排除ACI交换矩阵内转发 — L3转发故障：不同BD中的两个终端

目录

简介

本文档介绍了解ACI L3转发场景并对其进行故障排除的步骤。

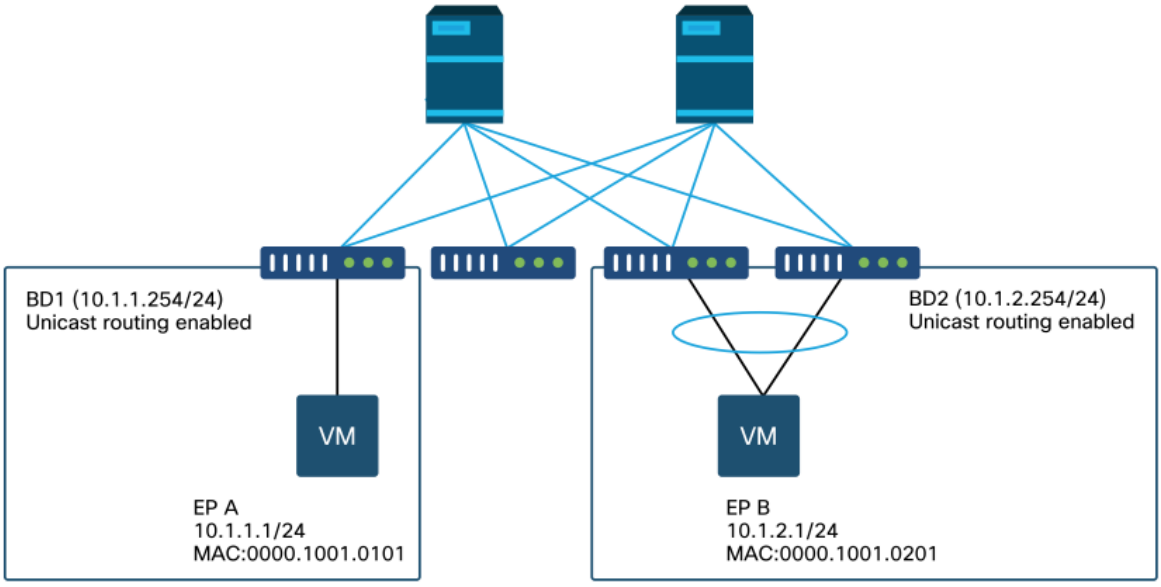
背景信息

本文档中的内容摘自 思科以应用为中心的基础设施故障排除（第二版） 书，特别是 **交换矩阵内转发 — L3转发：不同BD中的两个终端** 第章。

L3转发：不同BD中的两个终端

本章介绍不同网桥域中的终端无法相互通信的故障排除示例。这将是ACI交换矩阵路由的流。图1显示了拓扑。

不同网桥域中的终端



高级故障排除工作流程

以下是典型的故障排除步骤和验证命令：

第一次检查 — 验证编程

- 应该将BD普及网关推送到枝叶节点。
- 通往目的BD子网的路由应推送到枝叶节点。
- 应解析主机默认网关的ARP。

第二次检查 — 在枝叶节点上通过CLI验证学习和表条目

- 检查源枝叶节点和目标枝叶节点了解终端以及是否了解目标终端： 终端表 — “显示终端”。TEP目标 — “show interface tunnel <x>”。在“show ip route <TEP address> vrf overlay-1”命令中查找TEP目标。
- 检查主干节点了解终端： “show coop internal info”。

第三项检查 — 获取数据包并分析转发决策

- 使用ELAM（ELAM Assistant或CLI）验证帧是否存在。
- 或使用fTriage跟踪流量。

已知终端的故障排除工作流程

检查BD的普及网关

在本示例中，将使用以下源和目标终端：

- leaf1下的EP A 10.1.1.1。

- VPC对leaf3和leaf4下的EP B 10.1.2.1。

应看到以下沉浸式网关：

- 10.1.1.254/24，用于leaf1上的BD1网关。
- 10.1.2.254/24，用于leaf3和leaf4上的BD2网关。

可使用以下工具检查此情况：'show ip interface vrf <vrf name>'。

leaf1:

```
leaf1# show ip interface vrf Prod:VRF1
IP Interface Status for VRF "Prod:VRF1"
vlan7, Interface status: protocol-up/link-up/admin-up, iod: 106, mode: pervasive
 IP address: 10.1.1.254, IP subnet: 10.1.1.0/24
 IP broadcast address: 255.255.255.255
 IP primary address route-preference: 0, tag: 0
```

leaf3和4:

```
leaf3# show ip interface  vrf Prod:VRF1
IP Interface Status for VRF "Prod:VRF1"
vlan1, Interface status: protocol-up/link-up/admin-up, iod: 159, mode: pervasive
 IP address: 10.1.2.254, IP subnet: 10.1.2.0/24
 IP broadcast address: 255.255.255.255
 IP primary address route-preference: 0, tag: 0

leaf4# show ip interface  vrf Prod:VRF1
IP Interface Status for VRF "Prod:VRF1"
vlan132, Interface status: protocol-up/link-up/admin-up, iod: 159, mode: pervasive
 IP address: 10.1.2.254, IP subnet: 10.1.2.0/24
 IP broadcast address: 255.255.255.255
 IP primary address route-preference: 0, tag: 0
```

请注意，leaf3和leaf4具有相同的普及网关地址，但可能会看到不同的SVI VLAN封装。

- leaf3使用VLAN 1。
- leaf4使用VLAN 132。

这是因为VLAN 1或VLAN 132是枝叶上的本地VLAN。

如果未将沉浸式网关IP地址推送到枝叶，请在APIC GUI中验证不存在阻止部署VLAN的故障。

检查枝叶上的路由表

Leaf1在子网10.1.2.0/24中没有任何端点，但是它必须具有到该子网的路由才能到达该子网：

```
leaf1# show ip route 10.1.2.0/24 vrf Prod:VRF1
IP Route Table for VRF "Prod:VRF1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.1.2.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.8.65%overlay-1, [1/0], 00:22:37, static, tag 4294967294
       recursive next hop: 10.0.8.65/32%overlay-1
```

请注意，标记为“沉浸式”和“直接”的路由的下一跳为10.0.8.65。这是所有主干上存在的任播v4环回地

址。

```
leaf1# show isis  dteps  vrf overlay-1 | egrep 10.0.8.65
10.0.8.65          SPINE   N/A             PHYSICAL,PROXY-ACAST-V4
```

同样，leaf3和leaf4应具有通往10.1.1.0/24的路由。

```
leaf3# show ip route 10.1.1.1 vrf Prod:VRF1
IP Route Table for VRF "Prod:VRF1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.1.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.8.65%overlay-1, [1/0], 00:30:25, static, tag 4294967294
       recursive next hop: 10.0.8.65/32%overlay-1
```

如果缺少这些路由，很可能是因为在BD1中的EPG和BD2中的EPG之间没有合同。如果枝叶下的BD1中没有本地终端，则BD1普及网关不会推送到枝叶。如果EPG中有本地终端与BD1中的另一个EPG有合同，则会在枝叶上获取BD1子网。

默认网关IP的ARP解析

由于本地终端所在的枝叶应该具有沉浸式网关，因此对沉浸式网关的ARP请求应始终由本地枝叶解决。可以使用以下命令在本地枝叶上检查此情况：

```
leaf1# show ip arp internal event-history event  | egrep 10.1.1.1
  [116] TID 26571:arp_handle_arp_request:6135: log_collect_arp_pkt; sip = 10.1.1.1; dip =
10.1.1.254;interface = Vlan7; phy_inteface = Ethernet1/3; flood = 0; Info = Sent ARP response.
   [116] TID 26571:arp_process_receive_packet_msg:8384: log_collect_arp_pkt; sip = 10.1.1.1; dip
= 10.1.1.254;interface = Vlan7; phy_interface = Ethernet1/3;Info = Received arp request
```

入口枝叶源IP和MAC终端学习

对于第3层转发，ACI将执行第3层源IP学习和目标IP查找。获取的IP地址范围设为VRF。

这可以在EPG的“操作”选项卡中的GUI上检查。请注意，此处的IP和MAC都已获知。

EPG运营端点

EPG运营端点 — 详细信息

End Point	MAC	IP	Learning Source	Hosting Server	Reportin Controlle Name	Interface	Multicas Address	Encap
EP-00:00:10:01:01:01	00:00:10:01:01:01	10.1.1.1	learned	---	---	Pod-1/Node-101/eth1/3 (learned)	---	vlan-2501
EP-00:00:10:01:01:02	00:00:10:01:01:02	10.1.1.2,...	learned	---	---	Pod-1/Node-103-104/N3k-3-VPC3-4 (learned)	---	vlan-2501

检查本地终端是否已在本地枝叶上获取。在leaf1上检查IP 10.1.1.1已获知：

```
leaf1# show endpoint ip 10.1.1.1
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+------------------------------------+--------------+-----------------+-------------+----------
---+
    VLAN/                           Encap           MAC Address       MAC Info/       Interface
    Domain                          VLAN            IP Address        IP Info
+------------------------------------+--------------+-----------------+-------------+----------
---+
46                                      vlan-2501    0000.1001.0101 L
eth1/3
Prod:VRF1                               vlan-2501          10.1.1.1 L
                    eth1/3
```

如上所示，终端内容为：

- BD（BD的内部VLAN为46），采用EPG的VLAN封装(vlan-2501)和通过eth1/3获取的MAC地址
- 使用IP 10.1.1.1的VRF(Prod:VRF1)

这可以理解为相当于传统网络中的ARP条目。ACI不会在终端的ARP表中存储ARP信息。终端仅在终端表中可见。

枝叶上的ARP表仅用于L3Out下一跳。

```
leaf1# show ip arp vrf Prod:VRF1
Flags: * - Adjacencies learnt on non-active FHRP router
       + - Adjacencies synced via CFSoE
       # - Adjacencies Throttled for Glean
       D - Static Adjacencies attached to down interface    IP ARP Table for context Prod:VRF1
Total number of entries: 0
Address         Age       MAC Address     Interface
<NO ENTRY >
```

入口枝叶目标IP查找 — 已知远程终端

假设目标IP已知（已知单播），下面是目标IP 10.1.2.1的“show endpoint”输出。这是一个远程学习，因为它不驻留在枝叶1上，具体指向本地学习它的隧道接口（隧道4）。

远程终端仅包含IP或MAC，并且不能在同一条目中同时包含两者。仅当在本地获取终端时，才会发生同一终端中的MAC地址和IP地址。

```
leaf1# show endpoint ip 10.1.2.1
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
```

```
L - local            E - shared-service
+----------------------------------+--------------+-----------------+-------------+----------
---+
    VLAN/                          Encap           MAC Address       MAC Info/       Interface
    Domain                         VLAN            IP Address        IP Info
+----------------------------------+--------------+-----------------+-------------+----------
---+
Prod:VRF1                                                  10.1.2.1 p
tunnel4

leaf1# show interface tunnel 4
Tunnel4 is up
  MTU 9000 bytes, BW 0 Kbit
  Transport protocol is in VRF "overlay-1"
  Tunnel protocol/transport is ivxlan
  Tunnel source 10.0.88.95/32 (lo0)
  Tunnel destination 10.0.96.66
  Last clearing of "show interface" counters never
  Tx
  0 packets output, 1 minute output rate 0 packets/sec
  Rx
  0 packets input, 1 minute input rate 0 packets/sec
```

目的TEP是枝叶3和4 VPC对的任播TEP，通过到主干的上行链路获知。

```
leaf1# show ip route 10.0.96.66 vrf overlay-1
IP Route Table for VRF "overlay-1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.0.96.66/32, ubest/mbest: 4/0
  *via 10.0.88.65, eth1/49.10, [115/3], 02w06d, isis-isis_infra, isis-l1-int
  *via 10.0.128.64, eth1/51.8, [115/3], 02w06d, isis-isis_infra, isis-l1-int
  *via 10.0.88.64, eth1/52.126, [115/3], 02w06d, isis-isis_infra, isis-l1-int
  *via 10.0.88.94, eth1/50.128, [115/3], 02w06d, isis-isis_infra, isis-l1-int
```

可以使用“show system internal epm endpoint ip <ip>”命令收集IP 10.1.2.1的其他终端信息。

```
leaf1# show system internal epm endpoint ip 10.1.2.1
MAC : 0000.0000.0000 ::: Num IPs : 1
IP# 0 : 10.1.2.1 ::: IP# 0 flags :  ::: l3-sw-hit: No
Vlan id : 0 ::: Vlan vnid : 0 ::: VRF name : Prod:VRF1
BD vnid : 0 ::: VRF vnid : 2097154
Phy If : 0 ::: Tunnel If : 0x18010004
Interface : Tunnel4
Flags : 0x80004420 ::: sclass : 32771 ::: Ref count : 3
EP Create Timestamp : 10/01/2019 13:53:16.228319
EP Update Timestamp : 10/01/2019 14:04:40.757229
EP Flags : peer-aged|IP|sclass|timer|
::::
```

在输出检查中：

- VRF VNID已填充 — 这是用于将VXLAN中的帧封装到交换矩阵的VNID。
- MAC地址是0000.0000.0000，因为远程IP条目上从未填充MAC地址。
- BD VNID未知，因为对于路由帧，入口枝叶充当路由器并执行MAC重写。这意味着远程枝叶无法看到目标的BD，只有VRF。

现在，该帧将封装到发往远程TEP 10.0.96.66的VXLAN ID为2097154（即VRF的VNID）的VXLAN帧中。它将在overlay-1路由表中路由（IS-IS路由）并到达目标TEP。此处它将到达leaf3或

leaf4，因为10.0.96.66是leaf3和leaf4 VPC对的任播TEP地址。

出口枝叶上的源IP学习

此处的输出来自leaf3，但在leaf4上类似。当数据包到达leaf3（目标枝叶和TEP的所有者）时，枝叶将学习VRF中数据包的源IP。

```
leaf3# show endpoint ip 10.1.1.1
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+----------------------------------+--------------+-----------------+-------------+----------
---+
    VLAN/                          Encap           MAC Address       MAC Info/       Interface
    Domain                         VLAN            IP Address        IP Info
+----------------------------------+--------------+-----------------+-------------+----------
---+
Prod:VRF1                                                  10.1.1.1 p
tunnel26

leaf3# show interface tunnel 26
Tunnel26 is up
  MTU 9000 bytes, BW 0 Kbit
  Transport protocol is in VRF "overlay-1"
  Tunnel protocol/transport is ivxlan
  Tunnel source 10.0.88.91/32 (lo0)
  Tunnel destination 10.0.88.95
  Last clearing of "show interface" counters never
  Tx
  0 packets output, 1 minute output rate 0 packets/sec
  Rx
  0 packets input, 1 minute input rate 0 packets/sec
```

目的TEP 10.0.88.95是leaf1的TEP地址，通过到主干的所有上行链路获知。

出口枝叶上的目标IP查找

最后一步是出口枝叶查找目标IP。查看10.1.2.1的终端表。

这提供了以下信息：

- 出口枝叶知道目标10.1.2.1（类似于路由表中的/32主机路由），并且路由是在正确的VRF中获知的。
- 出口枝叶知道MAC 0000.1001.0201（终端信息）。
- 出口枝叶知道发往10.1.2.1的流量必须封装在vlan-2502中，然后通过port-channel 1(po1)发送。

```
leaf3# show endpoint ip 10.1.2.1
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+----------------------------------+--------------+-----------------+-------------+----------
```

```
---+
    VLAN/                            Encap           MAC Address       MAC Info/       Interface
    Domain                           VLAN            IP Address        IP Info
+-----------------------------------+---------------+-----------------+--------------+----------
---+
2                                         vlan-2502    0000.1001.0201 LpV
po1
Prod:VRF1                                 vlan-2502          10.1.2.1 LpV
po1
```

按数据路径进行分类

在APIC中使用fTriage跟踪数据路径流。请记住，fTriage依赖于ELAM，因此它需要真正的数据流。这样可以确认完整的数据路径，同时确认数据包已退出枝叶3端口1/16上的交换矩阵。

```
apic1# ftriage route -ii LEAF:101 -sip 10.1.1.1 -dip 10.1.2.1
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "InProgress", "pid": "6888",
"apicId": "1", "id": "0"}}}
Starting ftriage
Log file name for the current run is: ftlog_2019-10-01-21-17-54-175.txt
2019-10-01 21:17:54,179 INFO     /controller/bin/ftriage route -ii LEAF:101 -sip 10.1.1.1 -dip
10.1.2.1
2019-10-01 21:18:18,149 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-01 21:18:39,194 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-leaf1
Ingress: Eth1/3 Egress: Eth1/51 Vnid: 2097154
2019-10-01 21:18:39,413 INFO     ftriage:     main:242  ingress encap string vlan-2501
2019-10-01 21:18:39,419 INFO     ftriage:     main:271  Building ingress BD(s), Ctx
2019-10-01 21:18:41,240 INFO     ftriage:     main:294  Ingress BD(s) Prod:BD1
2019-10-01 21:18:41,240 INFO     ftriage:     main:301  Ingress Ctx: Prod:VRF1
2019-10-01 21:18:41,349 INFO     ftriage:   pktrec:490  bdsol-aci32-leaf1: Collecting transient
losses snapshot for LC module: 1
2019-10-01 21:19:05,747 INFO     ftriage:     main:933  SIP 10.1.1.1 DIP 10.1.2.1
2019-10-01 21:19:05,749 INFO     ftriage:  unicast:973  bdsol-aci32-leaf1: <- is ingress node
2019-10-01 21:19:08,459 INFO     ftriage:  unicast:1215 bdsol-aci32-leaf1: Dst EP is remote
2019-10-01 21:19:09,984 INFO     ftriage:     misc:657  bdsol-aci32-leaf1:
DMAC(00:22:BD:F8:19:FF) same as RMAC(00:22:BD:F8:19:FF)
2019-10-01 21:19:09,984 INFO     ftriage:     misc:659  bdsol-aci32-leaf1: L3 packet getting
routed/bounced in SUG
2019-10-01 21:19:10,248 INFO     ftriage:     misc:657  bdsol-aci32-leaf1: Dst IP is present in
SUG L3 tbl
2019-10-01 21:19:10,689 INFO     ftriage:     misc:657  bdsol-aci32-leaf1: RwDMAC
DIPo(10.0.96.66) is one of dst TEPs ['10.0.96.66']
2019-10-01 21:20:56,148 INFO     ftriage:     main:622  Found peer-node bdsol-aci32-spine3 and
IF: Eth2/1 in candidate list
2019-10-01 21:21:01,245 INFO     ftriage:     node:643  bdsol-aci32-spine3: Extracted Internal-
port GPD Info for lc: 2
2019-10-01 21:21:01,245 INFO     ftriage:     fcls:4414 bdsol-aci32-spine3: LC trigger ELAM with
IFS: Eth2/1 Asic :0 Slice: 0 Srcid: 32
2019-10-01 21:21:33,894 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-spine3
Ingress: Eth2/1 Egress: LC-2/0 FC-22/0 Port-1 Vnid: 2097154
2019-10-01 21:21:33,895 INFO     ftriage:   pktrec:490  bdsol-aci32-spine3: Collecting transient
losses snapshot for LC module: 2
2019-10-01 21:21:54,487 INFO     ftriage:      fib:332  bdsol-aci32-spine3: Transit in spine
2019-10-01 21:22:01,568 INFO     ftriage:  unicast:1252 bdsol-aci32-spine3: Enter
dbg_sub_nexthop with Transit inst: ig infra: False glbs.dipo: 10.0.96.66
2019-10-01 21:22:01,682 INFO     ftriage:  unicast:1417 bdsol-aci32-spine3: EP is known in COOP
(DIPo = 10.0.96.66)
2019-10-01 21:22:05,713 INFO     ftriage:  unicast:1458 bdsol-aci32-spine3: Infra route
10.0.96.66 present in RIB
2019-10-01 21:22:05,713 INFO     ftriage:     node:1331 bdsol-aci32-spine3: Mapped LC interface:
LC-2/0 FC-22/0 Port-1 to FC interface: FC-22/0 LC-2/0 Port-1
```

```
2019-10-01 21:22:10,799 INFO     ftriage:     node:460  bdsol-aci32-spine3: Extracted GPD Info
for fc: 22
2019-10-01 21:22:10,799 INFO     ftriage:     fcls:5748 bdsol-aci32-spine3: FC trigger ELAM with
IFS: FC-22/0 LC-2/0 Port-1 Asic :0 Slice: 2 Srcid: 24
2019-10-01 21:22:29,322 INFO     ftriage:  unicast:1774 L3 packet Seen on FC of node: bdsol-
aci32-spine3  with Ingress: FC-22/0 LC-2/0 Port-1  Egress: FC-22/0 LC-2/0 Port-1 Vnid: 2097154
2019-10-01 21:22:29,322 INFO     ftriage:   pktrec:487  bdsol-aci32-spine3: Collecting transient
losses snapshot for FC module: 22
2019-10-01 21:22:31,571 INFO     ftriage:     node:1339 bdsol-aci32-spine3: Mapped FC interface:
FC-22/0 LC-2/0 Port-1 to LC interface: LC-2/0 FC-22/0 Port-1
2019-10-01 21:22:31,572 INFO     ftriage:  unicast:1474 bdsol-aci32-spine3: Capturing Spine
Transit pkt-type L3 packet on egress LC on Node: bdsol-aci32-spine3 IFS: LC-2/0 FC-22/0 Port-1
2019-10-01 21:22:31,991 INFO     ftriage:     fcls:4414 bdsol-aci32-spine3: LC trigger ELAM with
IFS: LC-2/0 FC-22/0 Port-1 Asic :0 Slice: 1 Srcid: 0
2019-10-01 21:22:48,952 INFO     ftriage:  unicast:1510 bdsol-aci32-spine3: L3 packet Spine
egress  Transit pkt Seen on bdsol-aci32-spine3 Ingress: LC-2/0 FC-22/0 Port-1 Egress: Eth2/3
Vnid: 2097154
2019-10-01 21:22:48,952 INFO     ftriage:   pktrec:490  bdsol-aci32-spine3: Collecting transient
losses snapshot for LC module: 2
2019-10-01 21:23:50,748 INFO     ftriage:     main:622  Found peer-node bdsol-aci32-leaf3 and
IF: Eth1/51 in candidate list
2019-10-01 21:24:05,313 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-leaf3
Ingress: Eth1/51 Egress: Eth1/12 (Po1) Vnid: 11365
2019-10-01 21:24:05,427 INFO     ftriage:   pktrec:490  bdsol-aci32-leaf3: Collecting transient
losses snapshot for LC module: 1
2019-10-01 21:24:24,369 INFO     ftriage:     nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4326 scope:34 filter:65534
2019-10-01 21:24:25,698 INFO     ftriage:     main:522  Computed egress encap string vlan-2502
2019-10-01 21:24:25,704 INFO     ftriage:     main:313  Building egress BD(s), Ctx
2019-10-01 21:24:27,510 INFO     ftriage:     main:331  Egress Ctx Prod:VRF1
2019-10-01 21:24:27,510 INFO     ftriage:     main:332  Egress BD(s): Prod:BD2
2019-10-01 21:24:30,536 INFO     ftriage:  unicast:1252 bdsol-aci32-leaf3: Enter dbg_sub_nexthop
with Local inst: eg infra: False glbs.dipo: 10.0.96.66
2019-10-01 21:24:30,537 INFO     ftriage:  unicast:1257 bdsol-aci32-leaf3: dbg_sub_nexthop
invokes dbg_sub_eg for vip
2019-10-01 21:24:30,537 INFO     ftriage:  unicast:1784 bdsol-aci32-leaf3: <- is egress node
2019-10-01 21:24:30,684 INFO     ftriage:  unicast:1833 bdsol-aci32-leaf3: Dst EP is local
2019-10-01 21:24:30,685 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: EP if(Po1) same as
egr if(Po1)
2019-10-01 21:24:30,943 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: Dst IP is present in
SUG L3 tbl
2019-10-01 21:24:31,242 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: RW seg_id:11365 in
SUG same as EP segid:11365
2019-10-01 21:24:37,631 INFO     ftriage:     main:961  Packet is Exiting fabric with peer-
device: bdsol-aci32-n3k-3 and peer-port: Ethernet1/12
```

使用ELAM Assistant应用在出口枝叶上捕获数据包

下面是来自主干的leaf3上使用ELAM Assistant应用捕获的数据包。这表明：

- 来自外部第4层信息的VNID(VNID为2097154)。
- 外部L3报头源TEP和目标TEP。

ELAM Assistant — L3流出口枝叶（第1部分）

Device Type	LEAF
Packet Direction	egress (spine LC -> leaf)
Inconming I/F	eth1/51

L2 Header	
Destination MAC	000C.0C0C.0C0C
Source MAC	000C.0C0C.0C0C
Access Encap VLAN	No VLAN Tag
CoS	No VLAN Tag (= No CoS)

L3 Header	
L3 Type	IPv4
Destination IP	10.1.2.1
Source IP	10.1.1.1
IP Protocol	0x1 (ICMP)
DSCP	0
TTL	254
Don't Fragment Bit	0x0 (0x0)
IP Checksum	Unsupported for ELAM with VxLAN data
IP Packet Length	Unsupported for ELAM with VxLAN data

ELAM Assistant — L3流出口枝叶（第2部分）

L2 Header (Outer VxLAN)	
Destination MAC	000C.0C0C.0C0C
Source MAC	000D.0D0D.0D0D
Access Encap VLAN	2
CoS	0

L3 Header (Outer VxLAN)	
L3 Type	IPv4
Destination IP	10.0.96.66 (vPC (103_104))
Source IP	10.0.88.95 (bdsol-aci32-leaf1)
IP Protocol	0x11 (UDP)
DSCP	0
TTL	31
Don't Fragment Bit	0x0 (0x0)

L4 Header (Outer VxLAN)	
L4 Type	iVxLAN
DL (Don't Learn) Bit	0 (not set)
Src Policy Applied Bit	0 (Contract has yet to be applied)
Dst Policy Applied Bit	0 (Contract has yet to be applied)
Source EPG (sclass / src pcTag)	0x8002 / 32770 (Prod:App:EPG1)
VRF/BD VNID	2097154 (Prod:VRF1)

“数据包转发信息”部分证明它从port-channel 1获得

ELAM Assistant — L3出口枝叶 — 数据包转发信息

Packet Forwarding Information

	Forward Result
Destination Type	To a local port
Destination Logical Port	Po1
Destination Physical Port	eth1/12
Sent to SUP/CPU instead	no
SUP Redirect Reason (SUP code)	NONE

	Contract
Destination EPG pcTag (dclass)	32771 (null)
Source EPG pcTag (sclass)	32770 (null)
Contract was applied	1 (Contract was applied on this node)

	Drop
Drop Code	no drop

未知终端的故障排除工作流程

本部分显示入口枝叶不知道目标IP时的区别。

入口枝叶目标IP查找

第一步是检查是否存在目的IP的终端学习。

```
leaf1# show endpoint ip 10.1.2.1
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+----------------------------------+--------------+----------------+-------------+----------
--+
    VLAN/                          Encap           MAC Address       MAC Info/       Interface
    Domain                         VLAN            IP Address        IP Info
+----------------------------------+--------------+----------------+-------------+----------
--+
<NO ENTRY>
```

终端表中没有目的地的路由，因此下一步是检查路由表，查找到达目的地的最长前缀匹配路由：

```
leaf1# show ip route 10.1.2.1 vrf Prod:VRF1
IP Route Table for VRF "Prod:VRF1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.1.2.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.8.65%overlay-1, [1/0], 01:40:18, static, tag 4294967294
       recursive next hop: 10.0.8.65/32%overlay-1
```

落在/24 BD子网10.1.2.0/24上意味着枝叶将帧封装在VXLAN中，目标TEP为10.0.8.65（主干上的任播v4）。 帧将使用VXLAN ID，即VRF VNID。

主干上的COOP查找 — 目标IP已知

数据包将到达IP数据库中执行COOP查找的其中一个主干。必须验证源，并且需要从COOP数据库正确学习目的IP。

要在COOP数据库中查找IP，关键是VRF VNID(2097154示例中为10)

从下面的输出中，可以确认COOP数据库具有来自TEP 10.0.88.95(leaf1)的源IP的条目正确。

```
spine1# show coop internal info ip-db key 2097154 10.1.1.1
IP address : 10.1.1.1
Vrf : 2097154
Flags : 0
EP bd vnid : 15302583
EP mac :  00:00:10:01:01:01
Publisher Id : 10.0.88.95
Record timestamp : 10 01 2019 14:16:50 522482647
```

```
Publish timestamp : 10 01 2019 14:16:50 532239332
Seq No: 0
Remote publish timestamp: 01 01 1970 00:00:00 0
URIB Tunnel Info
Num tunnels : 1
      Tunnel address : 10.0.88.95
      Tunnel ref count : 1
```

以下输出显示，COOP数据库具有来自TEP 10.0.96.66（leaf3和4 VPC对的任播TEP）的目标IP条目正确

```
spine1# show coop internal info ip-db key 2097154 10.1.2.1
IP address : 10.1.2.1
Vrf : 2097154
Flags : 0
EP bd vnid : 15957974
EP mac :  00:00:10:01:02:01
Publisher Id : 10.0.88.90
Record timestamp : 10 01 2019 14:52:52 558812544
Publish timestamp : 10 01 2019 14:52:52 559479076
Seq No: 0
Remote publish timestamp: 01 01 1970 00:00:00 0
URIB Tunnel Info
Num tunnels : 1
      Tunnel address : 10.0.96.66
      Tunnel ref count : 1
```

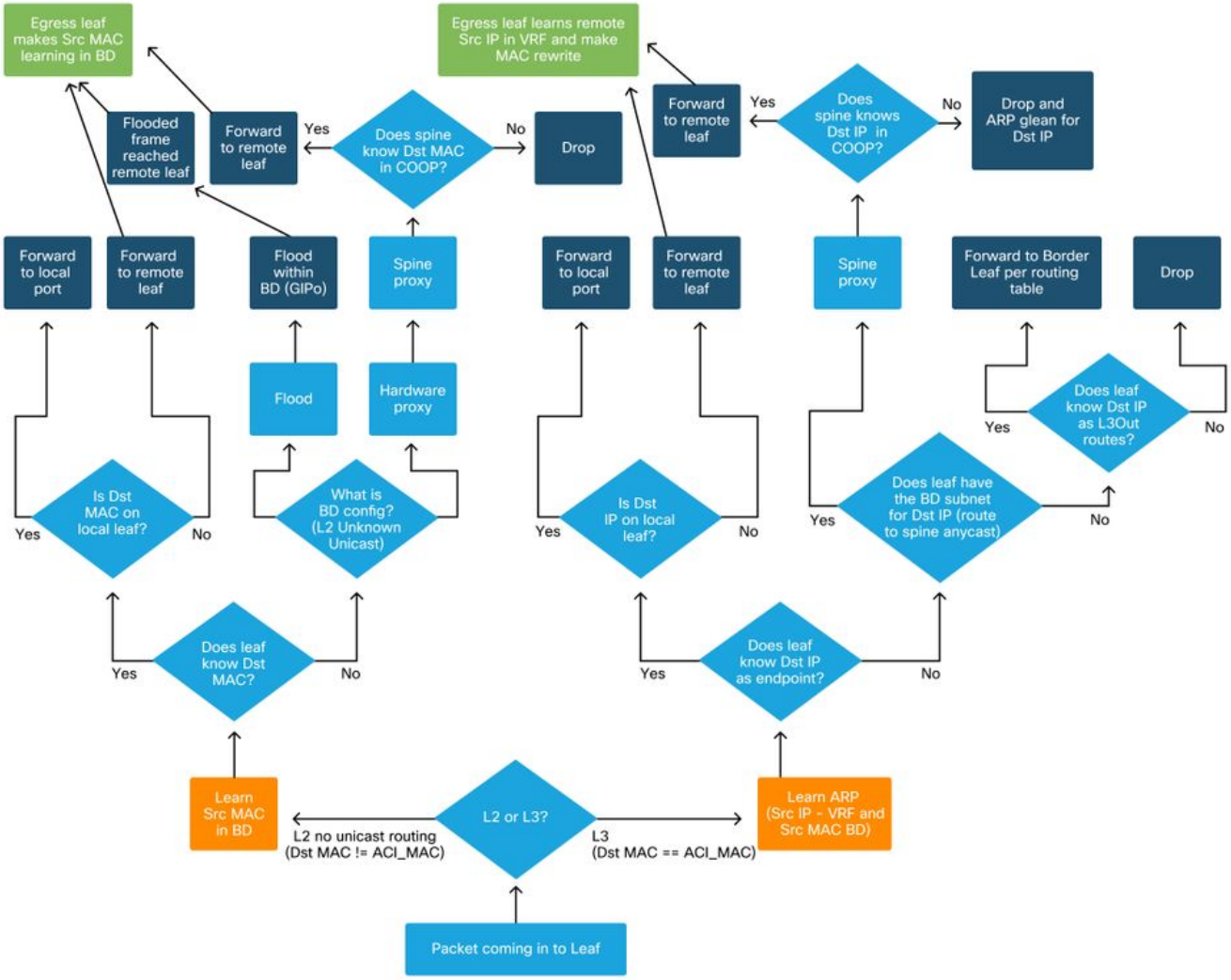
在此场景中，COOP知道目标IP，因此它将将VXLAN数据包中外部IP报头的目标IP重写为10.0.96.66，然后发送到leaf3或leaf4（取决于ECMP散列）。 请注意，VXLAN帧的源IP未更改，因此它仍然是leaf1 PTEP。

主干上的COOP查找 — 目标IP未知

如果目标IP的COOP条目未填充（静默终端或过期），主干将生成ARP收集器进行解析。有关详细信息，请参阅“多Pod转发”部分。

ACI转发摘要

下图总结了第2层和第3层的ACI转发。


Egress leaf makes Src MAC learning in BD
Egress leaf learns remote Src IP in VRF and make MAC rewrite
Flooded frame reached remote leaf
Forward to remote leaf
Yes
Does spine know Dst MAC in COOP?
No
Drop
Forward to remote leaf
Yes
Does spine knows Dst IP in COOP?
No
Drop and ARP glean for Dst IP
Forward to local port
Forward to remote leaf
Flood within BD (GIPo)
Spine proxy
Forward to local port
Forward to remote leaf
Spine proxy
Forward to Border Leaf per routing table
Drop
Flood
Hardware proxy
Does leaf know Dst IP as L3Out routes?
Yes
No
Is Dst MAC on local leaf?
Yes
No
What is BD config? (L2 Unknown Unicast)
Is Dst IP on local leaf?
Yes
No
Does leaf have the BD subnet for Dst IP (route to spine anycast)
Yes
No
Does leaf know Dst MAC?
Yes
No
Does leaf know Dst IP as endpoint?
Yes
No
Learn Src MAC in BD
L2 or L3?
L2 no unicast routing
(Dst MAC != ACI_MAC)
L3
(Dst MAC == ACI_MAC)
Learn ARP (Src IP - VRF and Src MAC BD)
Packet coming in to Leaf

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）。