VXLAN数据平面学习实验室（泛洪和学习机制）

目录

简介

本文档介绍如何使用VXLAN和Flood and Learn方法设置Nexus 9Kv交换机的CML。

先决条件

要求

Cisco 建议您了解以下主题：

- 了解路由和交换
- 组播路由概念，例如交汇点(RP)和平台无关组播(PIM)

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，枝叶和主干交换机均采用带Nexus 9000V交换机的思科建模实验室(CML)。

枝叶1	环回接口0 - 1.1.1.1	Loopback1 - 10.10.10.10

枝叶2	环回接口0 - 2.2.2.2	Loopb ack1 - 20.20.20.20
枝叶3	环回接口0 - 3.3.3.3	Loopback1 - 30.30.30.30
主干1	环回接口0 - 4.4.4.4	Loopback1 - 60.60.60.60 — 任播RP
主干2	环回接口0 - 5.5.5.5	Loopback1 - 60.60.60.60 — 任播RP
桌面子网	192.168.100.0/24	

使用的术语

虚拟可扩展局域网(VXLAN)隧道终端(VTEP) — 将MAC流量封装到IP流量中，并将MAC流量路由到其他VTEP。

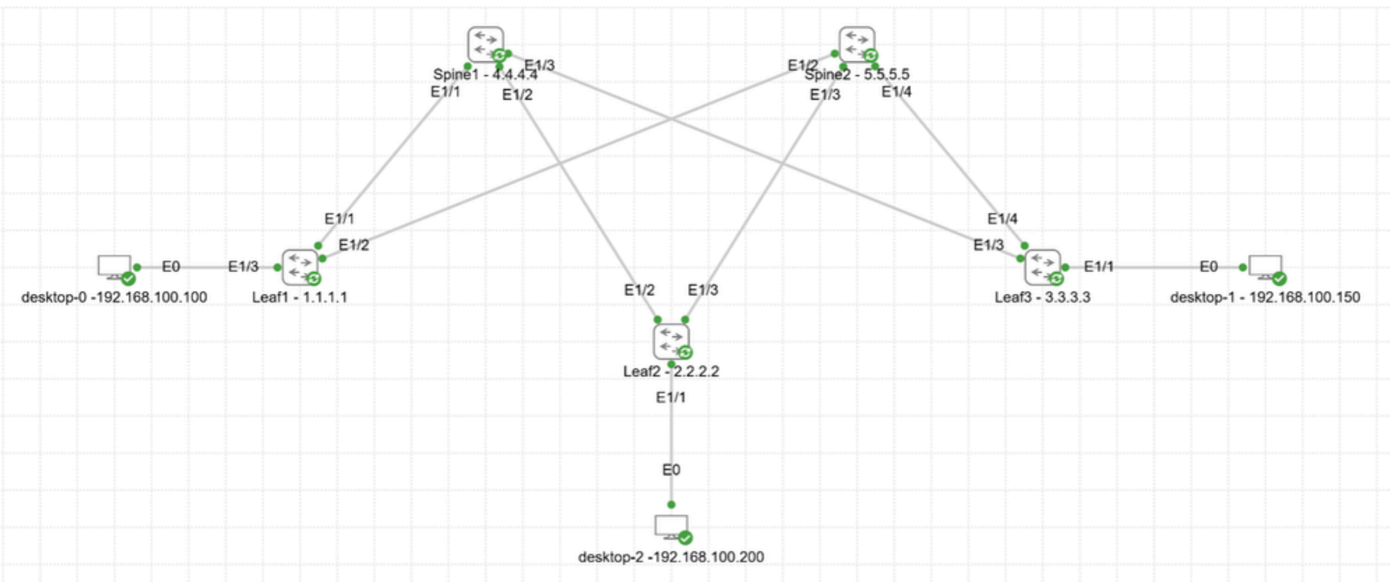
VXLAN网络标识符(VNID)- VXLAN报头中的ID，用于标识网络并且可以映射到VLAN。从转发角度来看，VNID是一个广播域。

网络虚拟接口(NVE) — 发生封装和解封的逻辑接口。

广播、未知单播和组播(BUM)

配置

网络图



网络连接图

配置

步骤1:

- 启用开放最短路径优先(OSPF)功能。
- 将环回添加到所有设备。
- 在以太网接口和环回接口上启用OSPF。

```
leaf1# show  running-config interface  ethernet 1/1

!Command: show running-config interface Ethernet1/1
!Running configuration last done at: Tue Dec 24 13:12:55 2024
!Time: Wed Dec 25 05:24:23 2024

version 9.3(8) Bios:version

interface Ethernet1/1
  no switchport
  ip address 70.0.0.2/30
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

leaf1# show  running-config interface  loopback 0

!Command: show running-config interface loopback0
!Running configuration last done at: Tue Dec 24 13:12:55 2024
!Time: Wed Dec 25 05:24:24 2024

version 9.3(8) Bios:version

interface loopback0
  ip address 1.1.1.1/32
  ip router ospf 1 area 0.0.0.0
```

在枝叶接口上启用OSPF

```
spine1# show  running-config interface ethernet 1/1

!Command: show running-config interface Ethernet1/1
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:25:46 2024

version 9.3(8) Bios:version

interface Ethernet1/1
  no switchport
  ip address 70.0.0.1/30
  ip router ospf 5 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

spine1# show  running-config interface loopback 0

!Command: show running-config interface loopback0
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:25:51 2024

version 9.3(8) Bios:version

interface loopback0
  ip address 4.4.4.4/32
  ip router ospf 5 area 0.0.0.0
```

在主干接口上启用OSPF

OSPF邻居关系在枝叶交换机和主干交换机之间建立。

```
spine1# show  ip ospf neighbors
 OSPF Process ID 5 VRF default
 Total number of neighbors: 3
 Neighbor ID     Pri State            Up Time  Address         Interface
 1.1.1.1           1 FULL/DR          16:22:51 70.0.0.2        Eth1/1
 2.2.2.2           1 FULL/DR          16:22:52 50.0.0.2        Eth1/2
 3.3.3.3           1 FULL/DR          16:22:52 30.0.0.2        Eth1/3
```

与枝叶交换机建立OSPF邻居

```
leaf1# show  ip ospf neighbors
 OSPF Process ID 1 VRF default
 Total number of neighbors: 2
 Neighbor ID     Pri State            Up Time  Address         Interface
 4.4.4.4           1 FULL/BDR         16:15:40 70.0.0.1        Eth1/1
 5.5.5.5           1 FULL/BDR         16:15:10 80.0.0.1        Eth1/2
```

使用主干交换机建立OSPF邻居

```
leaf1# ping 3.3.3.3 source-interface loopback 0
PING 3.3.3.3 (3.3.3.3): 56 data bytes
64 bytes from 3.3.3.3: icmp_seq=0 ttl=253 time=6.616 ms
64 bytes from 3.3.3.3: icmp_seq=1 ttl=253 time=6.695 ms
64 bytes from 3.3.3.3: icmp_seq=2 ttl=253 time=6.018 ms
64 bytes from 3.3.3.3: icmp_seq=3 ttl=253 time=6.52 ms
64 bytes from 3.3.3.3: icmp_seq=4 ttl=253 time=7.768 ms
```

从Leaf1到Leaf3的ping可达性

Step 2.

在枝叶交换机上添加将用于VXLAN的附加环回接口。此外，验证从所有枝叶交换机到交换矩阵的可达性。

```
leaf2# show  running-config interface loopback 1

!Command: show running-config interface loopback1
!Running configuration last done at: Wed Dec 25 05:41:46 2024
!Time: Wed Dec 25 05:42:33 2024

version 9.3(8) Bios:version

interface loopback1
  description Vxlan loopback
  ip address 20.20.20.20/32
  ip router ospf 2 area 0.0.0.0
```

VXLAN的环回

```
leaf2# ping 10.10.10.10 source 20.20.20.20
PING 10.10.10.10 (10.10.10.10) from 20.20.20.20: 56 data bytes
64 bytes from 10.10.10.10: icmp_seq=0 ttl=253 time=7.187 ms
64 bytes from 10.10.10.10: icmp_seq=1 ttl=253 time=6.248 ms
64 bytes from 10.10.10.10: icmp_seq=2 ttl=253 time=5.472 ms
64 bytes from 10.10.10.10: icmp_seq=3 ttl=253 time=4.741 ms
64 bytes from 10.10.10.10: icmp_seq=4 ttl=253 time=4.887 ms

--- 10.10.10.10 ping statistics ---
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min/avg/max = 4.741/5.707/7.187 ms
```

从Leaf2到Leaf1的重叠可达性

第 3 步：

在脊柱上配置PIM任意源组播(ASM)和任播RP:

- 启用PIM功能。
- 在所有底层链路上启用PIM。
- 在用于任播RP的主干上创建新的环回。
- 将此环回通告到OSPF。
- 配置主干上的任播RP（Nexus功能）。
- 在所有设备上配置RP。

任播RP:

任播RP是一种提供快速RP故障切换和RP负载共享的机制。任播RP涉及在两个或多个将用作RP的路由器上使用相同的IP地址(rp-address)。此IP地址必须在内部网关协议(IGP)中通告，以便其他路由器可以选择到rp地址的最佳路径。发生故障时，收敛时间将与IGP相同。

拥有多个RP和同一IP地址可确保源和接收器始终根据单播路由表路由到最近的RP。来自接收器的PIM加入消息可以发送到一个RP，而PIM指定的路由器将其本地源注册到另一个RP。
在不同的RP之间同步信息非常重要，因为某些发送方和接收方可以加入路由器1作为RP，而其他发送方和接收方可以加入路由器2作为RP。如果路由器没有关于所有源的完整信息，则组播通信可能会中断。为了解决此问题，需要一种机制来同步所有作为RP的路由器之间的源信息。
有两种协议可用于此目的：组播源发现协议(MSDP)和PIM。

```
spine1# show  running-config interface loopback 1

!Command: show running-config interface loopback1
!Running configuration last done at: Tue Dec 24 13:16:16 2024
!Time: Wed Dec 25 05:50:44 2024

version 9.3(8) Bios:version

interface loopback1
  description Anycast RP loopback
  ip address 60.60.60.60/32
  ip router ospf 5 area 0.0.0.0
  ip pim sparse-mode

spine1# show  running-config | section  rp
ip pim rp-address 60.60.60.60 group-list 224.0.0.0/4
ip pim anycast-rp 60.60.60.60 4.4.4.4
ip pim anycast-rp 60.60.60.60 5.5.5.5
```

60.60.60.60是任播RP IP，4.4.4.4/5.5.5.5是Spine 1和Spine 2的环回IP

```
leaf2# show  running-config interface  ethernet 1/2

!Command: show running-config interface Ethernet1/2
!Running configuration last done at: Wed Dec 25 05:41:46 2024
!Time: Wed Dec 25 05:51:18 2024

version 9.3(8) Bios:version

interface Ethernet1/2
  no switchport
  ip address 50.0.0.2/30
  ip router ospf 2 area 0.0.0.0
  ip pim sparse-mode
  no shutdown

leaf2# show  running-config | section rp-address
ip pim rp-address 60.60.60.60 group-list 224.0.0.0/4
```

枝叶交换机上的任播RP配置

```
spine1# show  ip pim neighbor
PIM Neighbor Status for VRF "default"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD
 ECMP Redirect
                                                         Priority Capable State
    Capable
70.0.0.2        Ethernet1/1          16:44:28  00:01:39  1        yes     n/a
  no
50.0.0.2        Ethernet1/2          16:44:28  00:01:44  1        yes     n/a
  no
30.0.0.2        Ethernet1/3          16:44:28  00:01:25  1        yes     n/a
  no
spine1# show  ip pim  rp
PIM RP Status Information for VRF "default"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

Anycast-RP 60.60.60.60 members:
  4.4.4.4*  5.5.5.5

RP: 60.60.60.60*, (0),
 uptime: 16:57:06   priority: 255,
 RP-source: (local),
 group ranges:
 224.0.0.0/4
```

已形成PIM邻居关系

注意：不要忘记将PIM放在用于枝叶交换机上VXLAN的环回上。

第 4 步：

- 启用VXLAN功能。
- 启用将VLAN映射到虚拟网络标识符(VNI)的功能。
- 创建NVE
- 配置接入端口到桌面。

```
feature vn-segment-vlan-based
feature nv overlay
```

启用VXLAN功能

```
!Command: show running-config interface nve1
!Running configuration last done at: Wed Dec 25 06:08:01 2024
!Time: Wed Dec 25 06:08:04 2024

version 9.3(8) Bios:version

interface nve1
  no shutdown
  source-interface loopback1
  member vni 10000
    mcast-group 239.0.0.1
```

创建NVE接口

```
vlan 10
  vn-segment 10000
```

VLAN到VN网段映射

```
leaf1# show  interface nve 1
nve1 is up
admin state is up,  Hardware: NVE
  MTU 9216 bytes
  Encapsulation VXLAN
  Auto-mdix is turned off
  RX
    ucast: 39 pkts, 3346 bytes - mcast: 0 pkts, 0 bytes
  TX
    ucast: 18 pkts, 2216 bytes - mcast: 0 pkts, 0 bytes
```

NVE接口的状态

从Desktop0 ping Desktop1和Desktop2，以验证其连通性。

从Desktop0向Desktop1发起地址解析协议(ARP)请求时，ARP数据包将发送到Leaf1。然后Leaf1将使用用于VNI1000的组播地址239.0.0.1将数据包转发到主干设备。主干设备会将数据包组播到属于同一VNI 10000的所有枝叶设备。

从Desktop0(192.168.100.100)对Desktop1(192.168.100.150)和Desktop2(192.168.100.200)执行ping操作。

```
inserthostname-here:~$ ping 192.168.100.150
PING 192.168.100.150 (192.168.100.150): 56 data bytes
64 bytes from 192.168.100.150: seq=0 ttl=42 time=8.477 ms
64 bytes from 192.168.100.150: seq=1 ttl=42 time=12.791 ms
64 bytes from 192.168.100.150: seq=2 ttl=42 time=8.352 ms
^C
--- 192.168.100.150 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max = 8.352/9.873/12.791 ms
inserthostname-here:~$
inserthostname-here:~$
inserthostname-here:~$ ping 192.168.100.200
PING 192.168.100.200 (192.168.100.200): 56 data bytes
64 bytes from 192.168.100.200: seq=0 ttl=42 time=15.432 ms
64 bytes from 192.168.100.200: seq=1 ttl=42 time=9.228 ms
64 bytes from 192.168.100.200: seq=2 ttl=42 time=7.133 ms
^C
--- 192.168.100.200 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max = 7.133/10.597/15.432 ms
inserthostname-here:~$ arp -a
? (192.168.100.150) at 52:54:00:05:84:a2 [ether]  on eth0
? (192.168.100.1) at 00:01:00:01:00:01 [ether]  on eth0
? (192.168.100.200) at 52:54:00:10:70:ae [ether]  on eth0
```

从Desktop0 ping Desktop1和Desktop2

LEAF1正在与LEAF3形成NVE对等体。

<#root>

```
leaf1# show  nve  peers

Interface Peer-IP                                 State LearnType Uptime   Route

r-Mac

--------- --------------------------------------  ----- --------- -------- -----

------------

nve1

30.30.30.30

                           Up     DP        00:10:23 n/a



leaf1# show  nve  vni 10000

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      10000    239.0.0.1         Up    DP   L2 [10]
```

LEAF3正在与LEAF1形成NVE对等体。

<#root>

```
leaf3# show  nve  peers
Interface Peer-IP                                 State LearnType Uptime   Route
r-Mac
--------- --------------------------------------  ----- --------- -------- -----
------------
nve1
```

10.10.10.10

```
                           Up    DP        00:10:56 n/a


leaf3# show  nve vni 10000
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      10000    239.0.0.1         Up    DP   L2 [10]
```

```
leaf1# show  mac address-table

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

---------+-----------------+--------+---------+------+----+------------------

*   10
```

5254.0004.8b92

```
   dynamic  0         F      F    Eth1/3                                      ------- MAC Address of De

*   10
```

5254.0005.84a2

```
   dynamic  0         F      F    nve1(30.30.30.30)                   ------- MAC Address of Desktop1 

G    -     5206.ab8a.1b08   static   -         F      F    sup-eth1(R)



leaf3# show  mac address-table

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

---------+-----------------+--------+---------+------+----+------------------

*   10
```

5254.0004.8b92

```
   dynamic  0         F      F    nve1(10.10.10.10)                ------- MAC Address of Desktop0 con

*   10
```

5254.0005.84a2

```
   dynamic  0         F      F    Eth1/1                                    ------- MAC Address of Desk

G    -     5206.0619.1b08   static   -         F      F    sup-eth1(R)
```

以下是从Leaf1向组播发起ARP数据包时的Wireshark快照。

No.	Time	Source	Destination	Protocol	Length	Info
7	5.105615	52:54:00:05:84:a2	52:54:00:04:8b:92	ARP	110	192.168.100.200 is at 52:54:00:05:84:a2
8	7.019252	52:54:00:04:8b:92	ff:ff:ff:ff:ff:ff	ARP	110	Who has 192.168.100.200? Tell 192.168.100.100

```
▶ Frame 8: 110 bytes on wire (880 bits), 110 bytes captured (880 bits)
▶ Ethernet II, Src: 52:06:ab:8a:1b:08 (52:06:ab:8a:1b:08), Dst: IPv4mcast_01 (01:00:5e:00:00:01)
▶ Internet Protocol Version 4, Src: 10.10.10.10, Dst: 239.0.0.1
▶ User Datagram Protocol, Src Port: 50384, Dst Port: 4789
▶ Virtual eXtensible Local Area Network
▶ Ethernet II, Src: RealtekU_04:8b:92 (52:54:00:04:8b:92), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
▶ Address Resolution Protocol (request)
```

显示ARP请求数据包进入组播组的Wireshark捕获

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）。