# 使用vPC的高级VXLAN:L2VNI和L3VNI的配置与验证

## 目录


简介

先决条件

- 要求
- 使用的组件

背景信息

配置

- 网络图
- 配置

确认


## 简介

本文档介绍如何使用Nexus 9Kv交换机设置实验，该实验使用具有虚拟端口通道(vPC)的高级虚拟可扩展局域网(VXLAN)。

## 先决条件

### 要求

Cisco 建议您了解以下主题：

- 了解路由和交换以及多协议标签交换(MPLS)技术
- 具有组播路由原理(例如交汇点(RP)和平台无关组播(PIM))的经验
- 了解边界网关协议(BGP)地址系列指示器(AFI)/后续地址系列指示器(SAFI)

### 使用的组件

本文档不限于特定的软件和硬件版本。

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

## 背景信息

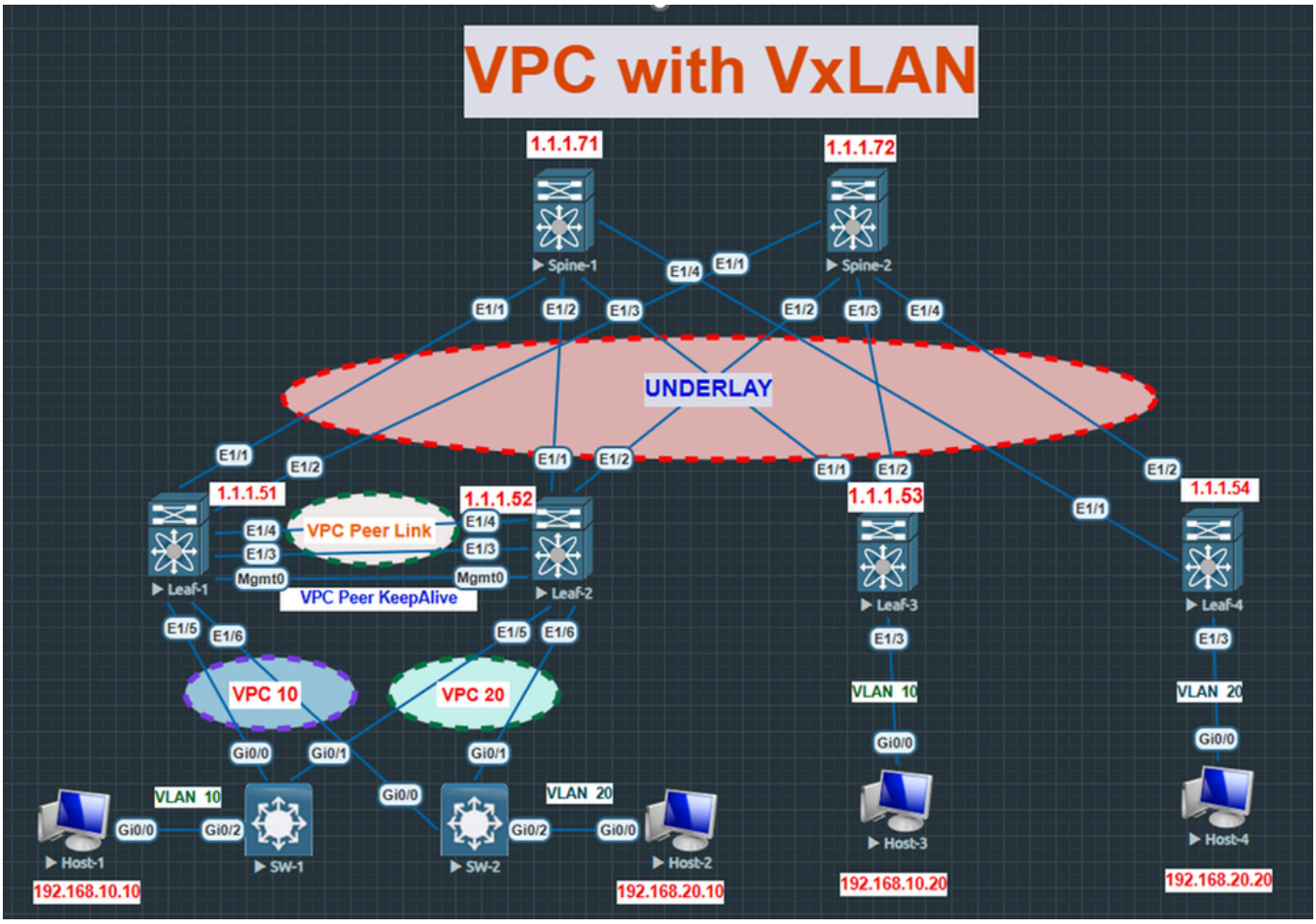
本文档还提供有关部署实验以及检验配置和操作的指导。

在本实验中，枝叶和主干均使用带有Nexus 9000V交换机的EveNg。

| | |
|---|---|
| 虚拟隧道终端(VTEP) | LEAF1、LEAF2、LEAF3、LEAF4 |
| vPC | LEAF1和LEAF2 |
| LEAF1主要和辅助环回IP | Loopback0 - 1.1.1.51、Loopback1 - 10.1.1.100 |
| LEAF2主环回和辅助环回IP | Loopback0 - 1.1.1.52、Loopback1 - 10.1.1.100 |
| LEAF3环回IP | 1.1.1.53 |
| LEAF4环回IP | 1.1.1.54 |
| SPINE1环回和任播RP | Loopback0 - 1.1.1.71、Loopback1 - 10.1.2.10（任播RP） |
| SPINE2环回和任播RP | Loopback0 - 1.1.1.72、Loopback1 - 10.1.2.10（任播RP） |
| 主机1 | 192.168.10.10(000.000.aaaa)(VLAN 10) |
| 主机2 | 192.168.20.10(0000.000.bbb)(VLAN 20) |
| 主机3 | 192.168.10.20(0000.000.cccc)(VLAN 10) |
| 主机4 | 192.168.20.20(0000.000.ddd)(VLAN 20) |
| VLAN 10 | L2VNI设100010 |
| VLAN 20 | L2VNI设100020 |
| VLAN 500 | L3VNI设50000 |

# 配置

## 网络图




## 配置

- 底层和PIM邻居已建立。

枝叶交换机：

```
feature ospf

router ospf UNDERLAY
  log-adjacency-changes

interface loopback0
  ip router ospf UNDERLAY area 0.0.0.0

interface Ethernet1/1
  ip ospf cost 4
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0

interface Ethernet1/2
  ip ospf cost 4
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
```

在枝叶交换机上启用开放最短路径优先(OSPF)

```
feature pim

ip pim rp-address 10.1.2.10 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8

vrf context TENANT1
  ip pim ssm range 232.0.0.0/8

interface Vlan10
  ip pim sparse-mode

interface Vlan20
  ip pim sparse-mode

interface loopback0
  ip pim sparse-mode

interface Ethernet1/1
  ip pim sparse-mode

interface Ethernet1/2
  ip pim sparse-mode
```

在枝叶交换机上启用PIM

```
LEAF-1# show ip ospf neighbors
 OSPF Process ID UNDERLAY VRF default
 Total number of neighbors: 2
 Neighbor ID     Pri State            Up Time  Address         Interface
 1.1.1.71          1 FULL/ -          04:32:03 192.168.11.1    Eth1/1
 1.1.1.72          1 FULL/ -          04:17:47 192.168.21.2    Eth1/2
LEAF-1# sh ip pim neighbor
PIM Neighbor Status for VRF "default"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD    ECMP Redirect
                                                         Priority Capable State     Capable
192.168.11.1    Ethernet1/1          04:32:14  00:01:30  1        yes     n/a     no
192.168.21.2    Ethernet1/2          04:17:58  00:01:44  1        yes     n/a     no
LEAF-1#
```

OSPF 邻居

主干交换机：

```
feature pim

ip pim rp-address 10.1.2.10 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
ip pim anycast-rp 10.1.2.10 1.1.1.71
ip pim anycast-rp 10.1.2.10 1.1.1.72
```

在主干交换机上启用PIM

- 底层和PIM邻居已建立。
- 两个主干交换机将是整个组播组224.0.0.0/4的相同任播RP。
- 枝叶和主干交换机之间的接口上的最大传输单位(MTU)设置为9000/9216。

首先，让我们在Leaf1和Leaf2之间设置vPC。

步骤1.启用vPC功能和域。

- 启用功能vPC和链路汇聚控制协议(LACP)。
- 配置vPC域。
- 管理0接口用作对等保持连接链路，Eth1/3和Eth1/4将成为vPC对等链路（端口通道1）的一部分。
- 确保将peer-switch命令配置为与下行交换机共享一个公用MAC地址。

```
feature lacp
feature vpc
```

在枝叶交换机上启用功能

```
LEAF-1# sh run vpc

!Command: show running-config vpc
!Running configuration last done at: Sat Dec 28 07:17:18 2024
!Time: Sat Dec 28 07:39:48 2024

version 7.0(3)I7(9) Bios:version
feature vpc

vpc domain 1
  peer-switch
  role priority 100
  peer-keepalive destination 192.168.0.52
  peer-gateway

interface port-channel1
  vpc peer-link
```

在枝叶交换机1上启用vPC

```
LEAF-2# sh run vpc

!Command: show running-config vpc
!Running configuration last done at: Sat Dec 28 07:17:14 2024
!Time: Sat Dec 28 07:40:20 2024

version 7.0(3)I7(9) Bios:version
feature vpc

vpc domain 1
  peer-switch
  role priority 200
  peer-keepalive destination 192.168.0.51
  peer-gateway

interface port-channel1
  vpc peer-link
```

在枝叶交换机2上启用vPC

步骤2.端口成员分配。

- 将端口成员分配给通道组并将它们包括在vPC中。在本例中，使用了两个vPC。vPC 20和vPC 10。

```
LEAF-1# sh run int port-channel 10, port-channel 20 membership

!Command: show running-config interface port-channel10, port-channel20 membership
!Running configuration last done at: Sat Dec 28 07:17:18 2024
!Time: Sat Dec 28 07:42:44 2024

version 7.0(3)I7(9) Bios:version

interface port-channel10
  switchport mode trunk
  vpc 10

interface Ethernet1/5

  switchport mode trunk
  channel-group 10 mode active

interface port-channel20
  switchport mode trunk
  vpc 20

interface Ethernet1/6

  switchport mode trunk
  channel-group 20 mode active

LEAF-1#
```

在枝叶交换机1上分配端口通道

```
LEAF-2# sh run int port-channel 10, port-channel 20 membership

!Command: show running-config interface port-channel10, port-channel20 membership
!Running configuration last done at: Sat Dec 28 07:17:14 2024
!Time: Sat Dec 28 07:43:16 2024

version 7.0(3)I7(9) Bios:version

interface port-channel10
  switchport mode trunk
  vpc 10

interface Ethernet1/5

  switchport mode trunk
  channel-group 10 mode active

interface port-channel20
  switchport mode trunk
  vpc 20

interface Ethernet1/6

  switchport mode trunk
  channel-group 20 mode active

LEAF-2#
```

在枝叶交换机2上分配端口通道

- 此处创建了一个vPC，对等体开始交换keepalive消息以验证可用性。

```
LEAF-1# show vpc
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                          : primary
Number of vPCs configured         : 2
Peer Gateway                      : Enabled
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)
Operational Layer3 Peer-router    : Disabled

vPC Peer-link status
---------------------------------------------------------------------
id    Port   Status Active vlans
--    ----   ------ --------------------------------------------------
1     Po1    up     1,10,20,500

vPC status
----------------------------------------------------------------------------
Id    Port          Status Consistency Reason                Active vlans
--    ------------  ------ ----------- ------                ---------------
10    Po10          up     success     success               1,10,20,500

20    Po20          up     success     success               1,10,20,500


Please check "show vpc consistency-parameters vpc <vpc-num>" for the
consistency reason of down vpc and for type-2 consistency reasons for
any vpc.

LEAF-1#
```

枝叶交换机1上的vPC状态

```
LEAF-2# sh vpc
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         : 2
Peer Gateway                      : Enabled
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)
Operational Layer3 Peer-router    : Disabled

vPC Peer-link status
---------------------------------------------------------------------
id     Port   Status Active vlans
--     ----   ------ -----------------------------------------------
1      Po1    up     1,10,20,500

vPC status
----------------------------------------------------------------------------
Id     Port          Status Consistency Reason                Active vlans
--     ------------  ------ ----------- ------                ---------------
10     Po10          up     success     success               1,10,20,500

20     Po20          up     success     success               1,10,20,500


Please check "show vpc consistency-parameters vpc <vpc-num>" for the
consistency reason of down vpc and for type-2 consistency reasons for
any vpc.

LEAF-2#
```

枝叶交换机2上的vPC状态

- VLAN 10、20、500已配置并通过vPC成员端口和vPC对等链路。

步骤3.配置辅助IP地址。

- 当vPC包含在VXLAN交换矩阵中时，两个vPC VTEP对等体开始使用虚拟IP(VIP)地址而不是物理IP地址(PIP)作为源地址。 这也意味着当BGP以太网VPN(EVPN)默认通告路由类型2（MAC/IP通告）和5（IP前缀路由）时，VIP用作下一跳。本例中的Loopback 0接口设置有两个IP地址：10.1.1.100/32(VIP)作为辅助IP，1.1.1.51/32(PIP)作为主要IP。
- 在这里，一个公用IP地址配置为loopback 0接口下的一个辅助地址。

```
LEAF-1# sh run int l0

!Command: show running-config interface loopback0
!Running configuration last done at: Sat Dec 28 07:51:58 2024
!Time: Sat Dec 28 07:55:26 2024

version 7.0(3)I7(9) Bios:version

interface loopback0
  ip address 1.1.1.51/32
  ip address 10.1.1.100/32 secondary
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

LEAF-1#
```

枝叶交换机1上的辅助IP

```
LEAF-2# sh run int l0

!Command: show running-config interface loopback0
!Running configuration last done at: Sat Dec 28 07:52:05 2024
!Time: Sat Dec 28 07:55:37 2024

version 7.0(3)I7(9) Bios:version

interface loopback0
  ip address 1.1.1.52/32
  ip address 10.1.1.100/32 secondary
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

LEAF-2#
```

枝叶交换机2上的辅助IP

步骤4.启用VXLAN和相关功能。

- 网络虚拟化(nV)重叠 — 启用VXLAN
- 功能nV重叠EVPN — 启用EVPN控制平面
- 功能交换矩阵转发 — 启用主机移动管理器
- 基于功能虚拟网络(VN)网段VLAN — 启用基于VLAN的VXLAN

```
LEAF-1# sh run | sec "feature|nv over"
nv overlay evpn
feature ospf
feature bgp
feature pim
feature fabric forwarding
feature interface-vlan
feature vn-segment-vlan-based
feature lacp
feature vpc
feature nv overlay
LEAF-1#
```

枝叶交换机上的功能

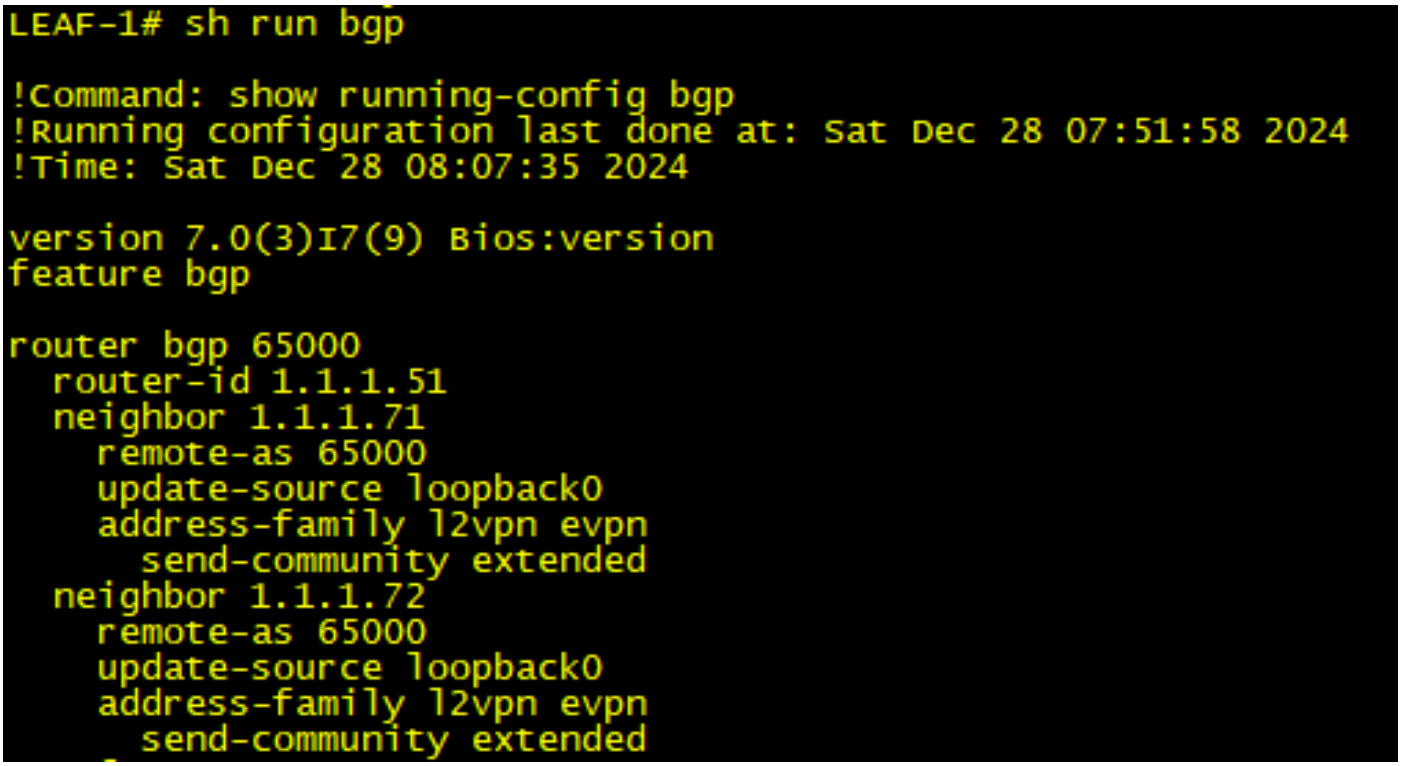
```
SPINE-1# sh run | sec "feature|nv over"
nv overlay evpn
feature ospf
feature bgp
feature pim
feature nv overlay
SPINE-1#
```

主干交换机上的功能

- 由于主干不需要了解客户端的VLAN信息，因此无需启用VN网段和交换矩阵功能。

步骤5.启用BGP邻居关系。

- 必须启用枝叶和主干交换机之间的BGP。主干将在实验中用作路由反射器。
- 虽然配置路由反射器(RR)是可选操作，但出于可扩展性的考虑，Cisco推荐使用RR。




在枝叶交换机上启用BGP

```
SPINE-1# sh run bgp

!Command: show running-config bgp
!Running configuration last done at: Sat Dec 28 07:16:33 2024
!Time: Sat Dec 28 08:08:21 2024

version 7.0(3)I7(9) Bios:version
feature bgp

router bgp 65000
  router-id 1.1.1.71
  neighbor 1.1.1.51
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
      route-reflector-client
  neighbor 1.1.1.52
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
      route-reflector-client
  neighbor 1.1.1.53
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
      route-reflector-client
  neighbor 1.1.1.54
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
      route-reflector-client


SPINE-1#
```

在主干交换机上启用BGP

```
LEAF-1# show bgp l2vpn evpn summary
BGP summary information for VRF default, address family L2VPN EVPN
BGP router identifier 1.1.1.51, local AS number 65000
BGP table version is 62, L2VPN EVPN config peers 2, capable peers 2
10 network entries and 13 paths using 2228 bytes of memory
BGP attribute entries [10/1600], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [4/16]

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
1.1.1.71        4 65000     146     121       62    0    0 01:45:52 3
1.1.1.72        4 65000     141     114       62    0    0 01:39:12 3
LEAF-1#
```

枝叶交换机上的BGP状态

```
SPINE-1# show bgp l2vpn evpn summary
BGP summary information for VRF default, address family L2VPN EVPN
BGP router identifier 1.1.1.71, local AS number 65000
BGP table version is 98, L2VPN EVPN config peers 4, capable peers 4
9 network entries and 9 paths using 2124 bytes of memory
BGP attribute entries [7/1120], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [0/0]

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
1.1.1.51        4 65000     147     124       98    0    0 01:46:29 2
1.1.1.52        4 65000     147     124       98    0    0 01:46:30 2
1.1.1.53        4 65000     128     155       98    0    0 02:01:15 1
1.1.1.54        4 65000     191     225       98    0    0 03:03:08 2
SPINE-1#
```

主干交换机上的BGP状态

步骤6.在枝叶交换机上启用VRF情景。VRF分离客户流量并促进两个不同的L2VNI之间通过L3VNI进行通信。

- 在VRF TENANT1下50000配L3VNI配置。

```
vrf context TENANT1
  vni 50000
  ip pim ssm range 232.0.0.0/8
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
```

L3VNI分配

步骤7.网络虚拟接口(NVE)、VXLAN标识符(VNI)和VLAN配置。

- 设置NVE接口，使用Loopback 0作为源。定义每个VNI的组播组，其中将传送第2层广播、未知单播和组播(BUM)流量，然后将VNI 100010和100020 ID附加到NVE接口。VXLAN报头包含VNI用于标识它所属的VXLAN网段的信息。
- L3VNI实例50000接到VRF实例(当将其发送到主干交换机时，VNI 50000已附加到VRF表中)。
- host-reachability protocol BGP命令在VXLAN隧道中激活EVPN地址系列，这意味着MAC地址和IP地址是通过控制平面中的BGP协议获知的，而不是在数据平面中的。
- 在NVE接口下配置suppress-arp。
- 将第2层和第3层VLAN连接到相关VNI。

抑制地址解析协议(ARP):

多协议(MP)-BGP EVPN控制平面提供称为ARP抑制的改进，以减少ARP请求中的广播流量带来的网络泛洪。每个VNI的VTEP都会为已知IP主机保留一个ARP抑制缓存表，并在该VNI启用ARP抑制时在VNI网段中保留与它们对应的MAC地址。其本地VTEP会拦截ARP请求，并在ARP抑制缓存表中查找ARP解析的IP地址，只要该VNI中的终端主机提交针对另一个终端主机IP地址的ARP请求。

本地VTEP代表远程终端主机在发现匹配项时发送ARP响应。然后，ARP响应将为本地主机提供远程主机的MAC地址。如果本地VTEP的ARP抑制表中没有ARP解析的IP地址，则ARP请求会泛洪到VNI中的其他VTEP。对于发送到静默网络主机的第一个ARP请求，可能会发生此ARP泛洪。

```
LEAF-1# sh run interface nve 1

!Command: show running-config interface nve1
!Running configuration last done at: Sat Dec 28 07:51:58 2024
!Time: Sat Dec 28 08:44:44 2024

version 7.0(3)I7(9) Bios:version

interface nve1
  no shutdown
  host-reachability protocol bgp
  source-interface loopback0
  member vni 50000 associate-vrf
  member vni 100010
    suppress-arp
    mcast-group 239.0.0.10
  member vni 100020
    suppress-arp
    mcast-group 239.0.0.20

LEAF-1#
```

NVE接口

```
LEAF-1# sh run vlan

!Command: show running-config vlan
!Running configuration last done at: Sat Dec 28 07:51:58 2024
!Time: Sat Dec 28 08:46:44 2024

version 7.0(3)I7(9) Bios:version
vlan 1,10,20,500
vlan 10
  vn-segment 100010
vlan 20
  vn-segment 100020
vlan 500
  vn-segment 50000


LEAF-1#
```

VLAN到VN网段的映射

- 通过向Spine发送PIM加入消息，NVE接口在启动时将分别加入组播组239.0.0.10和239.0.0.20。
- 您还可以在映像中看到其他(S、G)表(1.1.1.54、239.0.0.20)和(10.1.1.100、239.0.0.10/239.0.0.20)，这些表已经注册到来自不同枝叶交换机的主干。

```
LEAF-1# sh ip mroute summary
IP Multicast Routing Table for VRF "default"
Route Statistics unavailable - only liveness detected

Total number of routes: 7
Total number of (*,G) routes: 2
Total number of (S,G) routes: 4
Total number of (*,G-prefix) routes: 1
Group count: 2, rough average sources per group: 2.0

Group: 232.0.0.0/8, Source count: 0
Source          packets      bytes          aps   pps       bit-rate     oifs
(*,G)           0            0              0     0         0.000   bps  0

Group: 239.0.0.10/32, Source count: 2
Source          packets      bytes          aps   pps       bit-rate     oifs
(*,G)           1            100            100   0         0.000   bps  1
1.1.1.53        48           4644           96    0         78.267  bps  1
10.1.1.100      1124         113514         100   0         131.467 bps  1

Group: 239.0.0.20/32, Source count: 2
Source          packets      bytes          aps   pps       bit-rate     oifs
(*,G)           1            100            100   0         0.000   bps  1
1.1.1.54        51           4944           96    0         63.200  bps  1
10.1.1.100      1116         112729         101   0         70.667  bps  1
LEAF-1#
```

Mroute表

步骤8.启用EVPN实例。

- 为BGP下的EVPN和VRF启用EVPN实例以及地址系列。

```
LEAF-1# sh run bgp

!Command: show running-config bgp
!Running configuration last done at: Sat Dec 28 09:22:19 2024
!Time: Sat Dec 28 09:43:07 2024

version 7.0(3)I7(9) Bios:version
feature bgp

router bgp 65000
  router-id 1.1.1.51
  neighbor 1.1.1.71
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
  neighbor 1.1.1.72
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community extended
  vrf TENANT1
    address-family ipv4 unicast
      redistribute direct route-map REDIST
evpn
  vni 100010 l2
    rd auto
    route-target import auto
    route-target export auto
  vni 100020 l2
    rd auto
    route-target import auto
    route-target export auto
vrf context TENANT1
```

- route-map REDIST的唯一用途是允许所有内容。
- 使用redistribute direct命令，将连接的VRF感知路由提升为MP-BGP（第5类路由）。
- 上面显示的EVPN配置与BGP用于通告MAC路由（类型2路由）的network语句相同。

步骤9.在VRF下为终端主机配置每个VLAN的交换机虚拟接口(SVI)。

- 在每个枝叶交换机上，SVI配置为本地配置的VLAN，一个SVI配置为L3VNI VLAN，以实现对称路由信息库(RIB)。

对称肋：

- 当终端主机将数据包发送到不同的网络并接收至枝叶交换机时，它将首先被处理为L2VNI，然后使用VRF将其放置到L3VNI并发送到远程枝叶。
- 远程枝叶首先使用路由接收VRF表中的数据包，然后桥接到L2VNI并将其发送到终端主机。
- 通过这种方式，可以实现对称路由(B-R-R-B)。

```
LEAF-1# sh run interface vlan 10,vlan 20,vlan 500

!Command: show running-config interface Vlan10, Vlan20, Vlan500
!Running configuration last done at: Sat Dec 28 09:22:19 2024
!Time: Sat Dec 28 10:00:26 2024

version 7.0(3)I7(9) Bios:version

interface Vlan10
  no shutdown
  mtu 9216
  vrf member TENANT1
  no ip redirects
  ip address 192.168.10.254/24
  no ipv6 redirects
  ip pim sparse-mode
  fabric forwarding mode anycast-gateway

interface Vlan20
  no shutdown
  mtu 9216
  vrf member TENANT1
  no ip redirects
  ip address 192.168.20.254/24
  no ipv6 redirects
  ip pim sparse-mode
  fabric forwarding mode anycast-gateway

interface Vlan500
  no shutdown
  vrf member TENANT1
  no ip redirects
  ip forward
  no ipv6 redirects

LEAF-1#
```

- VLAN 500下的IP forward命令用于为所有VXLAN启用第3层转发。无需配置IP地址，因为它只处理从L2VNI表到L3VNI表的数据包。

```
LEAF-1# show bgp vpnv4 unicast vrf TENANT1
BGP routing table information for VRF default, address family VPNv4 Unicast
BGP table version is 15, Local Router ID is 1.1.1.51
Status: s-suppressed, x-deleted, S-stale, d-dampened, h-history, *-valid, >-best
Path type: i-internal, e-external, c-confed, l-local, a-aggregate, r-redist, I-injected
Origin codes: i - IGP, e - EGP, ? - incomplete, | - multipath, & - backup

   Network            Next Hop            Metric     LocPrf     Weight Path
Route Distinguisher: 1.1.1.51:3    (VRF TENANT1)
*>r192.168.10.0/24    0.0.0.0                  0        100      32768 ?
*>i192.168.10.20/32   1.1.1.53                          100          0 i
*>r192.168.20.0/24    0.0.0.0                  0        100      32768 ?
*>i192.168.20.20/32   1.1.1.54                          100          0 i

LEAF-1#
```

了解VRF TENANT1的BGP VPNv4路由

- 每个VLAN的IP地址对于所有枝叶交换机上的所有SVI都是通用的。这称为任播IP，它用于移动管理，其中终端可以无缝地与其他主机通信，而不会造成任何中断。

步骤10.为终端主机启用交换矩阵转发任播网关MAC。

- 它确保无缝的第3层网关冗余和连接到交换矩阵的设备的优化转发。
- 任播网关MAC地址是全局一致的MAC地址，用于交换矩阵中的所有第3层网关。
- 此概念与第一跳冗余协议(FHRP)中使用的概念相同，在该协议中，每个组都分配一个虚拟MAC。

```
LEAF-1# show running-config fabric forwarding

!Command: show running-config fabric forwarding
!Running configuration last done at: Sat Dec 28 09:22:19 2024
!Time: Sat Dec 28 10:08:08 2024

version 7.0(3)I7(9) Bios:version
nv overlay evpn
feature fabric forwarding

fabric forwarding anycast-gateway-mac 0000.1234.5678


interface Vlan10
  fabric forwarding mode anycast-gateway

interface Vlan20
  fabric forwarding mode anycast-gateway

LEAF-1#
```

启用交换矩阵转发

步骤11.对成员端口启用接入/中继VLAN。

vPC交换机：

```
LEAF-1# sh run int po10 membership

!Command: show running-config interface port-channel10 membership
!Running configuration last done at: Sat Dec 28 09:22:19 2024
!Time: Sat Dec 28 10:13:19 2024

version 7.0(3)I7(9) Bios:version

interface port-channel10
  switchport mode trunk
  vpc 10

interface Ethernet1/5

  switchport mode trunk
  channel-group 10 mode active

LEAF-1#
```

启用到vPC成员接口的中继端口

非vPC交换机：

```
LEAF-3# show running-config interface e1/3

!Command: show running-config interface Ethernet1/3
!Running configuration last done at: Sat Dec 28 09:28:18 2024
!Time: Sat Dec 28 10:14:42 2024

version 7.0(3)I7(9) Bios:version

interface Ethernet1/3
  switchport access vlan 10
  spanning-tree port type edge

LEAF-3#
```

启用连接到非vPC成员接口的中继端口

# 确认

- 检查ARP和MAC地址表。

```
LEAF-1# sh ip arp vrf TENANT1

Flags: * - Adjacencies learnt on non-active FHRP router
       + - Adjacencies synced via CFSoE
       # - Adjacencies Throttled for Glean
       CP - Added via L2RIB, Control plane Adjacencies
       PS - Added via L2RIB, Peer Sync
       RO - Re-Originated Peer Sync Entry
       D - Static Adjacencies attached to down interface

IP ARP Table for context TENANT1
Total number of entries: 2
Address         Age       MAC Address     Interface       Flags
192.168.20.10   00:00:36  0000.0000.bbbb  Vlan20
192.168.10.10   00:04:19  0000.0000.aaaa  Vlan10
LEAF-1#  sh ip arp suppression-cache deta

Flags: + - Adjacencies synced via CFSoE
       L - Local Adjacency
       R - Remote Adjacency
       L2 - Learnt over L2 interface
       PS - Added via L2RIB, Peer Sync
       RO - Dervied from L2RIB Peer Sync Entry

Ip Address      Age      Mac Address    Vlan Physical-ifindex    Flags    Remote Vtep Addrs

192.168.10.10   00:04:33 0000.0000.aaaa   10 port-channel10      L
192.168.10.20   00:55:53 0000.0000.cccc   10 (null)              R        1.1.1.53
192.168.20.10   00:00:50 0000.0000.bbbb   20 port-channel20      L
192.168.20.20   03:26:04 0000.0000.dddd   20 (null)              R        1.1.1.54
LEAF-1#
```

枝叶交换机1上的ARP和MAC表

```
LEAF-2# show ip arp vrf TENANT1

Flags: * - Adjacencies learnt on non-active FHRP router
       + - Adjacencies synced via CFSoE
       # - Adjacencies Throttled for Glean
       CP - Added via L2RIB, Control plane Adjacencies
       PS - Added via L2RIB, Peer Sync
       RO - Re-Originated Peer Sync Entry
       D - Static Adjacencies attached to down interface

IP ARP Table for context TENANT1
Total number of entries: 2
Address         Age       MAC Address     Interface       Flags
192.168.20.10   00:01:28  0000.0000.bbbb  vlan20          +
192.168.10.10   00:00:11  0000.0000.aaaa  vlan10          +
LEAF-2#
```

枝叶交换机2上的ARP和MAC表

- 两个对等体都维护ARP条目。
- 检查网络虚拟接口(NVI)状态。

vPC交换机：

```
LEAF-1# show nve peers
Interface Peer-IP          State LearnType Uptime   Router-Mac
--------- ---------------  ----- --------- -------- -----------------
nve1      1.1.1.53         Up    CP        01:09:04 5000.0003.0007
nve1      1.1.1.54         Up    CP        03:39:16 5000.0004.0007

LEAF-1# show nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      50000    n/a               Up    CP   L3 [TENANT1]
nve1      100010   239.0.0.10        Up    CP   L2 [10]            SA
nve1      100020   239.0.0.20        Up    CP   L2 [20]            SA

LEAF-1#
```

vPC交换机上的NVE对等体

非vPC交换机：

```
LEAF-3# show nve peers
Interface Peer-IP          State LearnType Uptime   Router-Mac
--------- ---------------  ----- --------- -------- -----------------
nve1      1.1.1.54         Up    CP        01:14:00 5000.0004.0007
nve1      10.1.1.100       Up    CP        01:14:16 5000.0001.0007

LEAF-3#
```

非vPC交换机上的NVE对等体

- 在这里，您会发现对等IP是10.1.1.100，而不是主环回IP地址，因此返回的数据包将针对该IP路由到任何vPC交换机。
- 检查BGP EVPN路由。

```
LEAF-1# show l2route evpn mac-ip all
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv(D):Del Pending (S):Stale (C):clear
(Ps):Peer Sync (Ro):Re-Originated
Topology    Mac Address    Prod   Flags         Seq No     Host IP         Next-Hops
----------- -------------- ------ ------------- ---------- --------------- --------------
10          0000.0000.aaaa HMM    --            0          192.168.10.10   Local
10          0000.0000.cccc BGP    --            0          192.168.10.20   1.1.1.53
20          0000.0000.bbbb HMM    --            0          192.168.20.10   Local
20          0000.0000.dddd BGP    --            0          192.168.20.20   1.1.1.54
LEAF-1#
```

BGP l2route EVPN MAC-IP

```
LEAF-1# show l2route evpn mac all

Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending
(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override
(Pf):Permanently-Frozen

Topology     Mac Address    Prod   Flags         Seq No     Next-Hops
-----------  -------------  -----  ------------  ---------  ----------------
10           0000.0000.aaaa Local  L,            0          Po10
10           0000.0000.cccc BGP    Spl           0          1.1.1.53
20           0000.0000.bbbb Local  L,            0          Po20
20           0000.0000.dddd BGP    SplRcv        0          1.1.1.54
500          5000.0003.0007 VXLAN  Rmac          0          1.1.1.53
500          5000.0004.0007 VXLAN  Rmac          0          1.1.1.54
LEAF-1#
```

BGP l2route EVPN MAC

```
LEAF-1# show bgp l2vpn evpn summary
BGP summary information for VRF default, address family L2VPN EVPN
BGP router identifier 1.1.1.51, local AS number 65000
BGP table version is 134, L2VPN EVPN config peers 2, capable peers 2
12 network entries and 15 paths using 2568 bytes of memory
BGP attribute entries [12/1920], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [4/16]

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
1.1.1.71        4 65000     312     263      134    0    0 03:46:01 3
1.1.1.72        4 65000     307     256      134    0    0 03:39:21 3
LEAF-1#
```

BGP EVPN摘要

```
LEAF-1# show bgp l2vpn evpn
BGP routing table information for VRF default, address family L2VPN EVPN
BGP table version is 146, Local Router ID is 1.1.1.51
Status: s-suppressed, x-deleted, S-stale, d-dampened, h-history, *-valid, >-best
Path type: i-internal, e-external, c-confed, l-local, a-aggregate, r-redist, I-injected
Origin codes: i - IGP, e - EGP, ? - incomplete, | - multipath, & - backup

   Network            Next Hop            Metric     LocPrf     Weight Path
Route Distinguisher: 1.1.1.51:32777    (L2VNI 100010)
*>l[2]:[0]:[0]:[48]:[0000.0000.aaaa]:[0]:[0.0.0.0]/216
                      10.1.1.100                             100      32768 i
*>l[2]:[0]:[0]:[48]:[0000.0000.aaaa]:[32]:[192.168.10.10]/272
                      10.1.1.100                             100      32768 i
*>i[2]:[0]:[0]:[48]:[0000.0000.cccc]:[32]:[192.168.10.20]/272
                      1.1.1.53                               100          0 i

Route Distinguisher: 1.1.1.51:32787    (L2VNI 100020)
*>l[2]:[0]:[0]:[48]:[0000.0000.bbbb]:[0]:[0.0.0.0]/216
                      10.1.1.100                             100      32768 i
*>i[2]:[0]:[0]:[48]:[0000.0000.dddd]:[0]:[0.0.0.0]/216
                      1.1.1.54                               100          0 i
*>l[2]:[0]:[0]:[48]:[0000.0000.bbbb]:[32]:[192.168.20.10]/272
                      10.1.1.100                             100      32768 i
*>i[2]:[0]:[0]:[48]:[0000.0000.dddd]:[32]:[192.168.20.20]/272
                      1.1.1.54                               100          0 i

Route Distinguisher: 1.1.1.53:32777
*>i[2]:[0]:[0]:[48]:[0000.0000.cccc]:[32]:[192.168.10.20]/272
                      1.1.1.53                               100          0 i
* i                   1.1.1.53                               100          0 i

Route Distinguisher: 1.1.1.54:32787
* i[2]:[0]:[0]:[48]:[0000.0000.dddd]:[0]:[0.0.0.0]/216
                      1.1.1.54                               100          0 i
*>i                   1.1.1.54                               100          0 i
* i[2]:[0]:[0]:[48]:[0000.0000.dddd]:[32]:[192.168.20.20]/272
                      1.1.1.54                               100          0 i
*>i                   1.1.1.54                               100          0 i

Route Distinguisher: 1.1.1.51:3    (L3VNI 50000)
*>i[2]:[0]:[0]:[48]:[0000.0000.cccc]:[32]:[192.168.10.20]/272
                      1.1.1.53                               100          0 i
*>i[2]:[0]:[0]:[48]:[0000.0000.dddd]:[32]:[192.168.20.20]/272
                      1.1.1.54                               100          0 i

LEAF-1#
```

BGP EVPN路由

- 常见的问题是枝叶交换机如何获取远程主机的MAC条目。此过程由无故ARP实现。网络端口激活后，它会立即发送ARP请求以验证IP地址的唯一性。然后，每台枝叶交换机记录MAC地址并将其包含在BGP更新数据包中。这允许其他枝叶交换机相应地更新其各自的MAC地址表。但可能出现终端主机不生成无故ARP（静默主机）的情况，在这种情况下，ARP请求将广播到枝叶，由于它是广播请求，枝叶交换机将生成组播请求到特定VNI的相应组。在本例中，它是239.0.0.10和239.0.0.20。

- 允许从同一VNI中的Host-1 ping Host-3，查看捕获。

```
HOST-1#ping 192.168.10.20 rep 2
Type escape sequence to abort.
Sending 2, 100-byte ICMP Echos to 192.168.10.20, timeout is 2 seconds:
!!
Success rate is 100 percent (2/2), round-trip min/avg/max = 11/11/12 ms
HOST-1#
```

从HOST-1 ping HOST-3

VXLAN上的互联网控制消息协议(ICMP)数据包：

```
> Frame 213: 164 bytes on wire (1312 bits), 164 bytes captured (1312 bits) on interface -, id 0
> Ethernet II, Src: 50:00:00:06:00:07 (50:00:00:06:00:07), Dst: 50:00:00:03:00:07 (50:00:00:03:00:07)
> Internet Protocol Version 4, Src: 10.1.1.100, Dst: 1.1.1.53
v User Datagram Protocol, Src Port: 50413, Dst Port: 4789
    Source Port: 50413
    Destination Port: 4789
    Length: 130
  > Checksum: 0x0000 [zero-value ignored]
    [Stream index: 24]
    [Stream Packet Number: 1]
  > [Timestamps]
    UDP payload (122 bytes)
v Virtual eXtensible Local Area Network
  > Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 100010
    Reserved: 0
> Ethernet II, Src: 00:00:00_00:aa:aa (00:00:00:00:aa:aa), Dst: 00:00:00_00:cc:cc (00:00:00:00:cc:cc)
v Internet Protocol Version 4, Src: 192.168.10.10, Dst: 192.168.10.20
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 100
    Identification: 0x0000 (0)
  > 000. .... = Flags: 0x0
    ...0 0000 0000 0000 = Fragment Offset: 0
    Time to Live: 255
    Protocol: ICMP (1)
    Header Checksum: 0x262a [validation disabled]
    [Header checksum status: Unverified]
    Source Address: 192.168.10.10
    Destination Address: 192.168.10.20
    [Stream index: 11]
> Internet Control Message Protocol
```

显示ICMP请求数据包通过L2VNI传输的Wireshark捕10010

- 如您所见，源IP是10.1.1.100，端口4789作为UDP目的地。
- 由于这是VNI内部通信，VLAN 10将使用VNI 100010，而VLAN 20将使用VNI 1000。
- 允许使用不同的VNI从Host-1 ping Host-4并查看捕获。

```
HOST-1#ping 192.168.20.20
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.20.20, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 11/13/21 ms
HOST-1#
```

从HOST-1 ping HOST-4

VXLAN上的ICMP数据包：

```
> Frame 27: 164 bytes on wire (1312 bits), 164 bytes captured (1312 bits) on interface -, id 0
> Ethernet II, Src: 50:00:00:05:00:07 (50:00:00:05:00:07), Dst: 50:00:00:04:00:07 (50:00:00:04:00:07)
> Internet Protocol Version 4, Src: 10.1.1.100, Dst: 1.1.1.54
v User Datagram Protocol, Src Port: 54712, Dst Port: 4789
    Source Port: 54712
    Destination Port: 4789
    Length: 130
  > Checksum: 0x0000 [zero-value ignored]
    [Stream index: 3]
    [Stream Packet Number: 1]
  > [Timestamps]
    UDP payload (122 bytes)
v Virtual eXtensible Local Area Network
  > Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 50000
    Reserved: 0
> Ethernet II, Src: 50:00:00:01:00:07 (50:00:00:01:00:07), Dst: 50:00:00:04:00:07 (50:00:00:04:00:07)
> Internet Protocol Version 4, Src: 192.168.10.10, Dst: 192.168.20.20
> Internet Control Message Protocol
```

显示ICMP请求数据包通过L3VNI传输的Wireshark捕50000

- 由于这是VNI间通信，因此将使用L50000VNI接口。
- 检查终端主机的ARP表。

```
HOST-1#sh ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.10.10           -   0000.0000.aaaa  ARPA   GigabitEthernet0/0
Internet  192.168.10.20          18   0000.0000.cccc  ARPA   GigabitEthernet0/0
Internet  192.168.10.254          3   0000.1234.5678  ARPA   GigabitEthernet0/0
HOST-1#
```

主机1 ARP条目

```
HOST-2#sh ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.20.10           -   0000.0000.bbbb  ARPA   GigabitEthernet0/0
Internet  192.168.20.20          44   0000.0000.dddd  ARPA   GigabitEthernet0/0
Internet  192.168.20.254          4   0000.1234.5678  ARPA   GigabitEthernet0/0
HOST-2#
```

主机2 ARP条目

```
HOST-3#sh ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.10.10         103   0000.0000.aaaa  ARPA   GigabitEthernet0/0
Internet  192.168.10.20           -   0000.0000.cccc  ARPA   GigabitEthernet0/0
Internet  192.168.10.254         10   0000.1234.5678  ARPA   GigabitEthernet0/0
HOST-3#
```

主机3 ARP条目

```
HOST-4#sh ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.20.10          43   0000.0000.bbbb  ARPA   GigabitEthernet0/0
Internet  192.168.20.20           -   0000.0000.dddd  ARPA   GigabitEthernet0/0
Internet  192.168.20.254          6   0000.1234.5678  ARPA   GigabitEthernet0/0
HOST-4#
```

主机4 ARP条目

```
HOST-4#tclsh
HOST-4(tcl)#set ip_list {192.168.10.10 192.168.10.20 192.168.20.10 192.168.20.20}
192.168.10.10 192.168.10.20 192.168.20.10 192.168.20.20
HOST-4(tcl)#foreach ip $ip_list {
HOST-4(tcl)#foreach ip $ip_list {
+>            puts "Pinging $ip rep 50 size 1500"
+>            set result [exec ping $ip]
+>            puts $result
+>        }
Pinging 192.168.10.10 rep 50 size 1500

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.10.10, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 12/14/16 ms
Pinging 192.168.10.20 rep 50 size 1500

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.10.20, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 10/12/15 ms
Pinging 192.168.20.10 rep 50 size 1500

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.20.10, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 8/11/17 ms
Pinging 192.168.20.20 rep 50 size 1500

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.20.20, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms

HOST-4(tcl)#
```

从HOST-4 ping所有其他终端主机

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）。