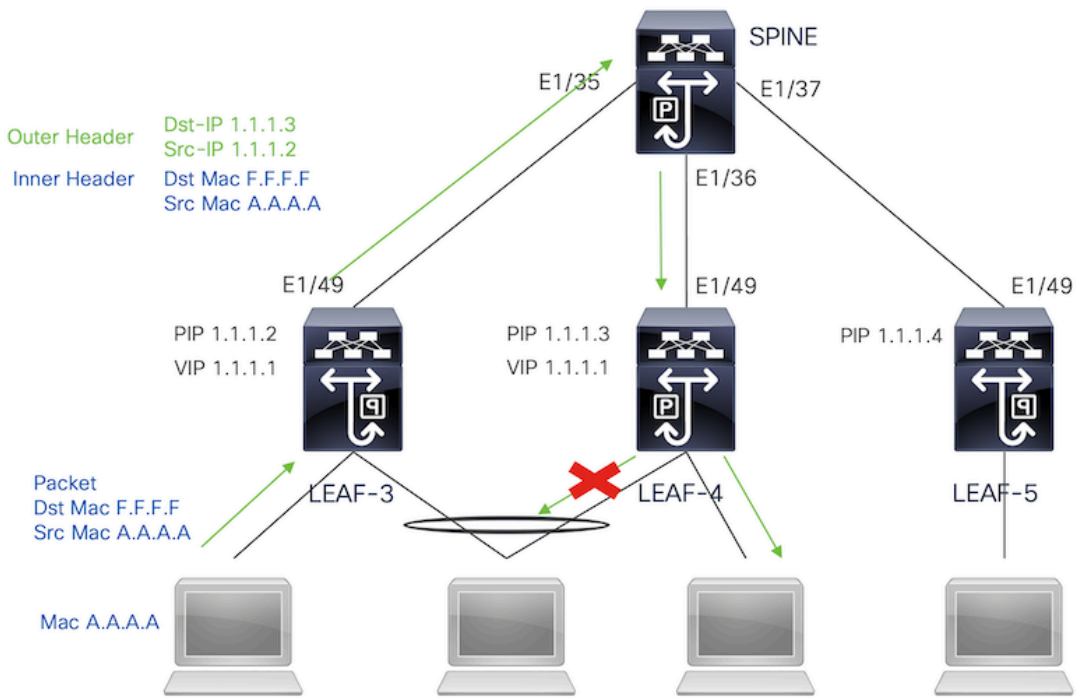
1. 到VNI泛洪列表中的所有远程VTEPS，包括本地接入端口
2. 到远程VPC对等设备

对于第一个副本，Nexus使用辅助IP地址的源IP和远程VTEP的目标IP以及本地接入端口封装流量。



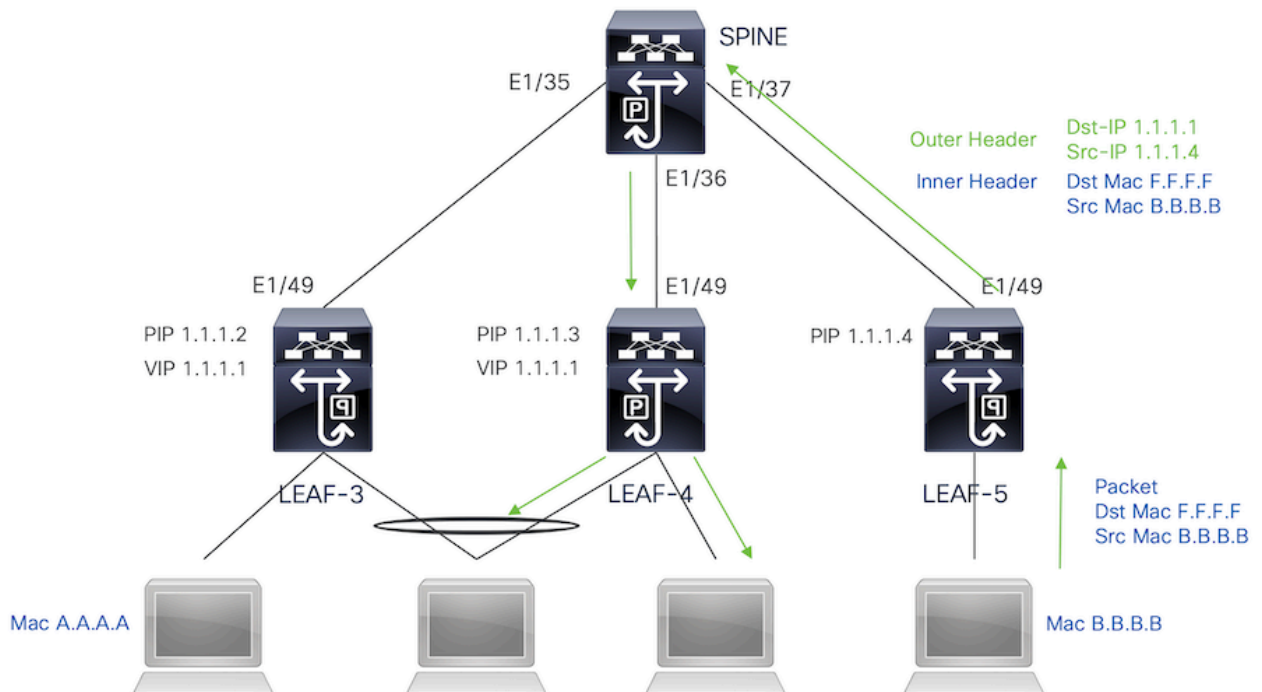
对于第二个副本，它将发送到远程VPC对等设备，源IP将是环回的主要IP，而目标IP是远程VPC对等设备的PIP。

从主干收到数据包后，远程VTEP将仅将数据包转发到孤立端口。



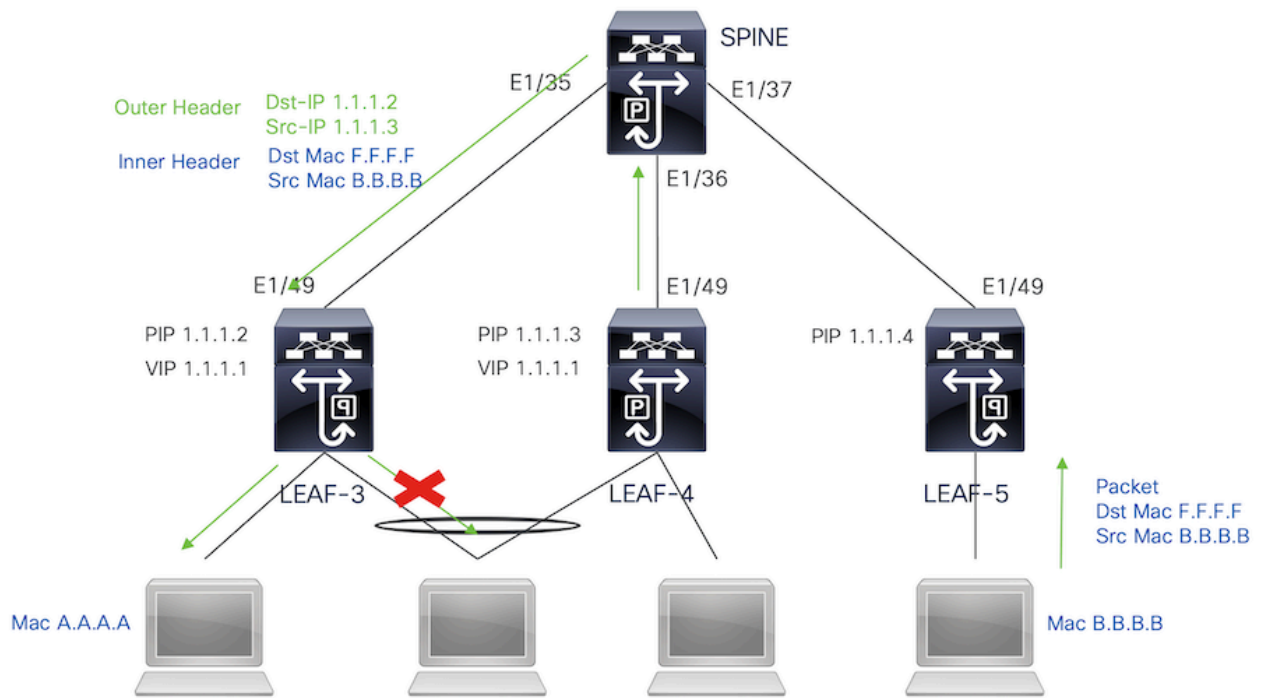
采用入口复制解除封装的广播、未知单播和组播流量

由于从另一个VTEP接收的BUM流量的目标IP是流量散列到其中一个VPC设备的VIP，因此它会解封数据包并将其发送到接入端口。



为了使流量到达远程VPC对等设备上连接的孤立端口，nexus会生成数据包的副本，并将仅使用主要IP地址作为源/目标IP将数据包发送到远程VPC。

在远程vpc对等设备上收到流量后，nexus会解封流量并仅将其转发到孤立端口。

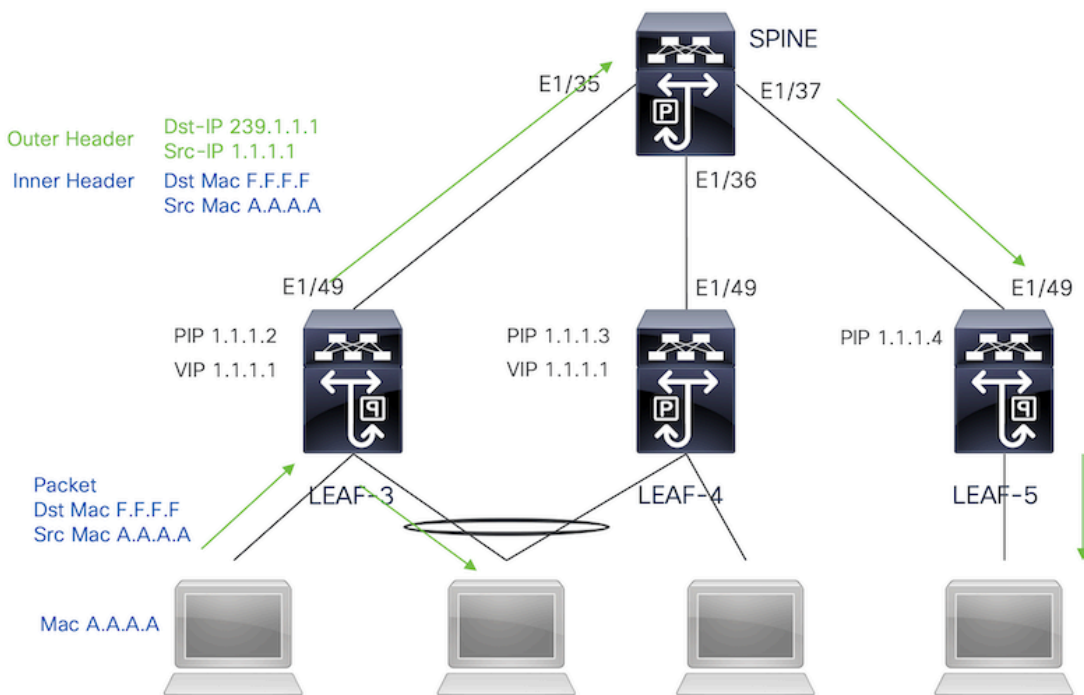


使用组播封装的广播、未知单播和组播流量

当nexus收到需要广播的数据包时，会生成该数据包的2个副本。

1. 数据包将发送到组播S，G条目中的所有OIF，包括本地接入端口
2. 到远程VPC对等设备

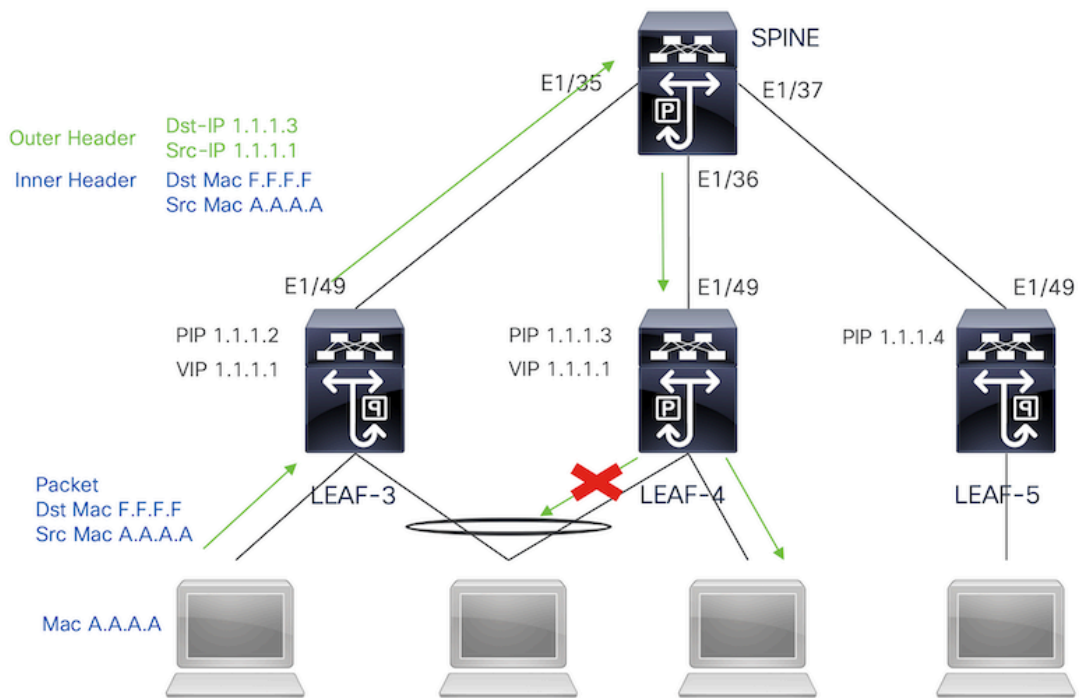
对于第一个副本，Nexus使用辅助IP地址的源IP和配置的组播组目标IP封装流量。



对于第二个副本，它将发送到远程VPC对等设备，源IP将是环回的辅助设备，目标IP是远程VPC对

等设备的PIP。

从主干收到数据包后，远程VTEP仅将数据包转发到孤立端口。

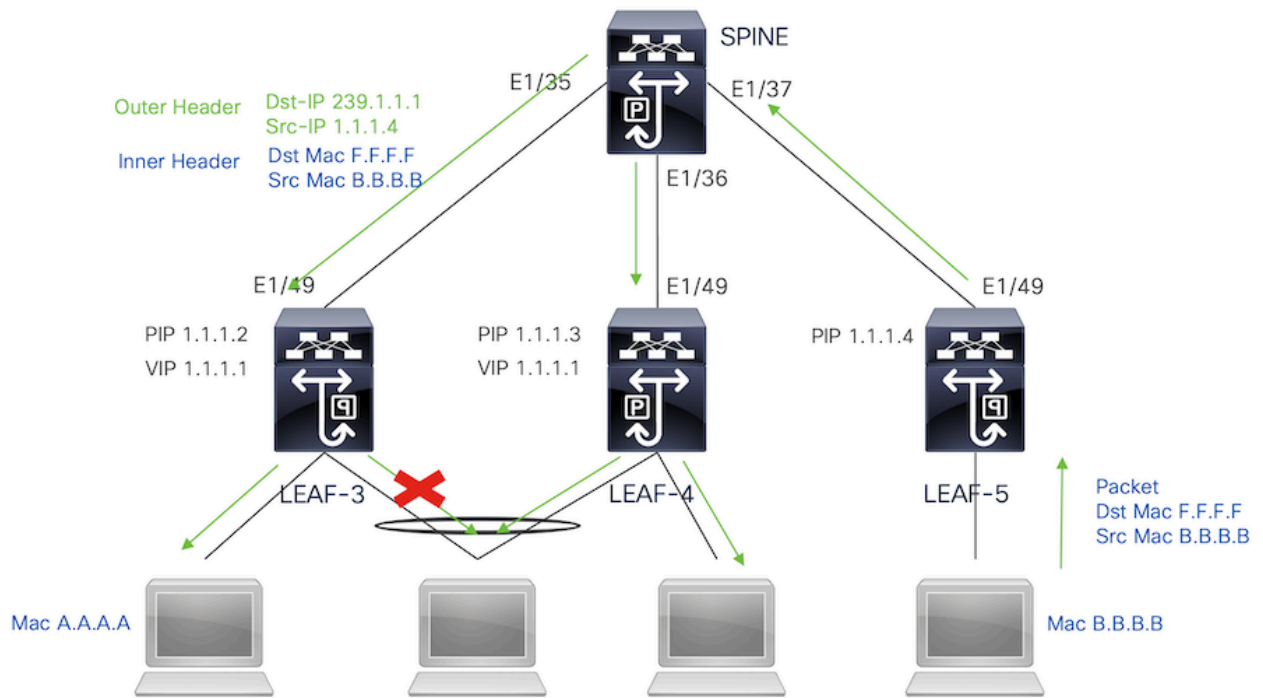


使用组播解封的广播、未知单播和组播流量

对于解封过程，数据包将到达两个VPC对等设备。只有一台VPC设备通过VPC端口通道转发流量。这由命令中显示的转发器决定。

```
module-1# show forwarding internal vpc-df-hash
```

```
VPC DF: FORWARDER
```



验证

要确保VPC已启动，请运行以下命令：

验证用于虚拟对等链路的IP地址的可达性。

```
LEAF-3# sh ip route 10.10.10.1
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'***' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.10.10.1/32, ubest/mbest: 1/0
    *via 192.168.120.1, Eth1/49, [110/3], 01:15:01, ospf-1, intra
```

```
LEAF-3# ping 10.10.10.1
PING 10.10.10.1 (10.10.10.1): 56 data bytes
64 bytes from 10.10.10.1: icmp_seq=0 ttl=253 time=0.898 ms
64 bytes from 10.10.10.1: icmp_seq=1 ttl=253 time=0.505 ms
64 bytes from 10.10.10.1: icmp_seq=2 ttl=253 time=0.433 ms
64 bytes from 10.10.10.1: icmp_seq=3 ttl=253 time=0.465 ms
64 bytes from 10.10.10.1: icmp_seq=4 ttl=253 time=0.558 ms
```

```
LEAF-3(config-if)# show vpc brief
```

Legend:

(*) - local vPC is down, forwarding via vPC peer-link

```
vPC domain id          : 1
Peer status            : peer adjacency formed ok <<<<
```

```

vPC keep-alive status           : peer is alive <<<<
Configuration consistency status : success
Per-vlan consistency status     : success
Type-2 consistency status       : success
vPC role                         : secondary
Number of vPCs configured       : 0
Peer Gateway                     : Disabled
Dual-active excluded VLANs      : -
Graceful Consistency Check      : Enabled
Auto-recovery status            : Disabled
Delay-restore status             : Timer is off.(timeout = 30s)
Delay-restore SVI status        : Timer is off.(timeout = 10s)
Delay-restore Orphan-port status : Timer is off.(timeout = 0s)
Operational Layer3 Peer-router  : Disabled
Virtual-peerlink mode           : Enabled <<<<<<<

```

vPC Peer-link status

```

-----
id  Port  Status  Active vlans
--  ---  -
1   Po1   up      1,10,50,600-604,608,610-611,614-618,638-639,
                                662-663,701-704

```

要检查VPC的角色，请运行命令：

```
LEAF-3(config-if)# sh 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             : d0:e0:42:e2:09:6f
vPC local role-priority          : 32667
vPC local config role-priority   : 32667
vPC peer system-mac             : 2c:4f:52:3f:46:df
vPC peer role-priority           : 32667
vPC peer config role-priority    : 32667

```

在对等链路端口通道中允许的所有VLAN必须映射到VNI，以防它们显示为不一致

```

LEAF-3(config-if)# show vpc virtual-peerlink vlan consistency
Following vlans are inconsistent
1 608 610 611 614 615 616 617 618 638 639 701 702 703 704

```

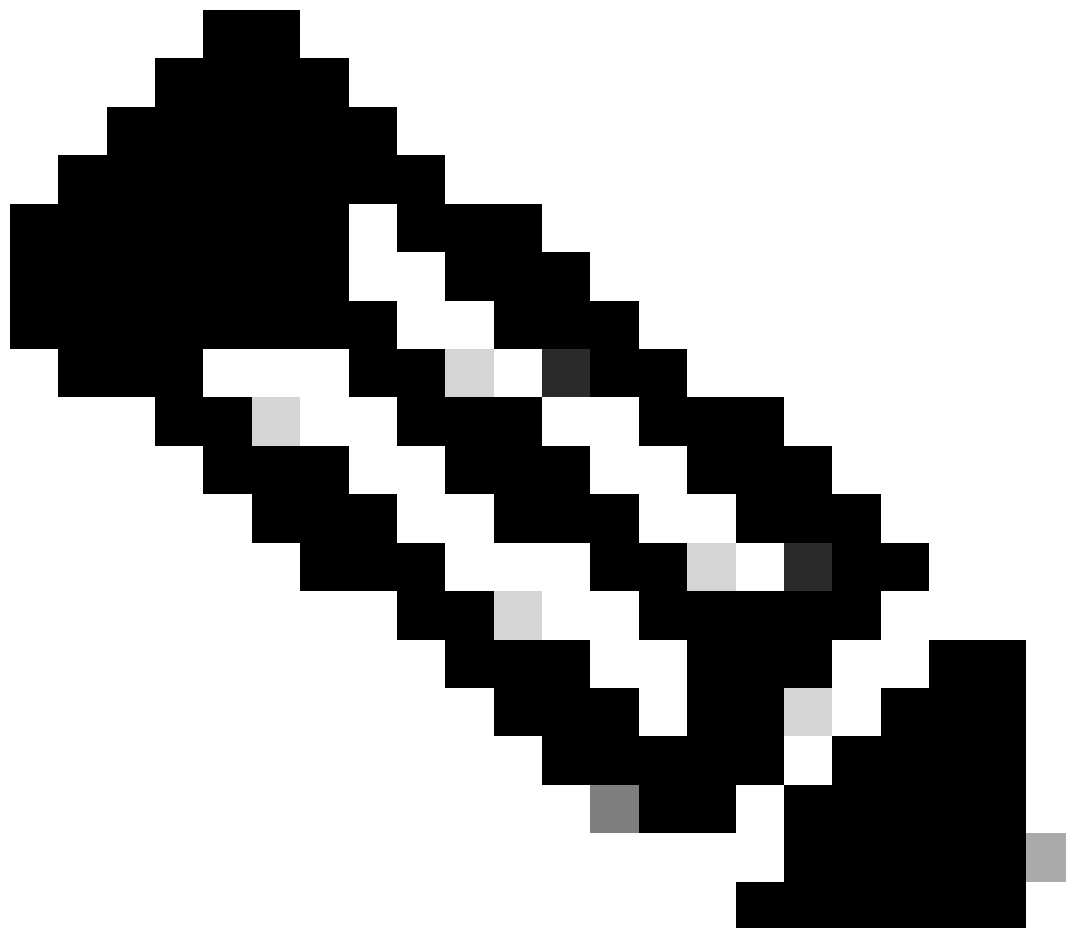
要确认上行链路的配置已正确编程，请运行以下命令：

```
LEAF-3(config-if)# show vpc fabric-ports
```

Number of Fabric port : 1
Number of Fabric port active : 1

Fabric	Ports	State

Ethernet	1/49	UP



注意：除非VPC已启动，否则NVE或与其关联的环回接口都将显示。

相关信息

- [思科技术支持和下载](#)

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）。