# 排除ACI外部转发故障

## 目录

# 简介

本文档介绍了解ACI中的L3out并对其进行故障排除的步骤

# 背景信息

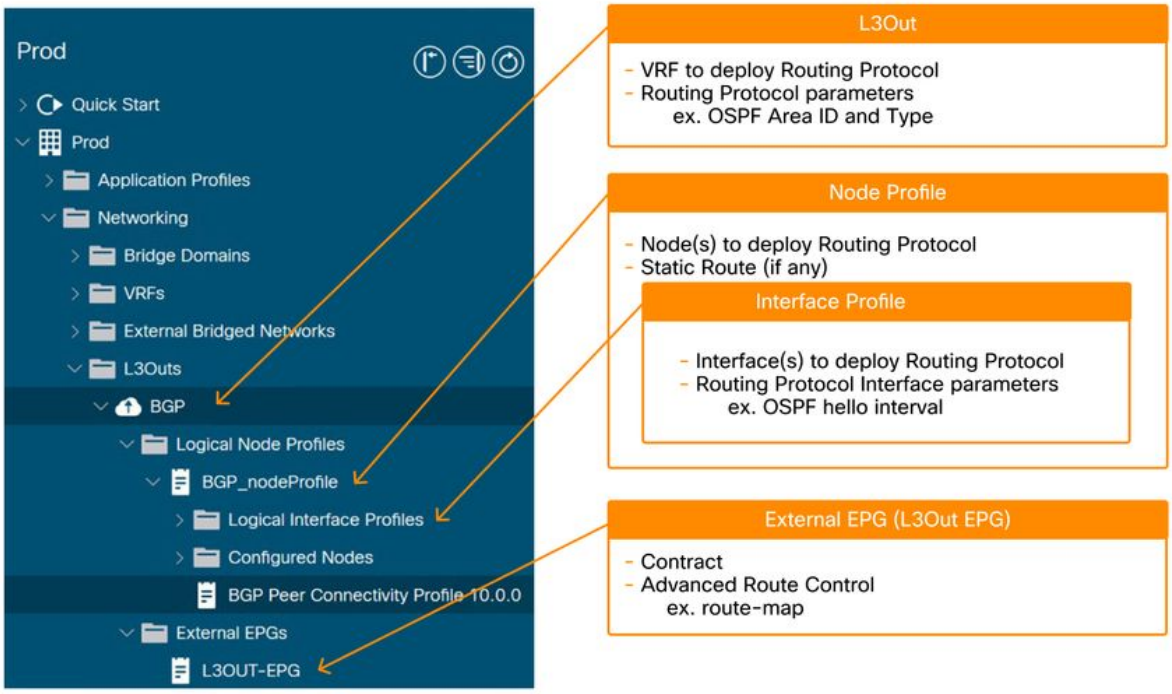
本文档中的内容摘自Cisco Application Centric Infrastructure第二版的故障排除，特别是**External Forwarding - Overview， External Forwarding - Adjacency， External Forwarding - Route advertisement， External Forwarding - Contract and L3out** 和**External Forwarding - Share L3out章**。

# 概述

## L3Out组件

下图显示了配置L3外部(L3Out)所需的主要构建基块。

## L3Out的主要组件

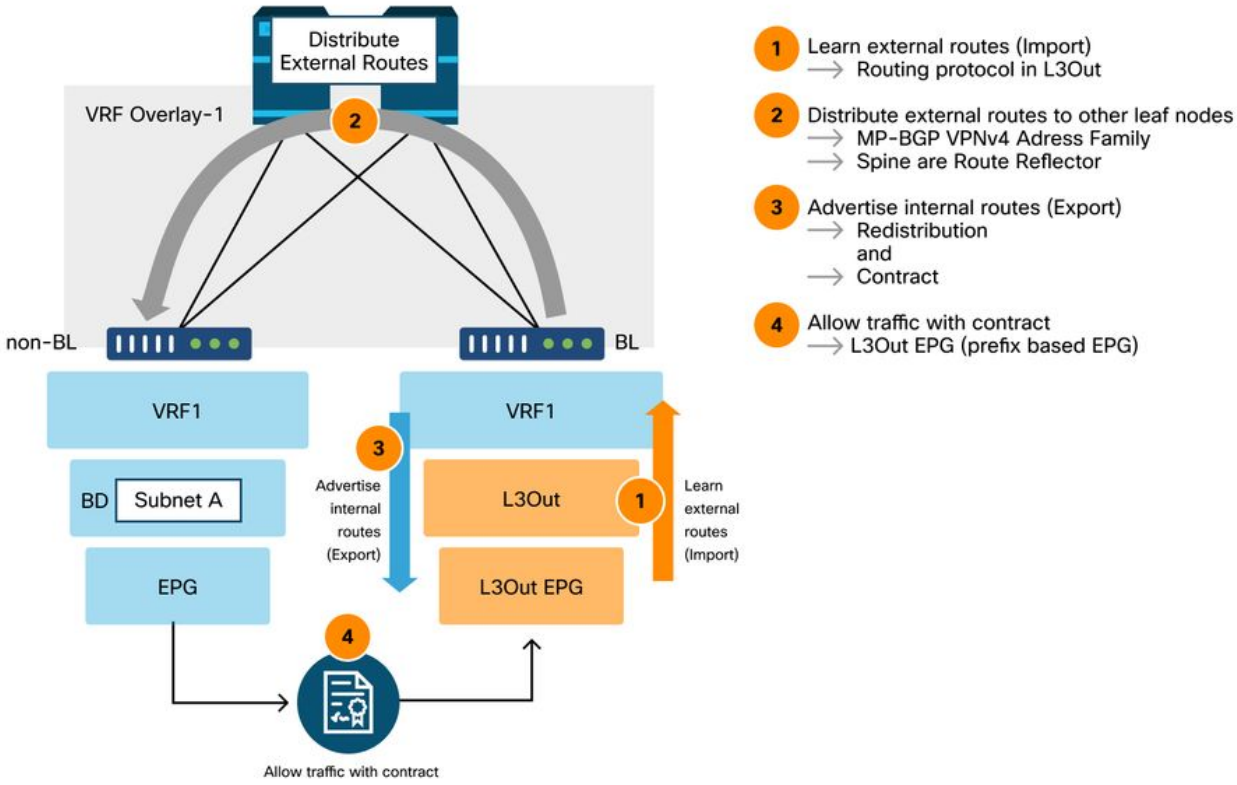



1. L3Out的根： 选择要部署的路由协议（例如OSPF、BGP）。选择VRF以部署路由协议。选择L3Out域以定义L3Out的可用枝叶接口和VLAN。
2. 节点配置文件： 选择枝叶交换机以部署路由协议。这些交换机通常称为“边界枝叶交换机”(BL)。为每个边界枝叶上的路由协议配置路由器ID(RID)。与普通路由器不同，ACI不会根据交换机上的IP地址自动分配路由器ID。配置静态路由。
3. 接口配置文件： 配置枝叶接口以运行路由协议。
即接口类型（SVI、路由端口、子接口）、接口ID和IP地址等。为接口级路由协议参数（例如hello间隔）选择策略。
4. 外部EPG(L3Out EPG): “外部EPG”是部署与L3Out关联的所有策略（例如IP地址或SVI）以建立邻居的硬性要求。稍后将介绍有关如何使用外部EPG的详细信息。

## 外部路由

下图显示外部工艺路线所涉及的高级工序。

## 高级外部路由流

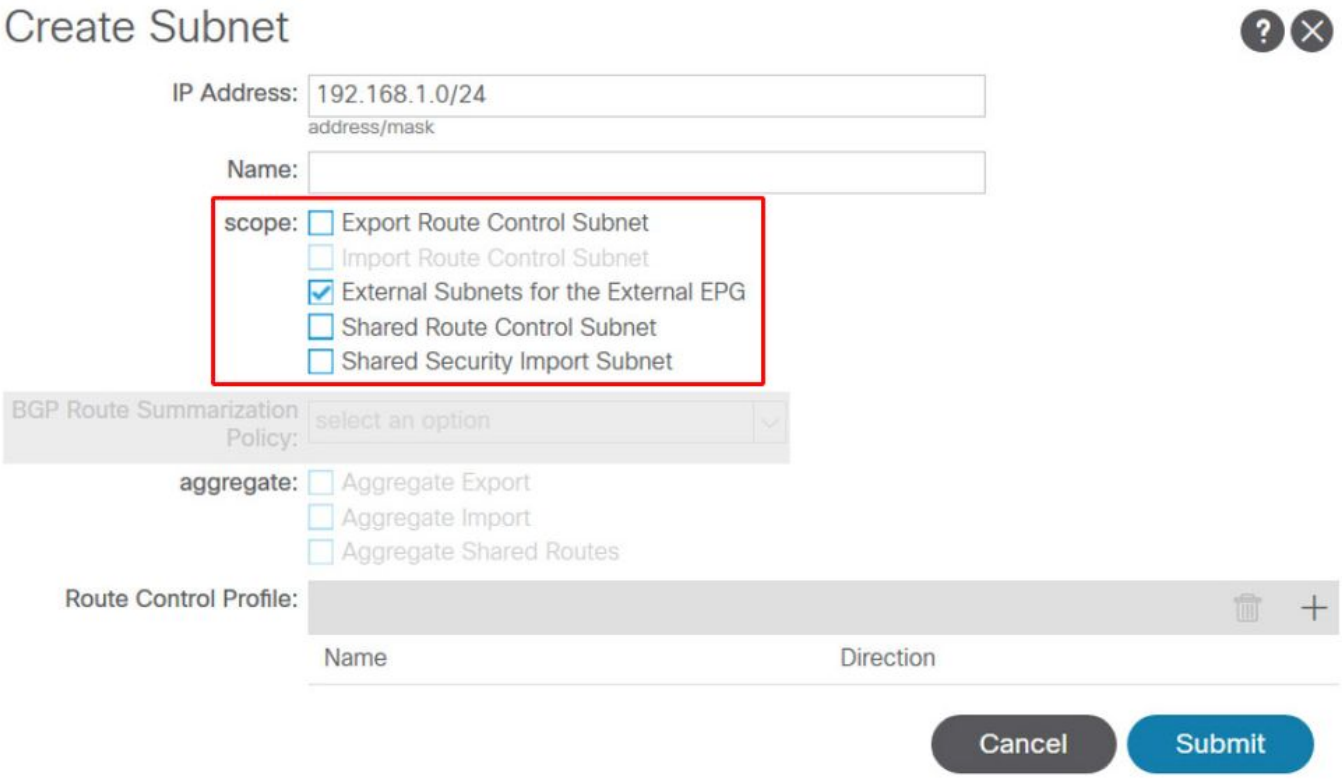



1. BL将与外部路由器建立路由协议邻接。
2. 从外部路由器接收路由前缀，并将其作为VPNv4地址系列路径注入到MP-BGP中。 至少必须将两个主干节点配置为BGP路由反射器，以便将外部路由反射到所有枝叶节点。
3. 从其他L3Out收到的内部前缀（BD子网）和/或前缀必须明确重分发到路由协议中才能通告给外部路由器。
4. 安全实施：l3Out至少包含一个L3Out EPG。必须在L3Out EPG上使用或提供合同（从类名称也称为l3extInstP），以允许流量进出L3Out。

## L3Out EPG配置选项

在L3Out EPG部分中，可以使用一系列“范围”和“聚合”选项定义子网，如下所示：

## 定义的L3Out子网包括“范围”定义




“范围”选项：

- **导出路由控制子网**：此范围是通过L3Out将子网从ACI通告（导出）到外部。虽然主要用于传输路由，但也可以用于通告BD子网，如“ACI BD子网通告”一节所述。
- **导入路由控制子网**：此范围涉及从L3Out学习（导入）外部子网。默认情况下，此范围处于禁用状态，因此呈灰色显示，并且边界枝叶(BL)从路由协议获取任何路由。当需要限制通过OSPF和BGP获知的外部路由时，可以启用此范围。EIGRP不支持此功能。要使用此范围，需要在给定L3Out上首先启用“导入路由控制实施”。
- **外部EPG的外部子网(import-security)：此作用域用于允许具有已配置子网的数据包通过合同从**L3Out或到L3Out。它根据子网将数据包分类到已配置的L3Out EPG，以便L3Out EPG上的合同可以应用于数据包。此范围是最长前缀匹配，而不是像路由表的其他范围那样完全匹配。如果在L3Out EPG A中为10.0.0.0/16配置了“外部EPG的外部子网”，则在该子网中具有IP的任何数据包（例如10.0.1.1）都将分类到L3Out EPG A中，以在其上使用合同。这并不意味着“外部EPG的外部子网”作用域将路由10.0.0.0/16安装在路由表中。它将创建一个不同的内部表，以便仅根据合同将子网映射到EPG(pcTag)。它对路由协议行为没有任何影响。此范围要在学习子网的L3Out上配置。
- **共享路由控制子网**：此范围是将外部子网泄漏到另一个VRF。ACI使用MP-BGP和路由目标将外部路由从一个VRF泄漏到另一个VRF。此作用域会创建带有子网的前缀列表，该前缀列表用作过滤器，用于导出/导入MP-BGP中具有路由目标的路由。此范围要在学习原始VRF中的子网的L3Out上配置。
- **共享安全导入子网**：此范围用于在数据包通过L3Out的VRF传输时，允许具有已配置子网的数据包。路由表中的路由会泄漏给另一个具有上述“共享路由控制子网”的VRF。但是，另一个VRF尚未知道泄漏的路由应属于哪个EPG。“共享安全导入子网”通知另一个VRF泄漏的路由所属的L3Out EPG。因此，仅当还使用“外部EPG的外部子网”时，才能使用此范围，否则原始VRF不知道该子网属于哪个L3Out EPG。此范围也是最长前缀匹配。

“聚合”选项：

- **Aggregate Export:**此选项只能用于0.0.0.0/0和“Export Route Control Subnet”。 当0.0.0.0/0同时

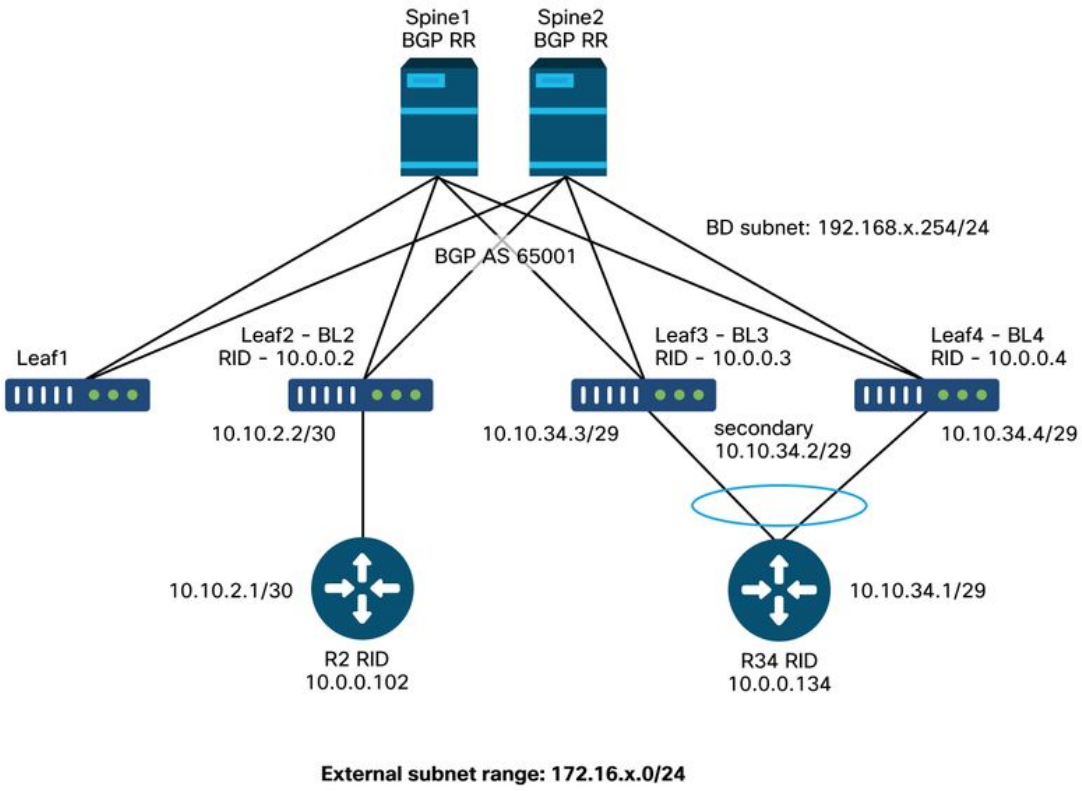
启用“导出路由控制子网”和“聚合导出”时；它会创建一个前缀列表并使用“0.0.0.0/0 le 32”匹配所有子网。因此，当L3Out需要向外部通告（导出）任何路由时，可以使用此选项。当需要更精细的聚合时，可以使用具有显式前缀列表的路由映射/配置文件。

- **Aggregate Import:**此选项只能用于0.0.0.0/0和“Import Route Control Subnet”。 当0.0.0.0/0同时启用“导入路由控制子网”和“聚合导入”时，它会创建前缀列表，其中包含“0.0.0.0/0 le 32”，该前缀列表匹配任何子网。因此，当L3Out需要从外部学习（导入）任何路由时，可以使用此选项。但是，通过禁用默认的“导入路由控制实施”，也可以完成同样的事情。当需要更精细的聚合时，可以使用具有显式前缀列表的路由映射/配置文件。
- **聚合共享路由**：此选项可用于具有“共享路由控制子网”的任何子网。 例如，如果同时为10.0.0.0/8启用了“共享路由控制子网”和“聚合共享路由”，则会创建前缀列表，其中包含“10.0.0.0/8 le 32”，该前缀列表与10.0.0.0/8、10.1.0.0/16等匹配。

### 本节中使用的L3Out拓扑

本节将使用以下拓扑：

### L3Out拓扑




# 邻接关系

本节介绍如何对L3Out接口上的路由协议邻接关系进行故障排除和验证。

下面是一些适用于所有ACI外部路由协议的检查参数：

- **路由器ID**：在ACI中，即使路由协议不同，每个L3Out也需要在同一枝叶上的相同VRF中使用相

同的路由器ID。此外，同一枝叶上只有一个L3Outs可以使用路由器ID（通常为BGP）创建环回。

- **MTU：虽然MTU的硬性要求仅适用于OSPF邻接，但建议匹配所有路由协议的MTU，以确保用于路由交换/更新的任何巨型数据包无需分段即可传输，因为大多数控制平面帧将使用DF（不要分段）位集发送，如果帧的大小超过接口的最大MTU，该位将丢弃该帧。**
- **MP-BGP路由器反射器**：如果不这样做，BGP进程不会在枝叶节点上启动。虽然OSPF或EIGRP仅建立邻居并不需要这样做，但是BL仍需要将外部路由分发到其他枝叶节点。
- **故障**：请务必在配置期间和完成之后检查故障。

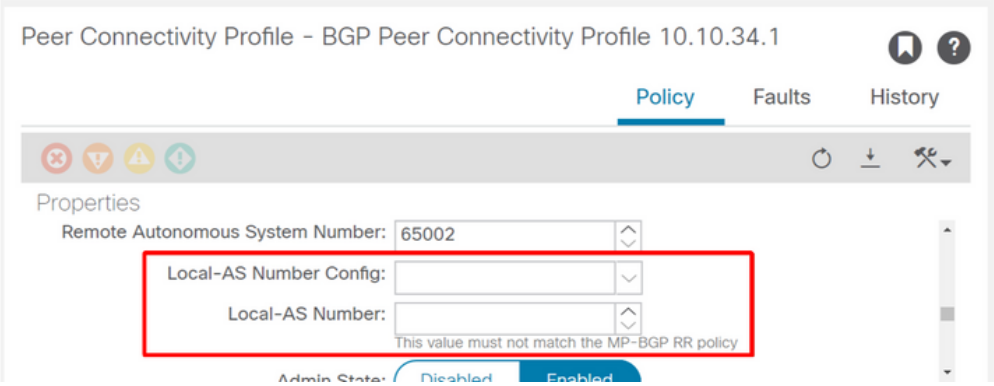
## 调试输出中显示“BGP

本节使用“概述”部分拓扑中BL3、BL4和R34上的环回之间的eBGP对等示例。R34上的BGP AS为65002。

建立BGP邻接关系时，请验证以下条件。

- 本地BGP AS编号（ACI BL端）。

## 对等连接配置文件 — Local-AS



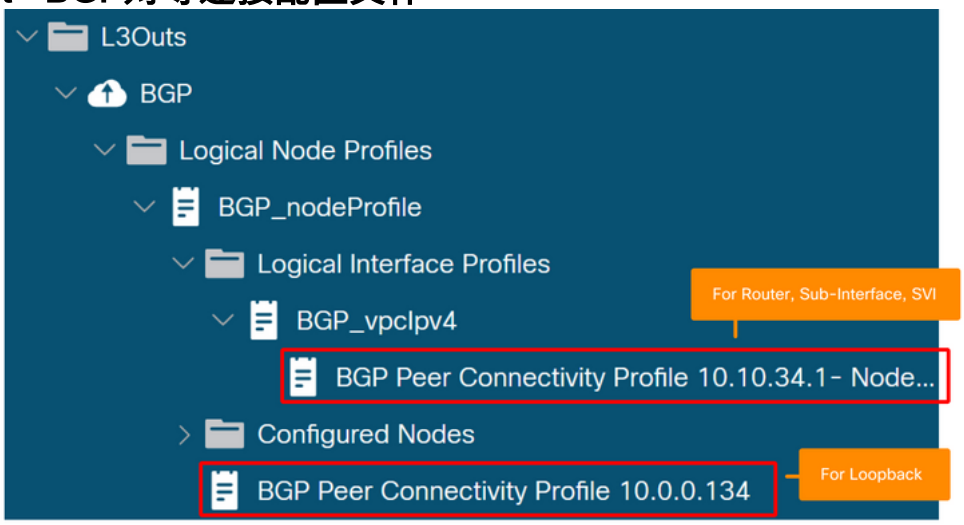

用户L3Out的BGP AS编号将自动与BGP路由反射器策略中配置的infra-MP-BGP的BGP AS相同。除非需要将ACI BGP AS伪装到外部，否则不需要在BGP对等连接配置文件中配置“本地AS”。这意味着外部路由器应指向BGP路由反射器中配置的BGP AS。

注意 — 需要本地AS配置的场景与独立NX-OS“local-as”命令相同。

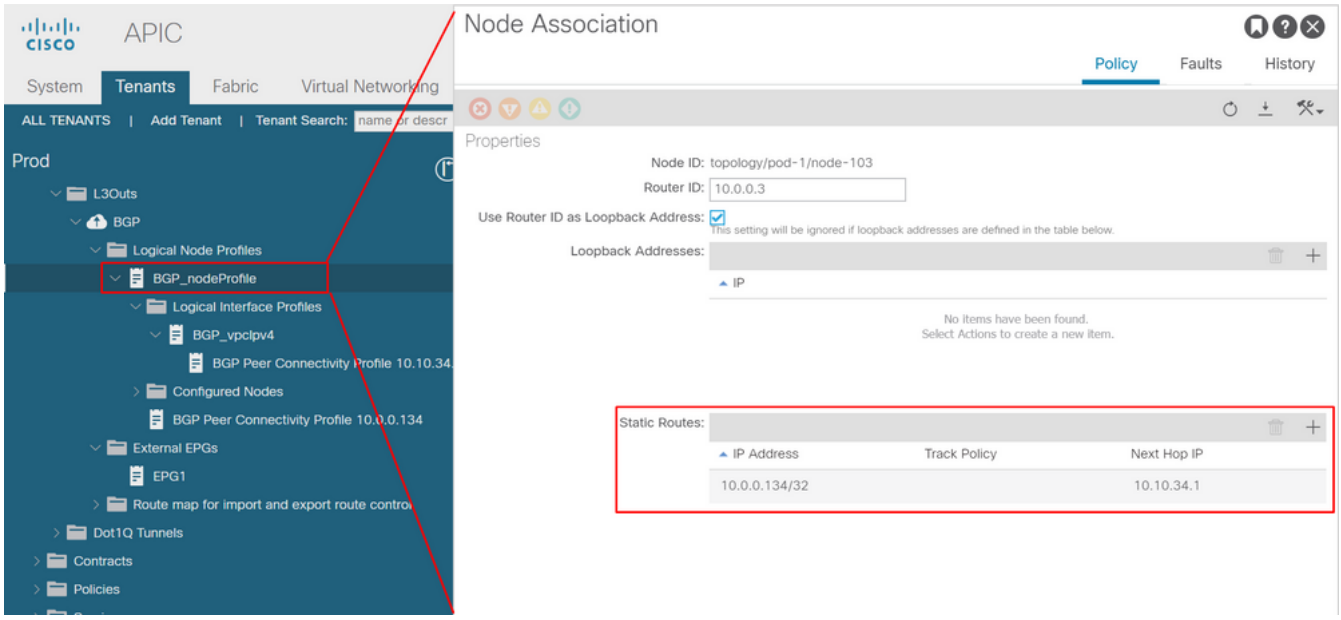
- 远程BGP AS编号（外部）**对等连接配置文件 — 远程AS**


仅当邻居的BGP AS不同于ACI BGP AS的eBGP时才需要远程BGP AS编号。BGP对等会话的

源IP。**L3Out - BGP对等连接配置文件**




ACI支持从典型ACI L3Out接口类型（路由、子接口、SVI）顶部的环回接口获取BGP会话。当需要从环回发起BGP会话时，请在逻辑节点配置文件下配置BGP对等连接**配置**文件。当BGP会话需要源自路由/子接口/SVI时，请在Logical **Interface** Profile下配置BGP对等连接配置**文**件。BGP对等IP可达性。**逻辑节点配置文件 — 节点关联**




当BGP对等IP是环回接口时，请确保BL和外部路由器可访问对等设备的IP地址。静态路由或OSPF可用于获得对等IP的可达性。**BGP CLI验证（带环回的eBGP示例）**以下步骤的CLI输出是从拓扑结构中的BL3的“概述”部分收集的。**1.检查BGP会话是否已建立**以下CLI输出中的“State/PfxRcd”表示已建立BGP会话。

```
f2-leaf3# show bgp ipv4 unicast summary vrf Prod:VRF1
BGP summary information for VRF Prod:VRF1, address family IPv4 Unicast
BGP router identifier 10.0.0.3, local AS number 65001

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.0.0.134      4 65002      10      10       10    0    0 00:06:39 0
```

如果“State/PfxRcd”显示“空闲”或“活动”，则尚未与对等设备交换BGP数据包。在这种情况下，请检查以下内容并进入下一步。

- 确保外部路由器正确指向ACI BGP AS(本地AS编号65001)。
- 确保ACI BGP对等连接配置文件指定了外部路由器从中获取BGP会话(10.0.0.134)的正确邻居IP。
- 确保ACI BGP对等连接配置文件指定了外部路由器的正确邻居AS(GUI中的远程自治系统编号，在CLI中显示为AS 65002)。

### 2.检查BGP邻居详细信息（BGP对等连接配置文件）

以下命令显示BGP邻居建立的关键参数。

- 邻居IP:10.0.0.134 的多播地址发送一次邻居消息。
- 邻居BGP AS:远程AS 65002。
- 源 IP：使用loopback3作为更新源。
- eBGP多跳：外部BGP对等体可能距离最多2跳。

```
f2-leaf3# show bgp ipv4 unicast neighbors vrf Prod:VRF1
BGP neighbor is 10.0.0.134,  remote AS 65002, ebgp link,  Peer index 1
 BGP version 4, remote router ID 10.0.0.134
 BGP state = Established, up for 00:11:18
 Using loopback3 as update source for this peer
 External BGP peer might be upto 2 hops away

...

  For address family: IPv4 Unicast
...
Inbound route-map configured is permit-all, handle obtained
Outbound route-map configured is exp-l3out-BGP-peer-3047424, handle obtained
Last End-of-RIB received 00:00:01 after session start
Local host: 10.0.0.3, Local port: 34873
 Foreign host: 10.0.0.134, Foreign port: 179
 fd = 64
```

一旦BGP对等体正确建立，“本地主机”和“外部主机”将显示在输出的底部。

### 3.检查BGP对等体的IP连通性

```
f2-leaf3# show ip route vrf Prod:VRF1
10.0.0.3/32, ubest/mbest: 2/0, attached, direct
   *via 10.0.0.3, lo3, [0/0], 02:41:46, local, local
   *via 10.0.0.3, lo3, [0/0], 02:41:46, direct
10.0.0.134/32, ubest/mbest: 1/0
   *via 10.10.34.1, vlan27, [1/0], 02:41:46, static    <--- neighbor IP reachability via static
route
10.10.34.0/29, ubest/mbest: 2/0, attached, direct
   *via 10.10.34.3, vlan27, [0/0], 02:41:46, direct
   *via 10.10.34.2, vlan27, [0/0], 02:41:46, direct
10.10.34.2/32, ubest/mbest: 1/0, attached
   *via 10.10.34.2, vlan27, [0/0], 02:41:46, local, local
10.10.34.3/32, ubest/mbest: 1/0, attached
   *via 10.10.34.3, vlan27, [0/0], 02:41:46, local, local
```

确保从ACI BGP的源IP对邻居IP执行ping操作。

```
f2-leaf3# iping 10.0.0.134 -V Prod:VRF1 -S 10.0.0.3
```

```
PING 10.0.0.134 (10.0.0.134) from 10.0.0.3: 56 data bytes
64 bytes from 10.0.0.134: icmp_seq=0 ttl=255 time=0.571 ms
64 bytes from 10.0.0.134: icmp_seq=1 ttl=255 time=0.662 ms
```

### 4.检查外部路由器上的相同内容

以下是外部路由器（独立NX-OS）上的配置示例。

```
router bgp 65002
 vrf f2-bgp
   router-id 10.0.0.134
   neighbor 10.0.0.3
     remote-as 65001
     update-source loopback134
     ebgp-multihop 2
     address-family ipv4 unicast
   neighbor 10.0.0.4
     remote-as 65001
     update-source loopback134
     ebgp-multihop 2
     address-family ipv4 unicast

interface loopback134
 vrf member f2-bgp
 ip address 10.0.0.134/32

interface Vlan2501
 no shutdown
 vrf member f2-bgp
 ip address 10.10.34.1/29

vrf context f2-bgp
 ip route 10.0.0.0/29 10.10.34.2
```

### 5.额外步骤 — tcpdump

在ACI枝叶节点上，tcpdump工具可以嗅探“kpm_inb”CPU接口，以确认协议数据包是否到达枝叶CPU。使用L4端口179(BGP)作为过滤器。

```
f2-leaf3# tcpdump -ni kpm_inb port 179
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on kpm_inb, link-type EN10MB (Ethernet), capture size 65535 bytes
20:36:58.292903 IP 10.0.0.134.179 > 10.0.0.3.34873: Flags [P.], seq 3775831990:3775832009, ack
807595300, win 3650, length 19: BGP, length: 19
20:36:58.292962 IP 10.0.0.3.34873 > 10.0.0.134.179: Flags [.], ack 19, win 6945, length 0
20:36:58.430418 IP 10.0.0.3.34873 > 10.0.0.134.179: Flags [P.], seq 1:20, ack 19, win 6945,
length 19: BGP, length: 19
20:36:58.430534 IP 10.0.0.134.179 > 10.0.0.3.34873: Flags [.], ack 20, win 3650, length 0
```

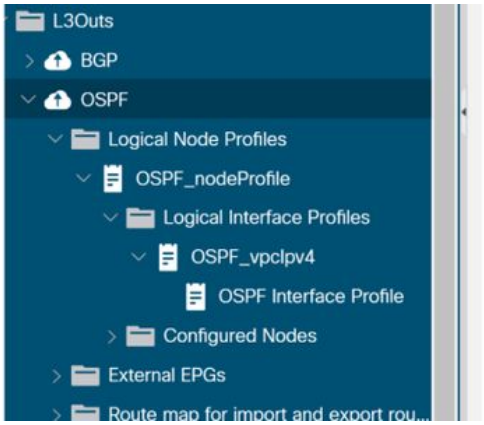
## OSPF

本部分使用OSPF区域ID 1(NSSA)的“概述”部分拓扑中BL3、BL4和R34之间的OSPF邻居关系示例。

以下是检查OSPF邻接关系建立的通用标准。

- OSPF区域ID和类型

## L3Out — OSPF接口配置文件 — 区域ID和类型





与任何路由设备一样，OSPF区域ID和类型需要在两个邻居上匹配。OSPF区域ID配置的一些特定于ACI的限制包括：

- 一个L3Out只能有一个OSPF区域ID。
- 仅当两个L3Outs位于两个不同的枝叶节点时，两个L3Outs才能在同一VRF中使用相同的OSPF区域ID。

虽然OSPF ID不需要是主干0，但在传输路由的情况下，同一枝叶上的两个OSPF L3Outs之间需要它；其中一个必须使用OSPF区域0，因为OSPF区域之间的任何路由交换都必须通过OSPF区域0。

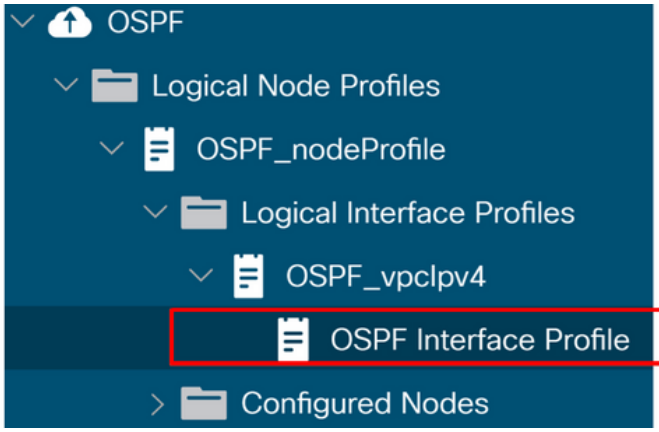
- MTU

## 逻辑接口配置文件 — SVI


ACI上的默认MTU为9000字节，而不是1500字节，后者通常用于传统路由设备。确保MTU与外部设备匹配。当OSPF邻居建立因MTU而失败时，它将停滞在EXCHANGE/DROTHER上。

- IP 子网掩码.OSPF要求邻居IP使用相同的子网掩码。
- OSPF接口配置文件。

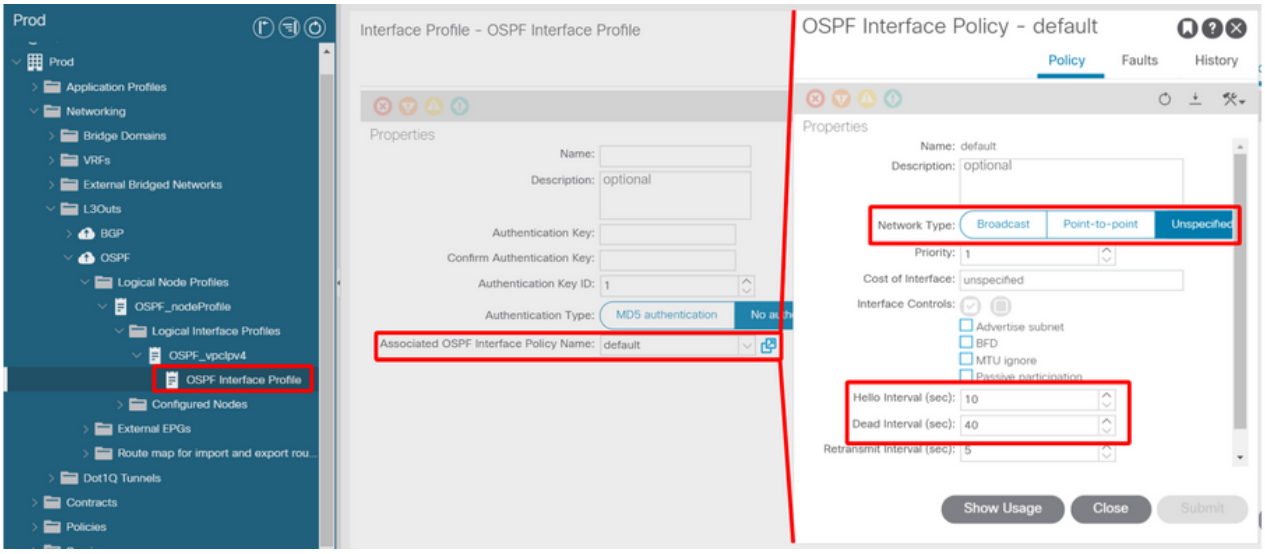
## OSPF接口配置文件



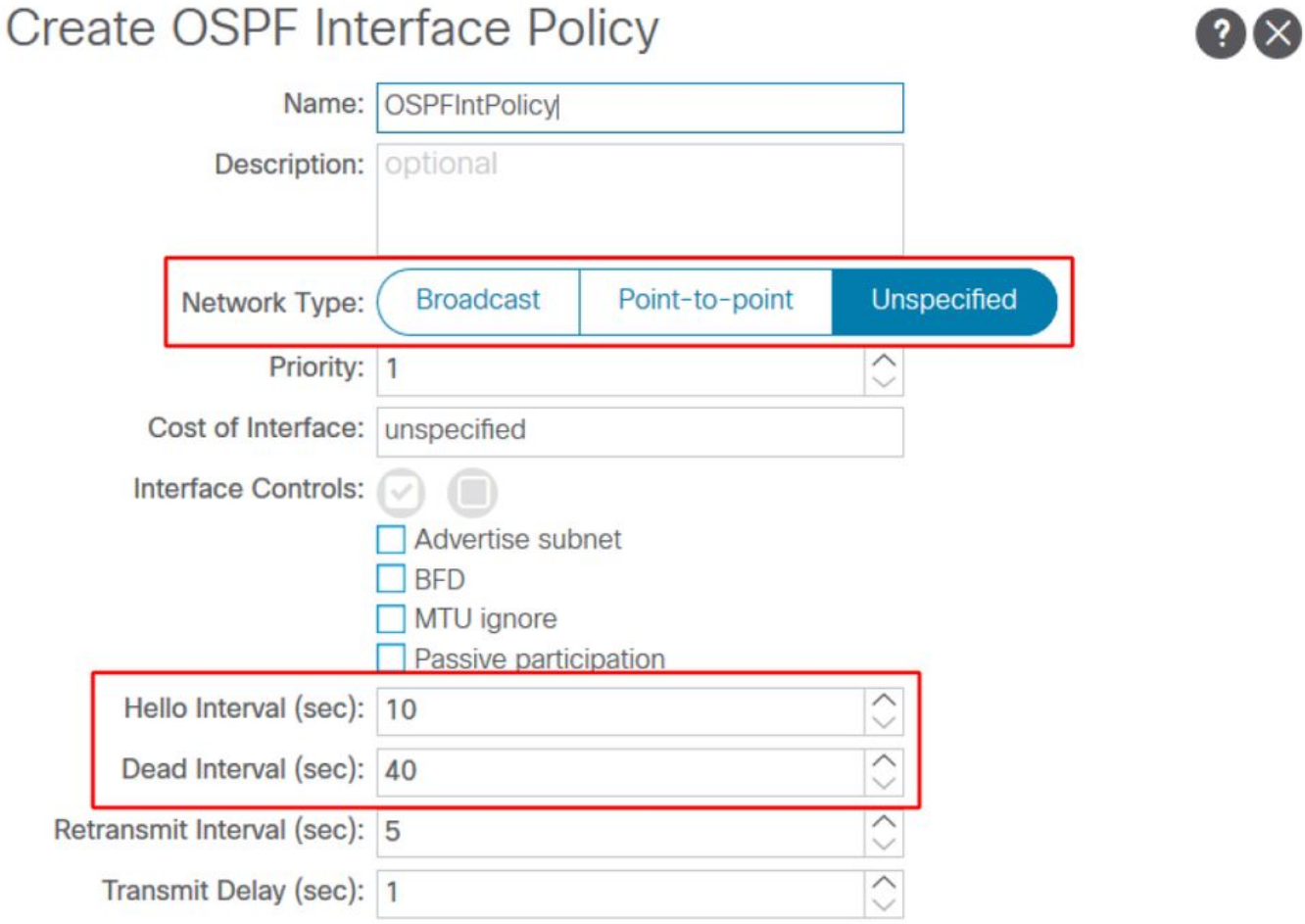

这相当于在独立NX-OS配置上启用OSPF的“ip router ospf <tag> area <area id>”。否则，枝叶接口不会加入OSPF。

- OSPF Hello/Dead计时器，网络类型

## OSPF接口配置文件 — Hello/Dead计时器和网络类型




## OSPF接口策略详细信息




OSPF要求每个邻居设备上的Hello计时器和Dead计时器匹配。这些配置在OSPF接口配置文件中。

确保OSPF接口网络类型与外部设备匹配。当外部设备使用点对点类型时，ACI端也需要显式配置点对点。OSPF接口配置文件中也配置了这些命令。

### OSPF CLI验证

以下步骤中的CLI输出是从拓扑结构中的BL3的“概述”部分收集的。

### 1.检查OSPF邻居状态

如果以下CLI中的“State”为“FULL”，则会正确建立OSPF邻居。否则，请继续下一步检查参数。

```
f2-leaf3# show ip ospf neighbors vrf Prod:VRF2
OSPF Process ID default VRF Prod:VRF2
Total number of neighbors: 2
Neighbor ID     Pri State            Up Time  Address         Interface
10.0.0.4          1 FULL/DR          00:47:30 10.10.34.4      Vlan28      <--- neighbor with BL4
10.0.0.134        1 FULL/DROTHER     00:00:21 10.10.34.1      Vlan28      <--- neighbor with R34
```

在ACI中，当将同一VLAN ID用于SVI时，BL将在外部路由器上彼此形成OSPF邻居关系。这是因为ACI具有称为L3Out BD（或外部BD）的内部泛洪域，适用于L3Out SVI中的每个VLAN ID。请注意，VLAN ID 28是一个称为PI-VLAN（独立于平台的VLAN）的内部VLAN，而不是用于线路的实际VLAN（接入封装VLAN）。使用以下命令验证访问封装VLAN('vlan-2502')。

```
f2-leaf3# show vlan id 28 extended
 VLAN Name                             Encap            Ports
 ---- -------------------------------- ---------------- ------------------------
 28   Prod:VRF2:l3out-OSPF:vlan-2502   vxlan-14942176,  Eth1/13, Po1
                                       vlan-2502
```

通过访问封装VLAN ID也可以获得相同的输出。

```
f2-leaf3# show vlan encap-id 2502 extended
 VLAN Name                             Encap            Ports
 ---- -------------------------------- ---------------- ------------------------
 28   Prod:VRF2:l3out-OSPF:vlan-2502   vxlan-14942176,  Eth1/13, Po1
                                       vlan-2502
```

### 2.检查OSPF区域

确保OSPF区域ID和类型与邻居相同。如果OSPF接口配置文件缺失，该接口不会加入OSPF，并且不会显示在OSPF CLI输出中。

```
f2-leaf3# show ip ospf interface brief vrf Prod:VRF2
OSPF Process ID default VRF Prod:VRF2
Total number of interface: 1
Interface               ID     Area            Cost   State    Neighbors Status
Vlan28                  94     0.0.0.1         4      BDR      2         up
f2-leaf3# show ip ospf vrf Prod:VRF2
Routing Process default with ID 10.0.0.3 VRF Prod:VRF2
...
  Area (0.0.0.1)
       Area has existed for 00:59:14
       Interfaces in this area: 1 Active interfaces: 1
       Passive interfaces: 0  Loopback interfaces: 0
       This area is a NSSA area
       Perform type-7/type-5 LSA translation
       SPF calculation has run 10 times
        Last SPF ran for 0.001175s
       Area ranges are
       Area-filter in 'exp-ctx-proto-3112960'
       Area-filter out 'permit-all'
       Number of LSAs: 4, checksum sum 0x0
```

### 3.检查OSPF接口详细信息

确保接口级别参数符合OSPF邻居建立要求，例如IP子网、网络类型、Hello/Dead计时器。请注意VLAN ID以指定SVI为PI-VLAN(vlan28)

```
f2-leaf3# show ip ospf interface vrf Prod:VRF2
Vlan28 is up, line protocol is up
   IP address 10.10.34.3/29, Process ID default VRF Prod:VRF2, area 0.0.0.1
   Enabled by interface configuration
   State BDR, Network type BROADCAST, cost 4
   Index 94, Transmit delay 1 sec, Router Priority 1
   Designated Router ID: 10.0.0.4, address: 10.10.34.4
   Backup Designated Router ID: 10.0.0.3, address: 10.10.34.3
   2 Neighbors, flooding to 2, adjacent with 2
   Timer intervals: Hello 10, Dead 40, Wait 40, Retransmit 5
     Hello timer due in 0.000000
```

```
   No authentication
   Number of opaque link LSAs: 0, checksum sum 0

f2-leaf3# show interface vlan28
Vlan28 is up, line protocol is up, autostate disabled
  Hardware EtherSVI, address is  0022.bdf8.19ff
  Internet Address is 10.10.34.3/29
  MTU 9000 bytes, BW 10000000 Kbit, DLY 1 usec
```

### 4.检查与邻居的IP连通性

虽然OSPF Hello数据包是链路本地组播数据包，但第一个OSPF LSDB交换所需的OSPF DBD数据包是单播数据包。因此，对于OSPF邻居关系的建立，还需要验证单播可达性。

```
f2-leaf3# iping 10.10.34.1 -V Prod:VRF2
PING 10.10.34.1 (10.10.34.1) from 10.10.34.3: 56 data bytes
64 bytes from 10.10.34.1: icmp_seq=0 ttl=255 time=0.66 ms
64 bytes from 10.10.34.1: icmp_seq=1 ttl=255 time=0.653 ms
```

### 5.在外部路由器上检查相同内容

以下是外部路由器（独立NX-OS）上的配置示例

```
router ospf 1
    vrf f2-ospf
    router-id 10.0.0.134
    area 0.0.0.1 nssa

interface Vlan2502
    no shutdown
    mtu 9000
    vrf member f2-ospf
    ip address 10.10.34.1/29
    ip router ospf 1 area 0.0.0.1
```

确保也在物理接口上验证MTU。

### 6.额外步骤 — tcpdump

在ACI枝叶节点上，用户可以在“kpm_inb”CPU接口上执行tcpdump，以验证协议数据包是否已到达枝叶CPU。虽然OSPF有多个过滤器，但IP协议号是最全面的过滤器。

- IP协议号：proto 89(IPv4)或ip6 proto 0x59(IPv6)
- 邻居的IP地址：host <ip>
- OSPF本地链路组播IP:host 224.0.0.5或host 224.0.0.6

```
f2-leaf3# tcpdump -ni kpm_inb proto 89
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on kpm_inb, link-type EN10MB (Ethernet), capture size 65535 bytes
22:28:38.231356 IP 10.10.34.4 > 224.0.0.5: OSPFv2, Hello, length 52
22:28:42.673810 IP 10.10.34.3 > 224.0.0.5: OSPFv2, Hello, length 52
22:28:44.767616 IP 10.10.34.1 > 224.0.0.5: OSPFv2, Hello, length 52
22:28:44.769092 IP 10.10.34.3 > 10.10.34.1: OSPFv2, Database Description, length 32
22:28:44.769803 IP 10.10.34.1 > 10.10.34.3: OSPFv2, Database Description, length 32
22:28:44.775376 IP 10.10.34.3 > 10.10.34.1: OSPFv2, Database Description, length 112
```

```
22:28:44.780959 IP 10.10.34.1 > 10.10.34.3: OSPFv2, LS-Request, length 36
22:28:44.781376 IP 10.10.34.3 > 10.10.34.1: OSPFv2, LS-Update, length 64
22:28:44.790931 IP 10.10.34.1 > 224.0.0.6: OSPFv2, LS-Update, length 64
```

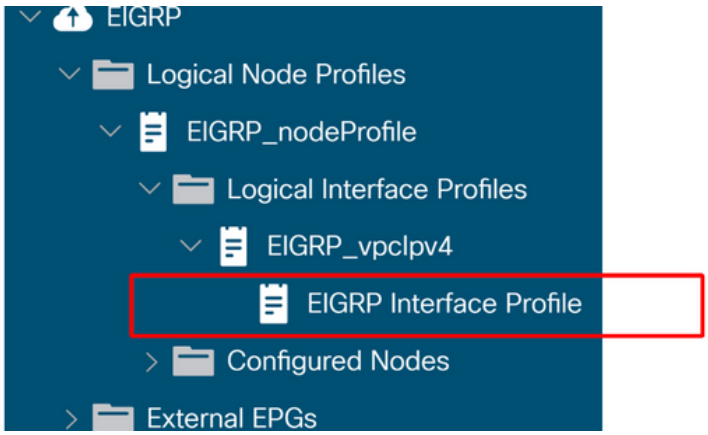
## EIGRP

本部分使用EIGRP AS 10的“概述”部分拓扑中的BL3、BL4和R34之间的EIGRP邻居关系示例。

以下是建立EIGRP邻接关系的常用标准。

- EIGRP AS:为L3Out分配了一个EIGRP AS。这必须与外部设备匹配。
- EIGRP接口配置文件。

## EIGRP接口配置文件




这相当于在独立NX-OS设备上配置“ip router eigrp <as>”。否则，枝叶接口不会加入EIGRP。

- MTU

尽管这不必匹配以简单建立EIGRP邻居关系，但EIGRP拓扑交换数据包可能会大于对等体之间接口上允许的最大MTU，并且由于这些数据包不允许分段，因此它们将被丢弃，因此EIGRP邻居关系将会抖动。

### EIGRP CLI验证

以下步骤中的CLI输出是从拓扑结构中的BL3的“概述”部分收集的。

#### 1.检查EIGRP邻居状态

```
f2-leaf3# show ip eigrp neighbors vrf Prod:VRF3
EIGRP neighbors for process 10 VRF Prod:VRF3
H   Address                 Interface       Hold  Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
0   10.10.34.4              vlan29           14   00:12:58  1    50    0   6    <--- neighbor
with BL4
1   10.10.34.1              vlan29           13   00:08:44  2    50    0   4    <--- neighbor
with R34
```

在ACI中，当外部路由器使用与SVI相同的VLAN ID时，BL将在外部路由器之上彼此形成EIGRP邻居关系。这是因为ACI具有称为L3Out BD（或外部BD）的内部泛洪域，适用于L3Out SVI中的每个

VLAN ID。

请注意，VLAN ID 29是称为PI-VLAN（独立于平台的VLAN）的内部VLAN，而不是有线上使用的实际VLAN（接入封装VLAN）。使用以下命令验证接入封装VLAN(vlan-2503)。

```
f2-leaf3# show vlan id 29 extended
 VLAN Name                            Encap            Ports
 ---- ------------------------------- ---------------- ------------------------
 29   Prod:VRF3:l3out-EIGRP:vlan-2503  vxlan-15237052,  Eth1/13, Po1
                                       vlan-2503
```

通过访问封装VLAN ID也可以获得相同的输出。

```
f2-leaf3# show vlan encap-id 2503 extended
 VLAN Name                            Encap            Ports
 ---- ------------------------------- ---------------- ------------------------
 29   Prod:VRF3:l3out-EIGRP:vlan-2503  vxlan-15237052,  Eth1/13, Po1
                                       vlan-2503
```

### 2.检查EIGRP接口详细信息

确保EIGRP在预期接口上运行。如果不是，请检查逻辑接口配置文件和EIGRP接口配置文件。

```
f2-leaf3# show ip eigrp interfaces vrf Prod:VRF3
EIGRP interfaces for process 10 VRF Prod:VRF3
                        Xmit Queue   Mean   Pacing Time   Multicast    Pending
Interface        Peers  Un/Reliable  SRTT   Un/Reliable   Flow Timer   Routes
vlan29             2        0/0         1       0/0          50             0
  Hello interval is 5 sec
  Holdtime interval is 15 sec
  Next xmit serial: 0
  Un/reliable mcasts: 0/2     Un/reliable ucasts: 5/10
  Mcast exceptions: 0    CR packets: 0    ACKs suppressed: 2
  Retransmissions sent: 2    Out-of-sequence rcvd: 0
  Classic/wide metric peers: 2/0
```

```
f2-leaf3# show int vlan 29
Vlan29 is up, line protocol is up, autostate disabled
  Hardware EtherSVI, address is  0022.bdf8.19ff
  Internet Address is 10.10.34.3/29
  MTU 9000 bytes, BW 10000000 Kbit, DLY 1 usec
```

### 3.在外部路由器上检查相同内容

以下为外部路由器（独立NX-OS）上的配置示例。

```
router eigrp 10
 vrf f2-eigrp

interface Vlan2503
 no shutdown
 vrf member f2-eigrp
 ip address 10.10.34.1/29
```

```
 ip router eigrp 10
```

### 4.额外步骤 — tcpdump

在ACI枝叶节点上，用户可以在“kpm_inb”CPU接口上执行tcpdump，以确认协议数据包是否到达枝叶CPU。使用IP协议编号88(EIGRP)作为过滤器。

```
f2-leaf3# tcpdump -ni kpm_inb proto 88
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on kpm_inb, link-type EN10MB (Ethernet), capture size 65535 bytes
23:29:43.725676 IP 10.10.34.3 > 224.0.0.10: EIGRP Hello, length: 40
23:29:43.726271 IP 10.10.34.4 > 224.0.0.10: EIGRP Hello, length: 40
23:29:43.728178 IP 10.10.34.1 > 224.0.0.10: EIGRP Hello, length: 40
23:29:45.729114 IP 10.10.34.1 > 10.10.34.3: EIGRP Update, length: 20
23:29:48.316895 IP 10.10.34.3 > 224.0.0.10: EIGRP Hello, length: 40
```

# 路由通告

本节重点介绍ACI中路由通告的验证和故障排除。具体来说，它查看的示例包括：

- 网桥域子网通告。
- 中转路由通告。
- 导入和导出路由控制。

本节讨论路由泄漏，因为它涉及共享L3Outs，将在后续章节中介绍。

## 网桥域路由通告工作流

在查看常见故障排除之前，用户应熟悉网桥域通告的运作原理。

当BD和L3Out在同一VRF中时，BD通告涉及：

- 在L3Out和内部EPG之间具有合同关系。
- 将L3Out关联到网桥域。
- 选择BD子网上的“Advertise External”。

此外，还可以使用导出路由配置文件控制网桥域通告，从而避免关联L3Out。但是，仍应选择“Advertise External”。这是一个不太常见的用例，因此将不在此讨论。

L3Out和EPG之间需要合同关系，以使BD普及静态路由推送到BL。实际的路由通告是通过将静态路由重分发到外部协议来处理的。最后，重分发路由映射将仅安装在与BD关联的L3Outs中。这样，路由不会通告给所有L3Outs。

在这种情况下，BD子网是192.168.1.0/24，应该通过OSPF L3Out通告该子网。

### 在应用L3Out和内部EPG之间的合同之前

```
leaf103# show ip route 192.168.1.0/24 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]
```

```
'%' in via output denotes VRF
Route not found
```

请注意，BD路由尚未出现在BL上。

### 在应用L3Out和内部EPG之间的合同后

此时尚未进行其他配置。L3Out尚未与BD关联，并且未设置“Advertise External”标志。

```
leaf103# show ip route 10.0.1.0/24 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]
'%' in via output denotes VRF
192.168.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive
   *via 10.0.120.34%overlay-1, [1/0], 00:00:08, static, tag 4294967294
        recursive next hop: 10.0.120.34/32%overlay-1
```

请注意，BD子网路由（由沉浸式标志表示）现在已部署在BL上。但请注意，该路由已标记。此标记值是在配置为“Advertise External”之前分配给BD路由的隐式值。 所有外部协议都拒绝重新分发此标记。

### 在BD子网中选择“Advertise External”后

L3Out尚未与BD关联。但请注意，标记已清除。

```
leaf103# show ip route 192.168.1.0/24 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]
'%' in via output denotes VRF
192.168.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive *via 10.0.120.34%overlay-1, [1/0],
00:00:06, static recursive next hop: 10.0.120.34/32%overlay-1
```

此时，路由仍然没有对外通告，因为没有路由映射和前缀列表匹配此前缀以重分发到外部协议。这可以通过以下命令进行验证：

```
leaf103# show ip ospf vrf Prod:Vrf1
Routing Process default with ID 10.0.0.3 VRF Prod:Vrf1
Stateful High Availability enabled
Supports only single TOS(TOS0) routes
Supports opaque LSA
Table-map using route-map exp-ctx-2392068-deny-external-tag
Redistributing External Routes from
  static route-map exp-ctx-st-2392068
  direct route-map exp-ctx-st-2392068
  bgp route-map exp-ctx-proto-2392068
  eigrp route-map exp-ctx-proto-2392068
  coop route-map exp-ctx-st-2392068
```

BD路由被编程为静态路由，因此请通过运行“show route-map <route-map name>”，然后在路由映射中存在的任何前缀列表上运行“show ip prefix-list <name>”来检查静态重分发路由映射。在下一步

中执行此操作。

### 将L3Out关联到BD之后

如前所述，此步骤会导致前缀列表与静态到外部协议重分发路由映射中安装的BD子网匹配。

```
leaf103# show route-map exp-ctx-st-2392068
route-map exp-ctx-st-2392068, deny, sequence 1
 Match clauses:
   tag: 4294967294
 Set clauses:

...
route-map exp-ctx-st-2392068, permit, sequence 15803
 Match clauses:
   ip address prefix-lists: IPv4-st16390-2392068-exc-int-inferred-export-dst
   ipv6 address prefix-lists: IPv6-deny-all
 Set clauses:
   tag 0
```

验证前缀列表：

```
leaf103# show ip prefix-list IPv4-st16390-2392068-exc-int-inferred-export-dst
ip prefix-list IPv4-st16390-2392068-exc-int-inferred-export-dst: 1 entries
  seq 1 permit 192.168.1.1/24
```

正在匹配BD子网以重分发到OSPF。

此时，从L3Out通告BD子网的配置和验证工作流程已完成。在此之后，验证将特定于协议。例如：

- 对于EIGRP，使用“show ip eigrp topology vrf <name>”检验是否正在拓扑表中安装路由
- 对于OSPF，使用“show ip ospf database vrf <name>”验证路由作为外部LSA安装在数据库表中
- 对于BGP，使用“show bgp ipv4 unicast vrf <name>”验证路由是否在BGP RIB中

### BGP 路由通告

对于BGP，隐式允许所有静态路由进行重分发。与BD子网匹配的路由映射应用于BGP邻居级别。

```
leaf103# show bgp ipv4 unicast neighbor 10.0.0.134 vrf Prod:Vrf1 | grep Outbound
 Outbound route-map configured is exp-l3out-BGP-peer-2392068, handle obtained
```

在上述示例中，10.0.0.134是在L3Out中配置的BGP邻居。

### EIGRP路由通告

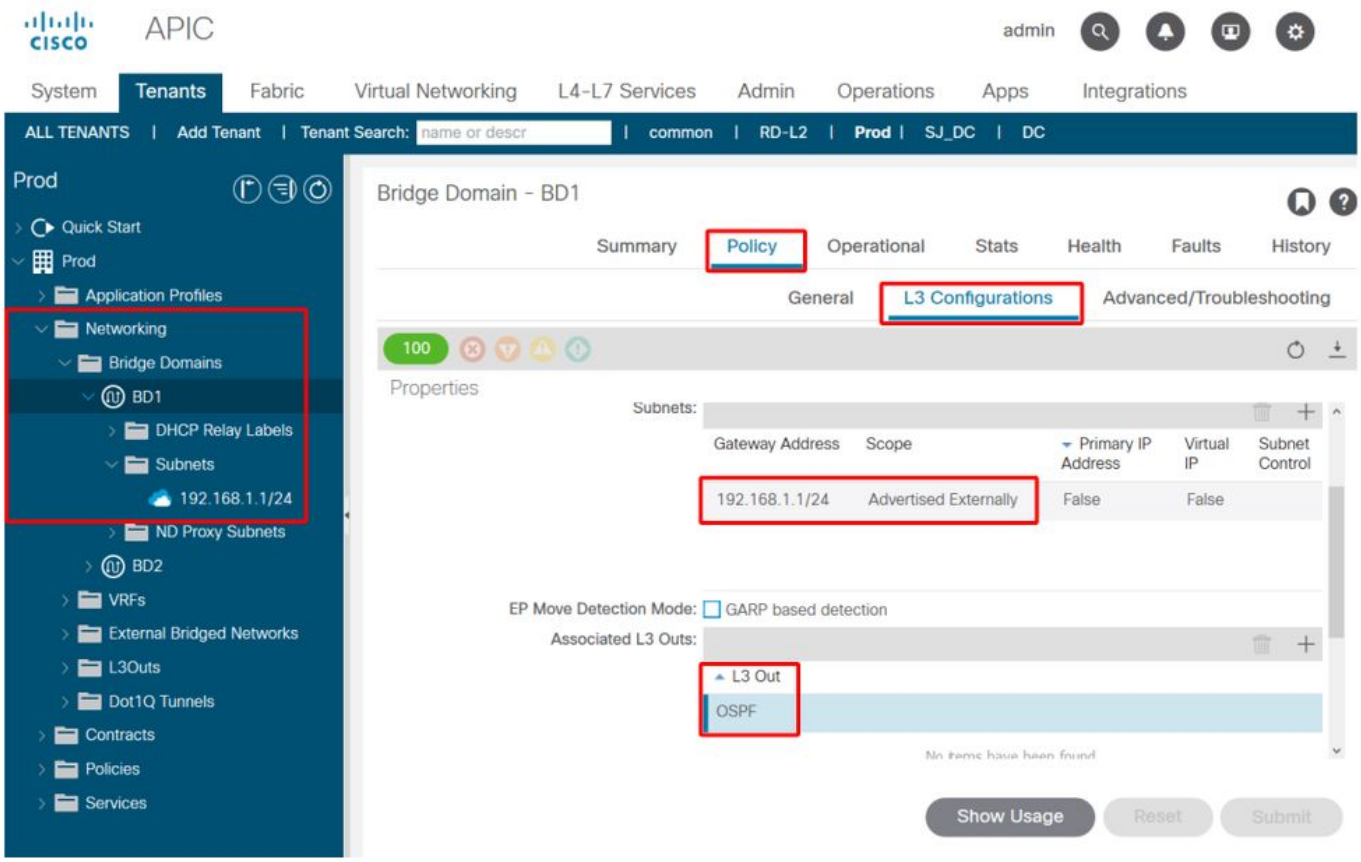
与OSPF类似，路由映射用于控制从Static到EIGRP的重分发。这样，只应重新分发与L3Out关联且设置为“外部通告”的子网。这可以通过以下命令验证：

```
leaf103# show ip eigrp vrf Prod:Vrf1
IP-EIGRP AS 100 ID 10.0.0.3 VRF Prod:Vrf1
 Process-tag: default
 Instance Number: 1
```

```
Status: running
Authentication mode: none
Authentication key-chain: none
Metric weights: K1=1 K2=0 K3=1 K4=0 K5=0
metric version: 32bit
IP proto: 88 Multicast group: 224.0.0.10
Int distance: 90 Ext distance: 170
Max paths: 8
Active Interval: 3 minute(s)
Number of EIGRP interfaces: 1 (0 loopbacks)
Number of EIGRP passive interfaces: 0
Number of EIGRP peers: 2
Redistributing:
  static route-map exp-ctx-st-2392068
  ospf-default route-map exp-ctx-proto-2392068
  direct route-map exp-ctx-st-2392068
  coop route-map exp-ctx-st-2392068
  bgp-65001 route-map exp-ctx-proto-2392068
```

最终工作BD配置如下所示。

## 网桥域L3配置




### 网桥域路由通告故障排除场景

在这种情况下，典型的症状通常是配置的BD子网不会从L3Out中通告。按照上一个工作流程了解哪个组件已损坏。

验证以下内容，在配置过低之前开始配置：

- EPG和L3Out之间是否有合同？

- L3Out是否与BD关联？
- BD子网是否设置为在外部进行通告？
- 外部协议邻接关系是否已启用？

**可能的原因：未部署BD**

此案例适用于多种不同的场景，例如：

- 内部EPG使用与按需选项的VMM集成，并且没有虚拟机终端连接到EPG的端口组。
- 已创建内部EPG，但尚未配置静态路径绑定，或者配置静态路径的接口已关闭。

在这两种情况下，都不会部署BD，因此BD静态路由不会推送到BL。此处的解决方案是在链接到此BD的EPG中部署一些活动资源，以便部署子网。

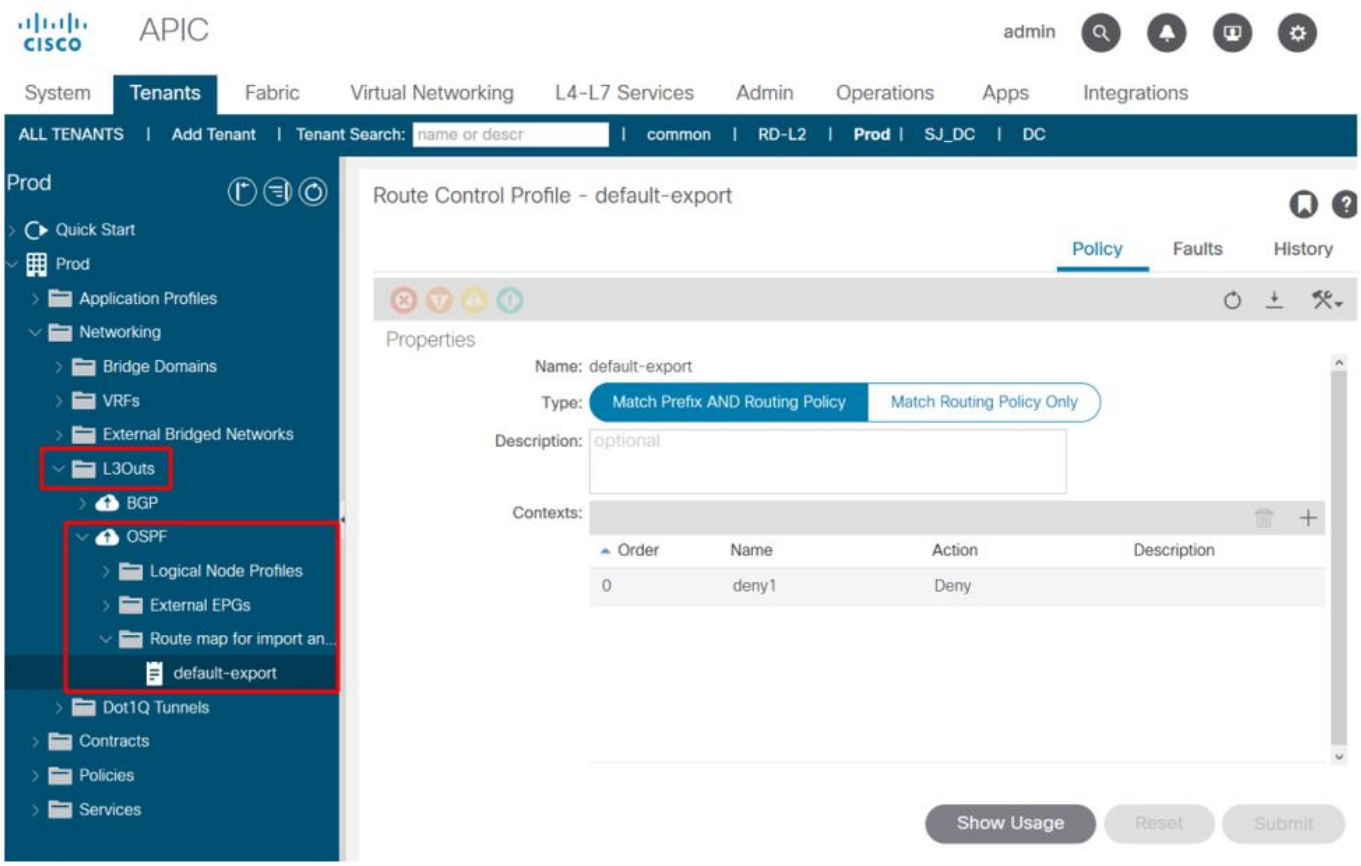
**可能的原因：OSPF L3Out配置为“Stub”或“NSSA”，且无重分发**

将OSPF用作L3Out协议时，仍然必须遵循基本OSPF规则。末节区域不允许重分布的LSA，但可以通告默认路由。NSSA区域确实允许重分发路径，但是必须在L3Out上选择“将重分发的LSA发送到NSSA区域”。或者，NSSA也可以通过禁用“Originate Summary LSA”来通告默认路由，这也是禁用“将重分发的LSA发送到NSSA区域”的典型场景。

**可能的原因：在L3Out下配置“拒绝”操作的“Default-Export”路由配置文件**

在L3Out下使用“default-export”或“default-import”名称配置路由配置文件时，它们会隐式应用于L3Out。此外，如果default-export route-profile设置为deny操作并配置为“Match Prefix and Routing Policy”，则应从此L3Out通告BD子网并隐式拒绝：

## 默认导出拒绝路由配置文件




如果选择了“仅匹配路由策略”(Match Routing Policy Only)选项，则default-export route-profile中的

前缀匹配不会隐式包括BD子网。

## 外部路由导入工作流

本节讨论ACI如何通过L3Out获取外部路由并将其分配到内部枝叶节点。在后面的章节中，还介绍中转和路由泄漏使用案例

与上一节一样，用户应了解更高级别的情况。

默认情况下，通过外部协议获知的所有路由会重分布到内部交换矩阵BGP进程中。无论在外部EPG下配置了哪些子网以及选择了哪些标志，此情况都是正确的。有两个例子表明并非如此。

- 如果顶级L3Out策略的“路由控制实施”选项设置为“导入”。 在这种情况下，路由导入模型将从阻止列表模型（仅指定不应允许的内容）转至允许列表模型（除非另行配置，否则所有内容都隐式拒绝）。
- 如果外部协议是EIGRP或OSPF，并且使用的Interleak Route-Profile与外部路由不匹配。

要将外部路由分配到内部枝叶，必须发生以下情况：

- 路由必须在BL上从外部路由器获知。要成为重新分发到交换矩阵MP-BGP进程的候选者，路由必须安装在路由表中，而不是仅安装在协议RIB中。
- 必须允许将路由重分发或通告到内部BGP进程。除非使用导入路由控制实施或互漏路由配置文件，否则应始终进行此操作。
- 必须配置BGP路由反射器策略并将其应用于应用于Pod配置文件的Pod策略组。如果未应用此项，则交换机上的BGP进程将不会初始化。

如果内部EPG/BD与L3Out位于同一VRF中，则内部EPG/BD使用外部路由只需上述三个步骤。

### 路由安装在BL路由表中

在本例中，应在BL 103和104上获知的外部路由是172.16.20.1/32。

```
leaf103# show ip route 172.16.20.1 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]
'%' in via output denotes VRF

172.16.20.1/32, ubest/mbest: 1/0
    *via 10.10.34.3, vlan347, [110/20], 00:06:29, ospf-default, type-2
```

很明显，它通过OSPF获知后即被安装到路由表中。如果此处未看到该协议，请检查单个协议并确保邻接关系已建立。路由重分布到BGP在检查未使用“导入”实施或互漏路由配置文件后，可以通过查看用于BGP重分布的外部协议的路由映射来验证重分布路由映射。请参阅以下命令：

```
leaf103# show bgp process vrf Prod:Vrf1

Information regarding configured VRFs:

BGP Information for VRF Prod:Vrf1
VRF Type                        : System
VRF Id                          : 85
```

```
VRF state                       : UP
VRF configured                  : yes
VRF refcount                    : 1
VRF VNID                        : 2392068
Router-ID                       : 10.0.0.3
Configured Router-ID            : 10.0.0.3
Confed-ID                       : 0
Cluster-ID                      : 0.0.0.0
MSITE Cluster-ID                : 0.0.0.0
No. of configured peers         : 1
No. of pending config peers     : 0
No. of established peers        : 1
VRF RD                          : 101:2392068
VRF EVPN RD                     : 101:2392068
...
    Redistribution
        direct, route-map permit-all
        static, route-map imp-ctx-bgp-st-interleak-2392068
        ospf, route-map permit-all
        coop, route-map exp-ctx-st-2392068
        eigrp, route-map permit-all
```

这里很明显，“permit-all”路由映射用于OSPF到BGP的重分发。这是默认设置。从此处可以验证BL并检查源自BGP的本地路由：

```
a-leaf101# show bgp ipv4 unicast 172.16.20.1/32 vrf Prod:Vrf1
BGP routing table information for VRF Prod:Vrf1, address family IPv4 Unicast
BGP routing table entry for 172.16.20.1/32, version 25 dest ptr 0xa6f25ad0
Paths: (2 available, best #2)
Flags: (0x80c0002 00000000) on xmit-list, is not in urib, exported
  vpn: version 16316, (0x100002) on xmit-list
Multipath: eBGP iBGP


  Advertised path-id 1, VPN AF advertised path-id 1
  Path type: redist 0x408 0x1 ref 0 adv path ref 2, path is valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (10.0.0.3)
      Origin incomplete, MED 20, localpref 100, weight 32768
      Extcommunity:
          RT:65001:2392068
          VNID:2392068
          COST:pre-bestpath:162:110

  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    10.0.64.64         10.0.72.66
  Path-id 2 not advertised to any peer
```

在上述输出中，0.0.0.0/0表示它源自本地。通告的对等体列表是交换矩阵中充当路由反射器的主干节点。

**检验内部枝叶上的路由**

BL应通过VPNv4 BGP地址系列将其通告给主干节点。主干节点应将其通告给任何已部署VRF的枝

叶节点（非路由泄漏示例中为true）。 在任何这些枝叶节点上运行“show bgp vpnv4 unicast <route> vrf overlay-1”以验证它是否在VPNv4中

使用以下命令验证内部枝叶上的路由。

```
leaf101# show ip route 172.16.20.1 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]
'%' in via output denotes VRF

172.16.20.1/32, ubest/mbest: 2/0
    *via 10.0.72.64%overlay-1, [200/20], 00:21:24, bgp-65001, internal, tag 65001
         recursive next hop: 10.0.72.64/32%overlay-1
    *via 10.0.72.67%overlay-1, [200/20], 00:21:24, bgp-65001, internal, tag 65001
         recursive next hop: 10.0.72.67/32%overlay-1
```

在上述输出中，路由是通过BGP获取的，下一跳应该是BL的物理TEP(PTEP)。

```
leaf101# acidiag fnvread
      ID   Pod ID                 Name    Serial Number         IP Address    Role        State
LastUpdMsgId
-------------------------------------------------------------------------------------------------
--------------
     103        1            a-leaf101      FDO20160TPS      10.0.72.67/32    leaf
active   0
     104        1            a-leaf103      FDO20160TQ0      10.0.72.64/32    leaf
active   0
```

**外部路由故障排除场景**

在这种情况下，内部枝叶(101)没有收到外部路由。

与往常一样，首先检查基本信息。确保：

- BL上的路由协议邻接关系已启用。
- BGP路由反射器策略应用于Pod策略组和Pod配置文件。

如果上述条件正确，下面是一些可能导致问题的更高级示例。

**可能的原因：未在内部枝叶上部署VRF**

在这种情况下，问题可能是在预期外部路由的内部枝叶上部署了资源的EPG。这可能是因为静态路径绑定仅在关闭接口上配置，或仅存在按需模式VMM集成EPG，且未检测到动态附件。

由于L3Out VRF未部署在内部枝叶上（在内部枝叶上使用“show vrf”进行验证），因此内部枝叶不会从VPNv4导入BGP路由。

要解决此问题，用户应在内部枝叶上的L3Out VRF中部署资源。

**可能的原因：正在使用导入路由实施**

如前所述，当启用导入路由控制实施时，L3Out仅接受明确允许的外部路由。通常，该功能以表映

射的形式实现。表映射位于协议RIB和实际路由表之间，因此它只影响路由表中的内容。

在下面的输出中，导入路由控制已启用，但没有任何明确允许的路由。请注意，LSA位于OSPF数据库中，但不位于BL上的路由表中：

```
leaf103# vsh -c "show ip ospf database external 172.16.20.1 vrf Prod:Vrf1"
        OSPF Router with ID (10.0.0.3) (Process ID default VRF Prod:Vrf1)

                Type-5 AS External Link States

Link ID         ADV Router      Age         Seq#       Checksum Tag
172.16.20.1     10.0.0.134      455         0x80000003 0xb9a0    0


leaf103# show ip route 172.16.20.1 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]
'%' in via output denotes VRF

Route not found
```

以下是当前安装的导致此行为的表映射：

```
leaf103# show ip ospf vrf Prod:Vrf1

 Routing Process default with ID 10.0.0.3 VRF Prod:Vrf1
 Stateful High Availability enabled
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
 Table-map using route-map exp-ctx-2392068-deny-external-tag
 Redistributing External Routes from..

leaf103# show route-map exp-ctx-2392068-deny-external-tag
route-map exp-ctx-2392068-deny-external-tag, deny, sequence 1
  Match clauses:
    tag: 4294967295
  Set clauses:
route-map exp-ctx-2392068-deny-external-tag, deny, sequence 19999
  Match clauses:
    ospf-area: 0.0.0.100
  Set clauses:
```

区域100（在此L3Out上配置的区域）中的任何学习内容都会被此表映射隐式拒绝，因此它不会安装到路由表中。

要解决此问题，用户应使用“Import Route Control Subnet”标记在外部EPG上定义子网，或创建与要安装的前缀匹配的导入路由配置文件。

- 请注意，EIGRP不支持导入实施。
- 另请注意，对于BGP，导入实施作为应用于BGP邻居的入站路由映射实施。检查“BGP路由通告”子部分以了解有关如何检查的详细信息。

**可能的原因：正在使用Interleak配置文件**

Interleak Route-Profiles用于EIGRP和OSPF L3Outs，用于控制从IGP重分发到BGP的内容，并允

许应用策略（如设置BGP属性）。

如果没有互漏路由配置文件，所有路由都会隐式导入BGP。

没有互漏路由配置文件：

```
leaf103# show bgp process vrf Prod:Vrf1

Information regarding configured VRFs:

BGP Information for VRF Prod:Vrf1
VRF Type                        : System
VRF Id                          : 85
VRF state                       : UP
VRF configured                  : yes
VRF refcount                    : 1
VRF VNID                        : 2392068
Router-ID                       : 10.0.0.3
Configured Router-ID            : 10.0.0.3
Confed-ID                       : 0
Cluster-ID                      : 0.0.0.0
MSITE Cluster-ID                : 0.0.0.0
No. of configured peers         : 1
No. of pending config peers     : 0
No. of established peers        : 1
VRF RD                          : 101:2392068
VRF EVPN RD                     : 101:2392068

...
    Peers      Active-peers    Routes     Paths      Networks   Aggregates
    1          1               7          11         0          0

    Redistribution
        direct, route-map permit-all
        static, route-map imp-ctx-bgp-st-interleak-2392068
        ospf, route-map permit-all
        coop, route-map exp-ctx-st-2392068
        eigrp, route-map permit-all
```

使用互漏路由配置文件：

```
a-leaf103# show bgp process vrf Prod:Vrf1

Information regarding configured VRFs:

BGP Information for VRF Prod:Vrf1
VRF Type                        : System
VRF Id                          : 85
VRF state                       : UP
VRF configured                  : yes
VRF refcount                    : 1
VRF VNID                        : 2392068
Router-ID                       : 10.0.0.3
Configured Router-ID            : 10.0.0.3
Confed-ID                       : 0
Cluster-ID                      : 0.0.0.0
MSITE Cluster-ID                : 0.0.0.0
No. of configured peers         : 1
No. of pending config peers     : 0
```

```
No. of established peers       : 1
VRF RD                         : 101:2392068
VRF EVPN RD                    : 101:2392068

...

    Redistribution
        direct, route-map permit-all
        static, route-map imp-ctx-bgp-st-interleak-2392068
        ospf, route-map imp-ctx-proto-interleak-2392068
        coop, route-map exp-ctx-st-2392068
        eigrp, route-map permit-all
```

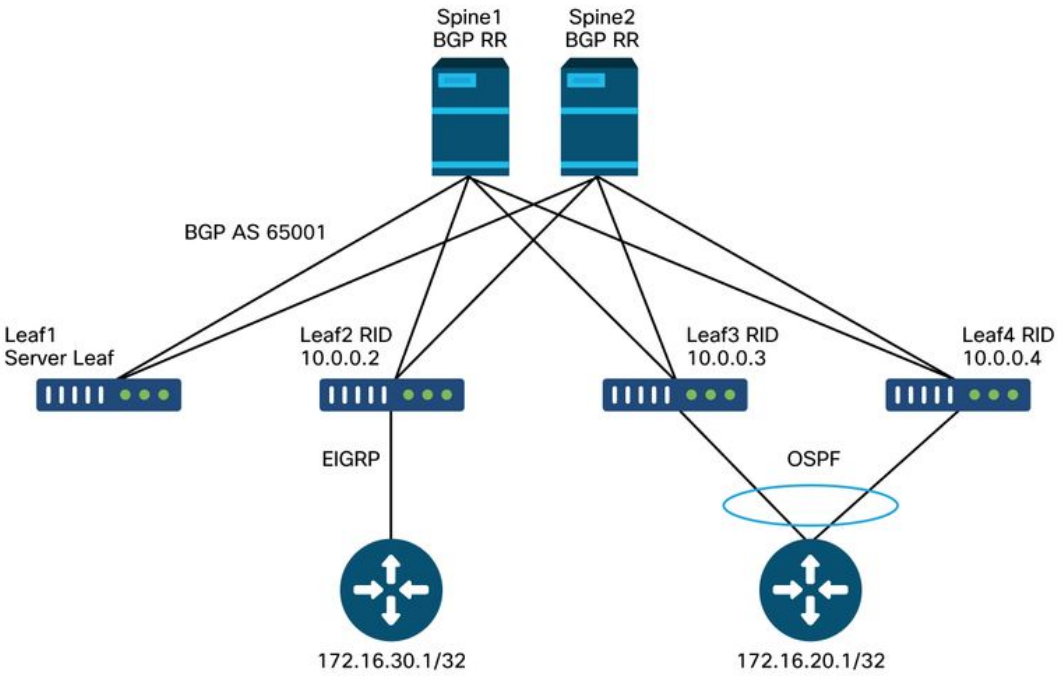
以上突出显示的路由映射将仅允许已配置的Interleak配置文件中明确匹配的内容。如果外部路由不匹配，则不会将其重新分配到BGP。

## 传输路由通告工作流

本节讨论如何从一个L3Out路由通告到另一个L3Out。这也包括需要通告L3Out上直接配置的静态路由的情况。它不会涉及每个具体的协议考虑事项，而是会涉及如何在ACI中实施该事项。此时不会进入VRF间传输路由。

此场景将使用以下拓扑：

## 传输路由拓扑




如何从OSPF获知172.16.20.1并将其通告到EIGRP的概要流程，以及整个流程和故障排除方案的验证将在下面讨论。

要将172.16.20.1路由通告到EIGRP，必须配置以下其中一项：

- 可以在EIGRP L3Out上使用“Export Route-Control Subnet”标志定义要通告的子网。如概述部分

所述，此标志主要用于中转路由，并定义应从该L3Out通告的子网。

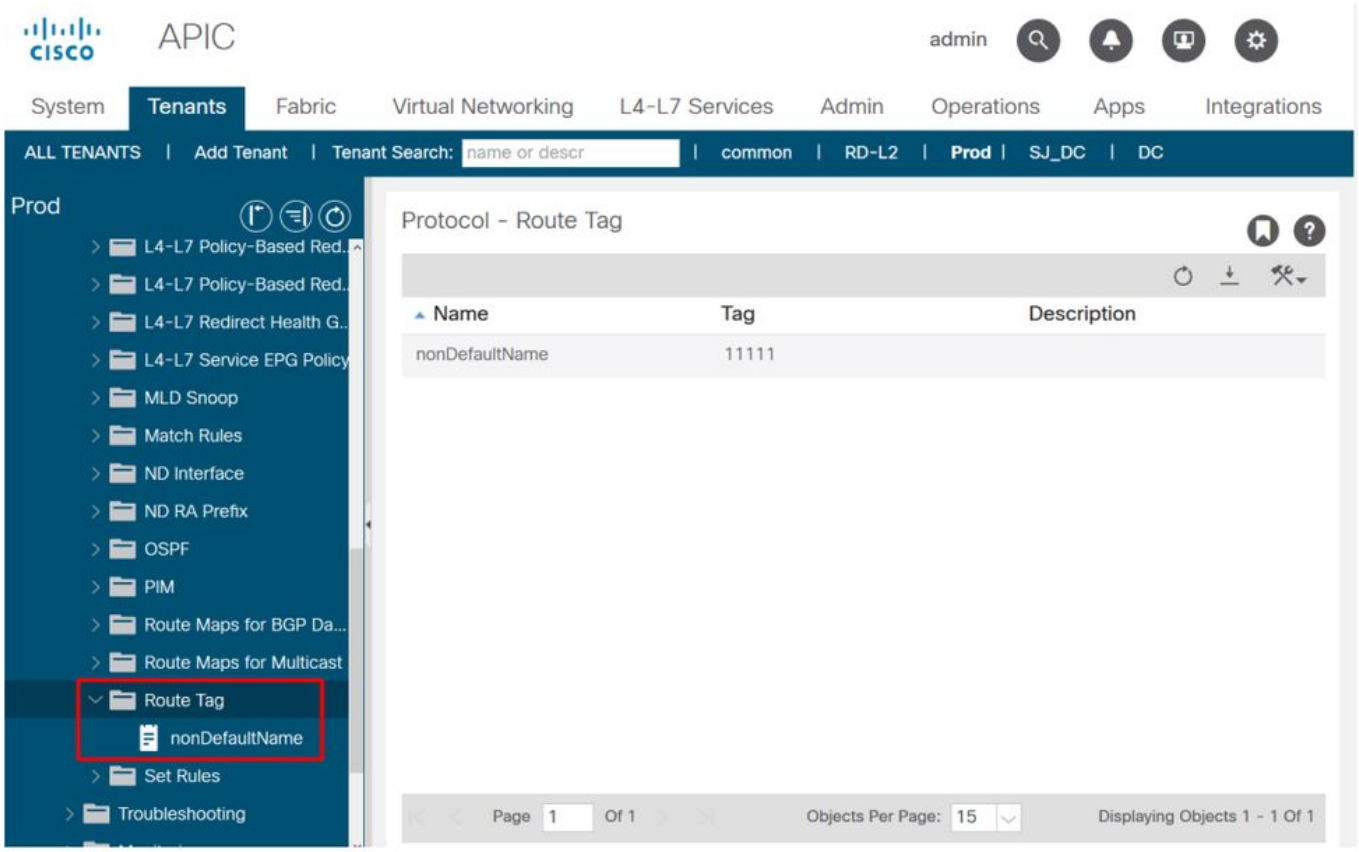
- 配置0.0.0.0/0并选择“Aggregate Export”和“Export Route Control Subnet”。 这会创建一个路由映射，用于重分发到与0.0.0.0/0及更具体的所有前缀（这是有效的匹配any）匹配的外部协议。请注意，当0.0.0.0/0与“Aggregate Export”一起使用时，不会为重分发匹配静态路由。这是为了防止无意中通告不应通告的BD路由。
- 最后，可以创建与要通告的前缀匹配的导出路由配置文件。使用此方法可以配置“Aggregate”选项，该选项带有除0.0.0.0/0外的前缀。

上述配置将导致中转路由被通告，但中转路由仍然需要实施安全策略以允许数据平面流量通过。与任何EPG到EPG通信一样，在允许流量之前必须签订合同。

> 请注意，不能在同一VRF中配置具有“外部EPG的外部子网”的重复外部子网。配置时，子网需要比0.0.0.0更具体。仅对接收路由的L3Out配置“外部EPG的子网”非常重要。请勿在应通告此路由的L3Out上配置此路由。

此外，还必须了解所有中转路由都标记有特定的VRF标记。默认情况下，此标记为4294967295。路由标记策略在“租户>网络>协议>路由标记：

## 路由标记策略




然后，此路由标记策略将应用于VRF。此标记的用途主要是防止环路。当中转路由通告回L3Out时，将应用此路由标记。如果随后收到具有相同路由标记的这些路由，则丢弃该路由。

### 检验通过OSPF的接收BL上是否存在路由

像最后一节一样，首先检验最初应接收正确路由的BL。

```
leaf103# show ip route 172.16.20.1 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]
'%' in via output denotes VRF

172.16.20.1/32, ubest/mbest: 1/0
    *via 10.10.34.3, vlan347, [110/20], 01:25:30, ospf-default, type-2
```

现在，假设通告L3Out位于不同的BL上（与拓扑中一样）（后面的场景将讨论它在同一个BL上的位置）。

### 验证该路由存在于接收OSPF BL的BGP中

对于要通告到外部EIGRP路由器的OSPF路由，需要将该路由通告到接收OSPF BL上的BGP。

```
leaf103# show bgp ipv4 unicast 172.16.20.1/32 vrf Prod:Vrf1
BGP routing table information for VRF Prod:Vrf1, address family IPv4 Unicast
BGP routing table entry for 172.16.20.1/32, version 30 dest ptr 0xa6f25ad0
Paths: (2 available, best #1)
Flags: (0x80c0002 00000000) on xmit-list, is not in urib, exported
  vpn: version 17206, (0x100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type: redist 0x408 0x1 ref 0 adv path ref 2, path is valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (10.0.0.3)
      Origin incomplete, MED 20, localpref 100, weight 32768
      Extcommunity:
          RT:65001:2392068
          VNID:2392068
          COST:pre-bestpath:162:110

  VRF advertise information:

  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    10.0.64.64         10.0.72.66
  Path-id 2 not advertised to any peer
```

该路由在BGP中。

### 在EIGRP BL上检验应通告其已安装的路由

```
leaf102# show ip route 172.16.20.1 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]
'%' in via output denotes VRF

172.16.20.1/32, ubest/mbest: 2/0
    *via 10.0.72.67%overlay-1, [200/20], 00:56:46, bgp-65001, internal, tag 65001
         recursive next hop: 10.0.72.67/32%overlay-1
    *via 10.0.72.64%overlay-1, [200/20], 00:56:46, bgp-65001, internal, tag 65001
```

```
        recursive next hop: 10.0.72.64/32%overlay-1
```

它安装在路由表中，具有指向原始边界枝叶节点的重叠下一跳。

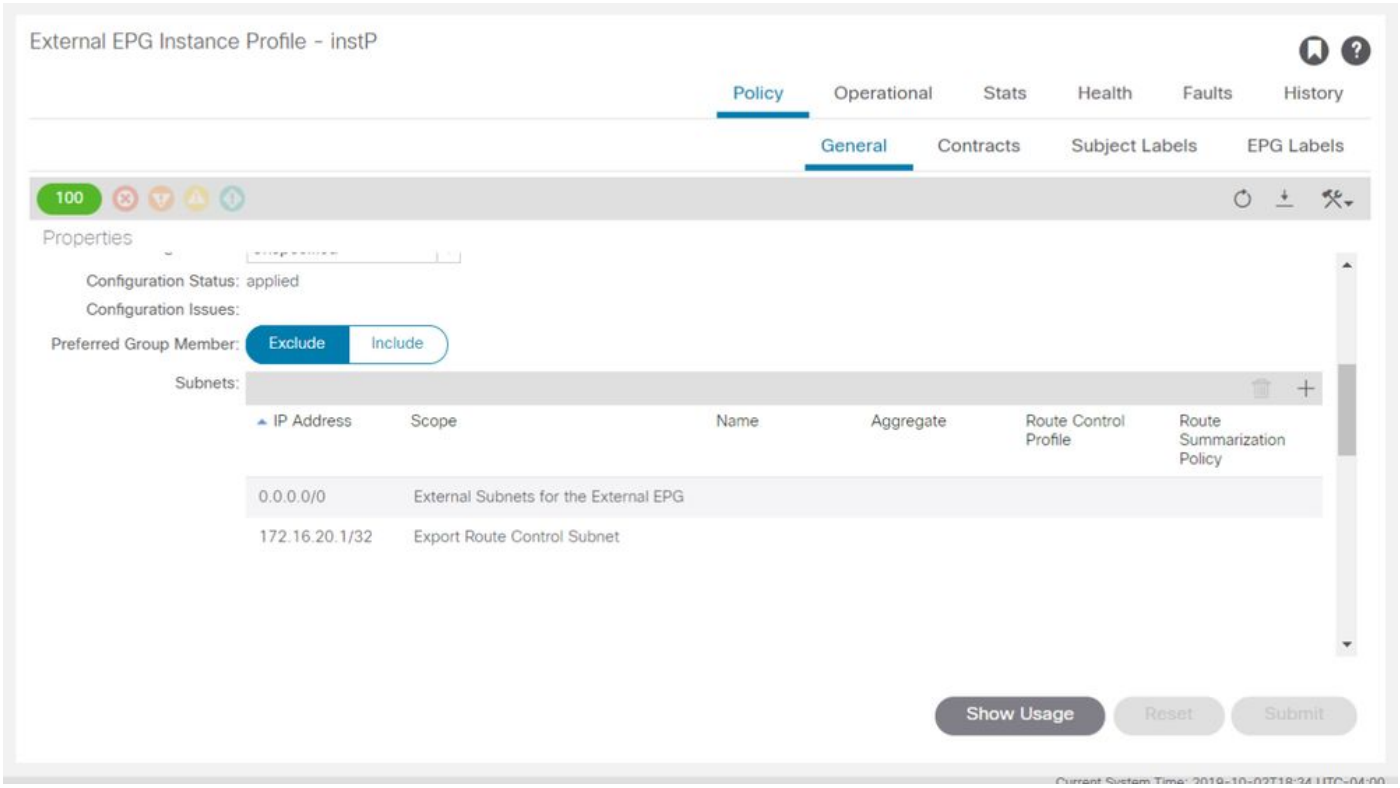
```
leaf102# acidiag fnvread

      ID   Pod ID                 Name    Serial Number         IP Address    Role        State
LastUpdMsgId
-------------------------------------------------------------------------------------------------
--------------
     103        1            a-leaf101      FDO20160TPS      10.0.72.67/32    leaf
active   0
     104        1            a-leaf103      FDO20160TQ0      10.0.72.64/32    leaf
active   0
```

### 检验路由是否已在BL上通告

由于在配置的子网上设置了“Export Route Control Subnet”标志，BL 102将通告路由：

## 导出路由控制




使用以下命令查看由于此“导出路由控制”标志创建的路由映射：

```
leaf102# show ip eigrp vrf Prod:Vrf1
IP-EIGRP AS 101 ID 10.0.0.2 VRF Prod:Vrf1
  Process-tag: default
  Instance Number: 1
  Status: running
  Authentication mode: none
  Authentication key-chain: none
  Metric weights: K1=1 K2=0 K3=1 K4=0 K5=0
  metric version: 32bit
```

```
  IP proto: 88 Multicast group: 224.0.0.10
  Int distance: 90 Ext distance: 170
  Max paths: 8
  Active Interval: 3 minute(s)
  Number of EIGRP interfaces: 1 (0 loopbacks)
  Number of EIGRP passive interfaces: 0
  Number of EIGRP peers: 1
  Redistributing:
    static route-map exp-ctx-st-2392068
    ospf-default route-map exp-ctx-proto-2392068
    direct route-map exp-ctx-st-2392068
    coop route-map exp-ctx-st-2392068
    bgp-65001 route-map exp-ctx-proto-2392068
```

要查找“BGP > EIGRP redistribution”，请查看路由映射。但是，无论源协议是OSPF、EIGRP还是BGP，路由映射本身都应相同。静态路由将使用不同的路由映射进行控制。

```
leaf102# show route-map exp-ctx-proto-2392068
route-map exp-ctx-proto-2392068, permit, sequence 15801
  Match clauses:
    ip address prefix-lists: IPv4-proto32771-2392068-exc-ext-inferred-export-dst
    ipv6 address prefix-lists: IPv6-deny-all
  Set clauses:
    tag 4294967295

a-leaf102# show ip prefix-list IPv4-proto32771-2392068-exc-ext-inferred-export-dst
ip prefix-list IPv4-proto32771-2392068-exc-ext-inferred-export-dst: 1 entries
   seq 1 permit 172.16.20.1/32
```

在上述输出中，在此前缀上设置VRF标记用于环路预防，并且使用“导出路由控制”配置的子网明确匹配。

### 接收和通告BL时的传输路由相同

如前所述，当接收和通告BL不同时，必须使用BGP通过交换矩阵通告路由。当BL相同时，可以在枝叶上的协议之间直接执行重分发或通告。

下面简要介绍如何实现此功能：

- **在同一枝叶上的两个OSPF L3Outs之间传输路由：路由通告通过应用于OSPF进程级别的“**area-filter”进行控制。区域0中的L3Out必须部署在枝叶上，因为路由是在区域之间通告的，而不是通过重分发进行通告。使用“show ip ospf vrf <name>”查看过滤器列表。使用“show route-map <filter name>”显示过滤器的内容。
- **在同一枝叶上OSPF和EIGRP L3Outs之间传输路由：路由通告通过重分发路由映射控制，通过“**show ip ospf”和“show ip eigrp”可以看到。 请注意，如果同一BL中存在多个OSPF L3Outs，则重新分发到这些OSPF L3Outs中仅一个的唯一方法是，另一个是末节或NSSA，禁用了“**Send redistributed LSAs into NSSA area”(将重分发的LSA发送到NSSA区**域)，以便它不允许任何外部LSA。
- **在同一枝叶上的OSPF或EIGRP与BGP之间传输路由：通**过重分配路由映射控制到IGP的路由通告。BGP中的路由通告通过直接应用于应发送路由的bgp邻居的出站路由映射进行控制。这可以通过“show bgp ipv4 unicast neighbor <neighbor address> vrf <name>进行验证 | grep Outbound'。
- **在同一枝叶上的两个BGP l3Outs之间传输路由：所有通告通过直接应用于路由应发送到的**

**bgp邻居的路由映射来控**制。这可以通过“show bgp ipv4 unicast neighbor <neighbor address> vrf <name>进行验证 | grep Outbound'。

**传输路由故障排除#1案：未通告中转路由**

此故障排除场景涉及应该通过一个L3Out获知的路由，而不是从另一个L3Out发送出去。

与以往一样，请先检查基本知识，然后再查看任何特定于ACI的内容。

- 协议邻接关系是否已建立？
- ACI应通告的路由是否首先从外部协议获知？
- 对于BGP，路径是否由于某些BGP属性而被丢弃？（as路径等）。
- 接收的L3Out是否包含在OSPF数据库、EIGRP拓扑表或BGP表中？
- BGP路由反射器策略是否应用到Pod策略组，是否应用到Pod配置文件？

如果所有基本协议验证都配置正确，下面是未通告中转路由的一些其他常见原因。

**可能的原因：无OSPF区域0**

如果受影响的拓扑涉及同一边界枝叶上的两个OSP L3Outs，则必须存在区域0，以便将路由从一个区域通告到另一个区域。有关详细信息，请参阅上面的“同一枝叶上两个OSPF L3Outs之间的传输路由”项目符号。

**可能的原因：OSPF区域是末节或NSSA**

如果OSPF L3Out配置了末节区域或未配置为通告外部LSA的NSSA区域，则会出现这种情况。使用OSPF时，外部LSA永远不会通告到末节区域。如果选择了“将重分发的LSA发送到NSSA区域”，则将其通告到NSSA区域。

**传输路由故障排除#2案：未收到中转路由**

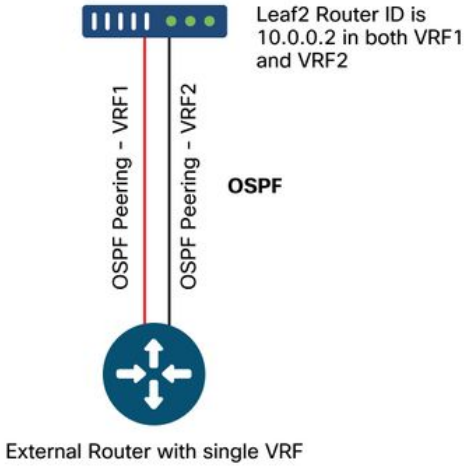
在这种情况下，问题在于ACI L3Out通告的某些路由未在另一个L3Out中返回。如果L3Outs位于两个独立的交换矩阵中，并且由外部路由器连接，或者L3Outs位于不同的VRF中，并且路由正由外部路由器在VRF之间传递，则此方案可能适用。

**可能的原因：多个VRF中的BL配置了相同的路由器ID**

从配置角度看，路由器ID不能在同一VRF中重复。但是，只要两个VRF没有连接到相同的路由协议域，通常可以在不同VRF中使用相同的路由器ID。

请考虑以下拓扑：

## 具有单个VRF的外部路由器 — 未收到传输路由




此处的问题是，ACI枝叶会看到收到其自己的路由器ID的LSA，从而导致这些数据包未安装在OSPF数据库中。

此外，如果发现使用VPC对时设置相同，则会在某些路由器上持续添加和删除LSA。例如，路由器会看到来自其VPC对等体的LSA和VRF以及来自其他VRF中发起的同一节点（具有相同的路由器ID）的LSA。

要解决此问题，用户应确保节点在具有L3Out的每个VRF内具有不同的唯一路由器ID。

**可能的原因：使用相同VRF标记从一个ACI交换矩阵中的一个L3Out接收的路由**

除非更改，否则ACI中的默认路由标记始终相同。如果路由从一个VRF或ACI交换矩阵中的一个L3Out通告到另一个VRF或ACI交换矩阵中的另一个L3Out，而不更改默认VRF标记，则接收BL将丢弃路由。

此场景的解决方案只是对ACI中的每个VRF使用唯一的路由标记策略。

**中转路由故障排除场景#3 — 意外通告的中转路由**

当中转路由通告到不打算通告它们的L3Out时，将会出现这种情况。

**可能的原因：0.0.0.0/0与“Aggregate Export”的用法**

当外部子网配置为0.0.0.0/0并带有“Export Route Control Subnet”和“Aggregate Export”时，结果是安装了匹配的所有重分发路由映射。在这种情况下，通过OSPF、EIGRP或BGP获知的BL上的所有路由都会从配置该路由的L3Out中通告。

以下是作为聚合导出结果部署到枝叶的路由映射：

```
leaf102# show ip eigrp vrf Prod:Vrf1
IP-EIGRP AS 101 ID 10.0.0.2 VRF Prod:Vrf1
 Process-tag: default
 Instance Number: 1
 Status: running
 Authentication mode: none
 Authentication key-chain: none
 Metric weights: K1=1 K2=0 K3=1 K4=0 K5=0
```

```
 metric version: 32bit
 IP proto: 88 Multicast group: 224.0.0.10
 Int distance: 90 Ext distance: 170
 Max paths: 8
 Active Interval: 3 minute(s)
 Number of EIGRP interfaces: 1 (0 loopbacks)
 Number of EIGRP passive interfaces: 0
 Number of EIGRP peers: 1
 Redistributing:
   static route-map exp-ctx-st-2392068
   ospf-default route-map exp-ctx-proto-2392068
   direct route-map exp-ctx-st-2392068
   coop route-map exp-ctx-st-2392068
   bgp-65001 route-map exp-ctx-proto-2392068
 Tablemap: route-map exp-ctx-2392068-deny-external-tag , filter-configured
 Graceful-Restart: Enabled
 Stub-Routing: Disabled
 NSF converge time limit/expiries: 120/0
 NSF route-hold time limit/expiries: 240/0
 NSF signal time limit/expiries: 20/0
 Redistributed max-prefix: Disabled
 selfAdvRtTag: 4294967295
leaf102# show route-map exp-ctx-proto-2392068
route-map exp-ctx-proto-2392068, permit, sequence 19801
 Match clauses:
   ip address prefix-lists: IPv4-proto32771-2392068-agg-ext-inferred-export-dst
   ipv6 address prefix-lists: IPv6-deny-all
 Set clauses:
   tag 4294967295



leaf102# show ip prefix-list IPv4-proto32771-2392068-agg-ext-inferred-export-dst
    ip prefix-list IPv4-proto32771-2392068-agg-ext-inferred-export-dst: 1 entries
seq 1 permit 0.0.0.0/0 le 32
```

这是导致涉及ACI环境的路由环路的第一大原因。

# 合同和L3Out

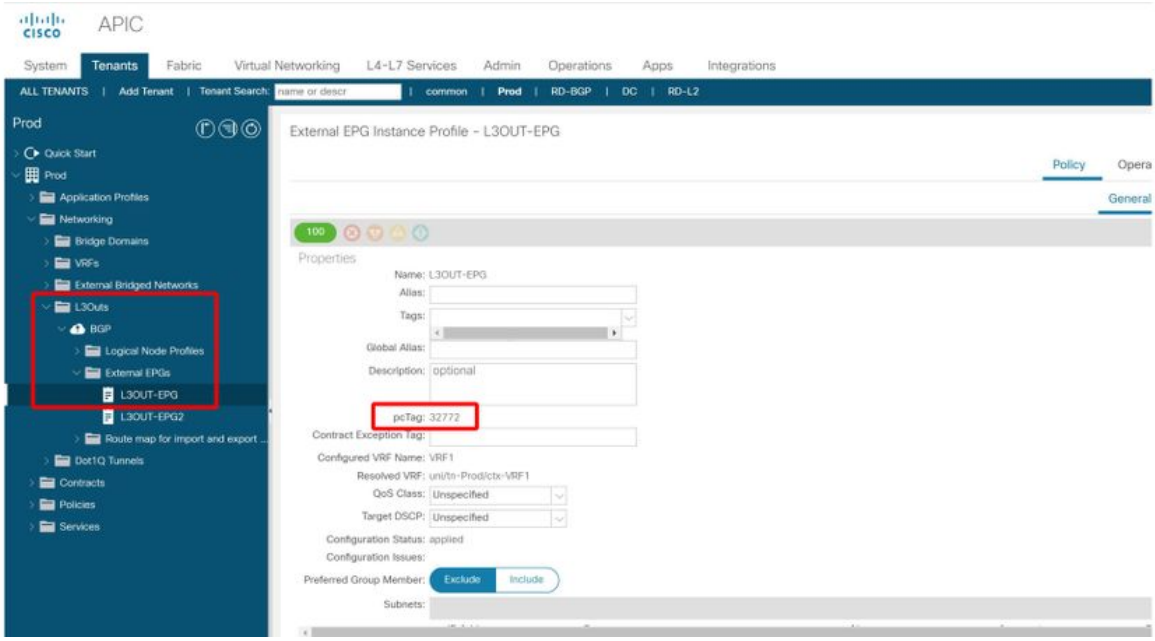
## L3Out上基于前缀的EPG

在内部EPG（非L3Out）中，合同在派生源EPG的pcTag和目标EPG的pcTag后实施。下行链路端口上接收的数据包的封装VLAN/VXLAN用于将数据包分类到EPG中，从而驱动此pcTag。每当学习MAC地址或IP地址时，都会学习该地址及其访问封装和关联的EPG pcTag。有关pcTag和合同实施的更多详细信息，请参阅“安全策略”一章。

L3Outs还使用位于“Tenant > Networking > L3OUT > Networks > L3OUT-EPG”下的L3Out EPG（外部EPG）驱动pcTag。 但是，L3Outs不依靠VLAN和接口对数据包进行此类分类。相反，分类基于“最长前缀匹配”形式的源前缀/子网。因此，L3Out EPG可以称为基于**前缀的EPG**。根据子网将数据包分类为L3Out后，它遵循与常规EPG类似的策略实施模式。

下图概述了可在GUI中找到给定L3Out EPG的pcTag。

## L3Out的pcTag的位置




用户负责定义基于前缀的EPG表。这是使用“外部EPG的外部子网”子网范围完成的。使用该作用域设置的每个子网将在静态最长前缀匹配(LPM)表中添加条目。此子网将指向用于该前缀内的任何IP地址的pcTag值。

可以在枝叶交换机上使用以下命令验证基于前缀的EPG子网的LPM表：

```
vsh -c 'show system internal policy-mgr prefix'
```

备注:

- LPM表条目范围设为VRF VNID。根据vrf_vnid/src pcTag/dst pcTag完成查找。
- 每个条目都指向一个pcTag。因此，两个L3Out EPG不能在同一VRF内使用具有相同掩码长度的同一子网。
- 子网0.0.0.0/0始终使用特殊的pcTag 15。因此，可以复制该子网，但只有在完全了解策略实施含义的情况下才能复制。
- 此表双向使用。 从L3Out到枝叶本地终端，源pcTag使用此表派生。从枝叶本地终端到L3Out，目标pcTag使用此表派生。
- 如果VRF具有“Policy Control Enforcement Direction”的“Ingress”实施设置，则LPM前缀表将存在于L3Out BL以及VRF中具有与L3Out合同的所有枝叶交换机上。

## 示例 1：具有特定前缀的单个L3Out

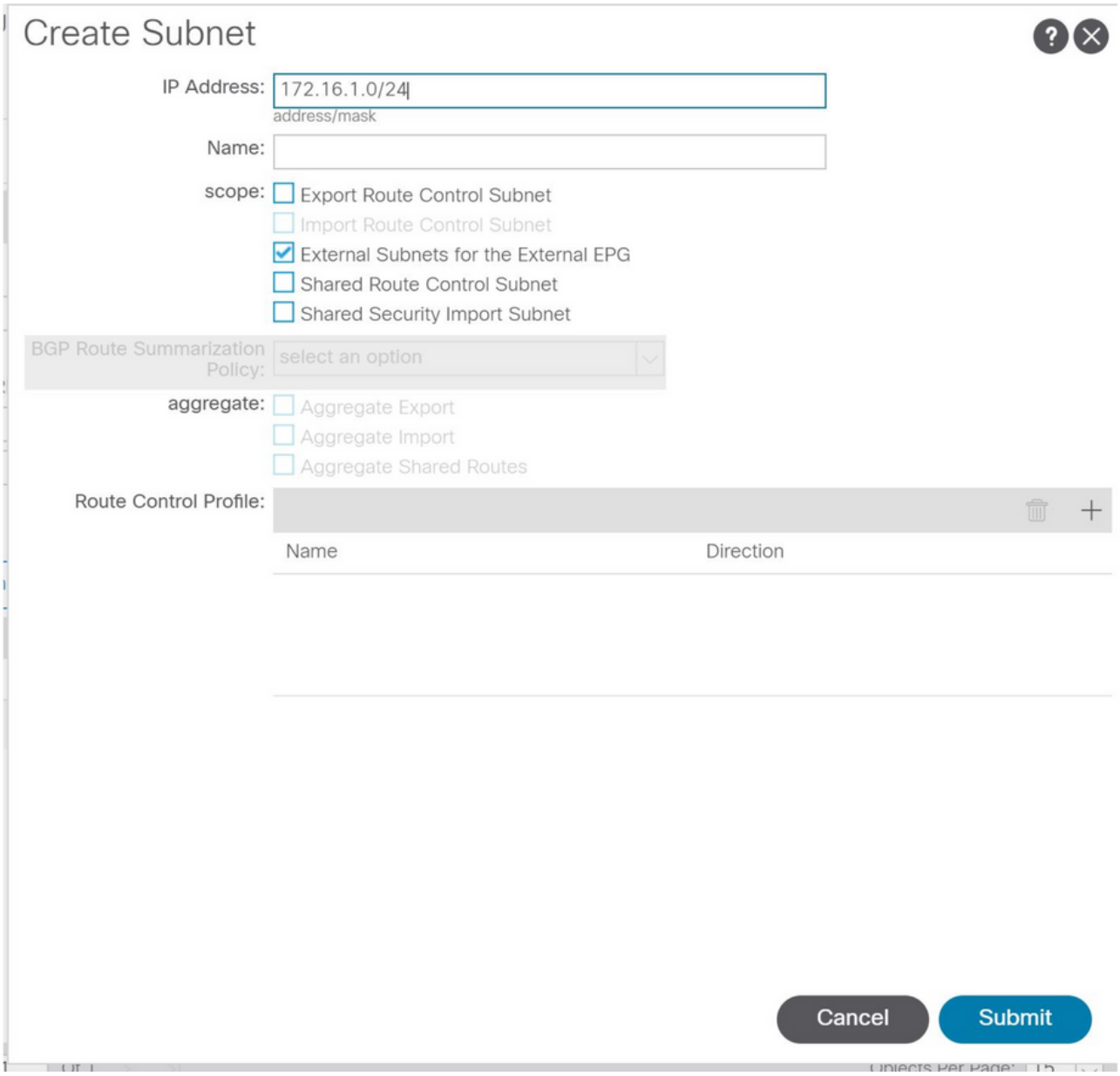
**场景:**vrf Prod:VRF1中的单个BGP L3Out和一个L3Out EPG。从外部源接收前缀172.16.1.0/24，因此必须将其分类到L3Out EPG中。

```
bdsol-aci32-leaf3# show ip route 172.16.1.0  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]
'%' in via output denotes VRF

172.16.1.0/24, ubest/mbest: 1/0
    *via 10.0.0.134%Prod:VRF1, [20/0], 00:56:14, bgp-132, external, tag 65002
         recursive next hop: 10.0.0.134/32%Prod:VRF1
```

首先，将子网添加到前缀表中。

## 具有“外部EPG的外部子网”范围的子网




验证具有L3Out VRF的枝叶交换机上的前缀列表的编程：

```
bdsol-aci32-leaf3# vsh -c ' show system internal  policy-mgr prefix ' |  egrep "Prod|==|Addr"
Vrf-Vni VRF-Id Table-Id Table-State  VRF-Name                     Addr
Class Shared Remote Complete
======= ======  =========== =======  ============================
================================ ====== ====== ====== ========
2097154 35     0x23          Up     Prod:VRF1
0.0.0.0/0   15      True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.1.0/24  32772   True   True   False
```

L3Out EPG的pcTag在vrf范32772中处于状2097154。

## 示例 2：带有多个前缀的单个L3Out

在前一个示例上扩展，在此场景中，L3Out接收多个前缀。当输入每个前缀在功能上合理时，另一个选项（取决于预期设计）是接受L3Out上收到的所有前缀。

可以使用'0.0.0.0/0'前缀完成此操作。

# Subnet - 0.0.0.0/0

Policy | Faults | History

## Properties

IP Address: 0.0.0.0/0
address/mask

Scope:
- [ ] Export Route Control Subnet
- [ ] Import Route Control Subnet
- [x] External Subnets for the External EPG
- [ ] Shared Route Control Subnet
- [ ] Shared Security Import Subnet

Aggregate:
- [ ] Aggregate Export
- [ ] Aggregate Import
- [ ] Aggregate Shared Routes

BGP Route Summarization Policy: select an option

Route Control Profile:

| Name | Direction |
| --- | --- |

No items have been found.
Select Actions to create a new item.

这会导致以下policy-mgr前缀表条目：

```
bdsol-aci32-leaf3# vsh -c ' show system internal  policy-mgr prefix ' |  egrep "Prod|==|Addr"
Vrf-Vni VRF-Id Table-Id Table-State  VRF-Name                     Addr
Class Shared Remote Complete
======= ======  =========== =======  ============================
================================ ====== ====== ====== ========
2097154 35     0x23          Up     Prod:VRF1
0.0.0.0/0   15      True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.1.0/24  32772   True   True   False
```

> 请注意，分配给0.0.0.0/0的pcTag使用值15，而不是32772。pcTag 15是保留的系统pcTag，仅与0.0.0.0/0一起使用，后者充当通配符以匹配L3Out上的所有前缀。

如果VRF具有单个L3Out和使用0.0.0.0/0的单个L3Out EPG，则策略前缀保持唯一，并且是最容易捕获所有内容的方法。

## 示例3a:VRF中的多个L3Out EPG

在这种情况下，同一VRF中有多个L3Out EPG。

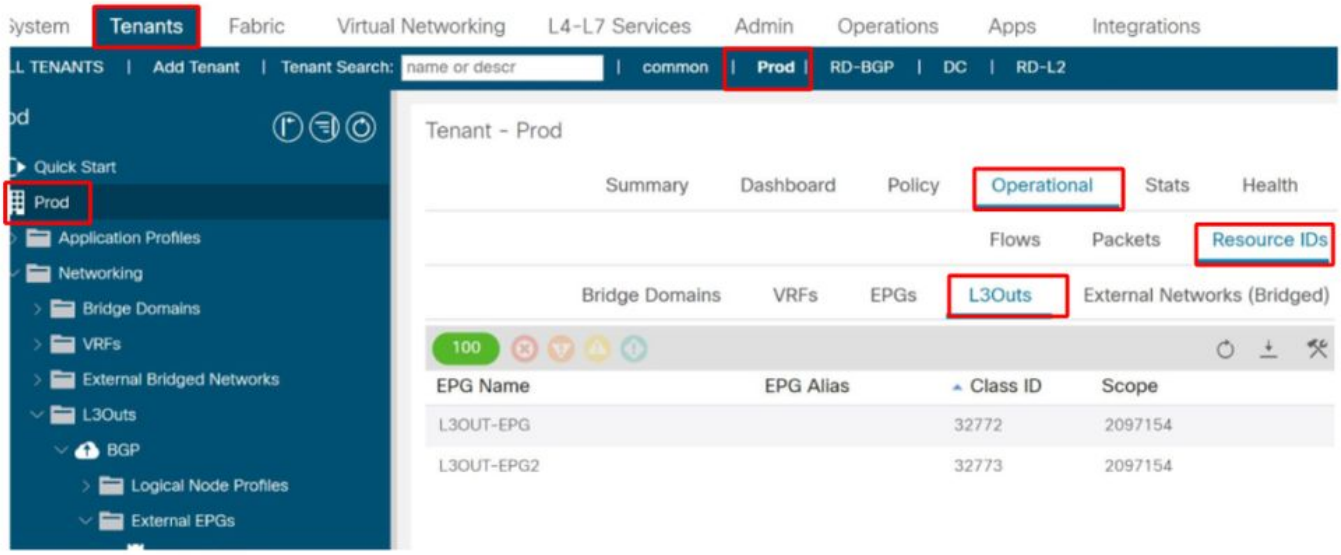
注意：从基于前缀的EPG角度来看，以下两个配置将产生等效的LPM策略管理器前缀表条目：

1. 两个L3Out，各有一个L3Out EPG。
2. 一个L3Out与两个L3Out EPG

在这两种情况下，L3Out EPG总数都是2。这意味着每个都有自己的pcTag和相关子网。

给定的L3Out EPG的所有pcTags都可以在GUI中查看“租户>操作>资源ID > L3Outs”

## 验证L3Out pcTag




在这种情况下，ACI交换矩阵从外部路由器接收多个前缀，L3Out EPG定义如下：

- 172.16.1.0/24分配给L3OUT-EPG。

- 172.16.2.0/24已分配给L3OUT-EPG2。
- 172.16.0.0/16分配给L3OUT-EPG(以捕获172.16.3.0/24前缀)。

要与此匹配，配置定义如下：

- L3OUT-EPG的子网172.16.1.0/24和172.16.0.0/16的范围都是“外部EPG的外部子网”。
- L3OUT-EPG2的子网172.16.2.0/24的范围为“外部EPG的外部子网”。

结果的前缀表条目为：

```
bdsol-aci32-leaf3# vsh -c 'show system internal  policy-mgr prefix' |  egrep "Prod|==|Addr"
Vrf-Vni VRF-Id Table-Id Table-State  VRF-Name                    Addr
Class Shared Remote Complete
======= ======  =========== =======  ============================
================================ ====== ====== ====== ========
2097154 35     0x23          Up     Prod:VRF1
0.0.0.0/0   15      True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.1.0/24  32772   True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.0.0/16  32772   True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.2.0/24  32773   True   True   False
```

172.16.2.0/24分配给pcTag 32773(L3OUT-EPG2),172.16.0.0/16分配给32772(L3OUT-EPG)。

在此方案中，172.16.1.0/24的条目是冗余的，因为/16超网被分配到同一个EPG。

当目标是将不同的合同应用于单个L3Out中的前缀组时，多个L3Out EPG非常有用。下一个示例将说明合同如何与多个L3Out EPG配合使用。

## 示例3b:具有不同合同的多个L3Out EPG

此方案包含以下设置：

- ICMP合同仅允许ICMP。
- 仅允许tcp目标端口80的HTTP合同。
- EPG1(pcTag 32770)提供L3OUT-EPG(pcTag 32772)使用的HTTP合同。
- EPG2(pcTag 32771)提供L3OUT-EPG2(pcTag 32773)使用的ICMP合同。

将使用上一个示例中的相同policymgr前缀：

- L3OUT-EPG中的172.16.1.0/24应允许HTTP到EPG1
- L3OUT-EPG2中的172.16.2.0/24应允许ICMP到EPG2

policy-mgr prefix and zoning-rules:

```
bdsol-aci32-leaf3# vsh -c ' show system internal  policy-mgr prefix ' |  egrep "Prod|==|Addr"
Vrf-Vni VRF-Id Table-Id Table-State  VRF-Name                    Addr
Class Shared Remote Complete
======= ======  =========== =======  ============================
================================ ====== ====== ====== ========
2097154 35     0x23          Up     Prod:VRF1
0.0.0.0/0   15      True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.1.0/24  32772   True   True   False
```

```
2097154 35     0x23          Up     Prod:VRF1
172.16.0.0/16  32772   True   True   False
2097154 35     0x23          Up     Prod:VRF1
172.16.2.0/24  32773   True   True   False


bdsol-aci32-leaf3# show zoning-rule scope 2097154
+---------+--------+--------+----------+----------------+---------+---------+------+----------+-
---------------------+
| Rule ID | SrcEPG | DstEPG | FilterID |      Dir       |  operSt |  Scope  | Name |  Action  |
Priority       |
+---------+--------+--------+----------+----------------+---------+---------+------+----------+-
---------------------+
|   4326  |   0    |   0    | implicit |    uni-dir     | enabled | 2097154 |      | deny,log |
any_any_any(21)    |
|   4335  |   0    | 16387  | implicit |    uni-dir     | enabled | 2097154 |      |  permit  |
any_dest_any(16)   |
|   4334  |   0    |   0    | implarp  |    uni-dir     | enabled | 2097154 |      |  permit  |
any_any_filter(17)  |
|   4333  |   0    |   15   | implicit |    uni-dir     | enabled | 2097154 |      | deny,log |
any_vrf_any_deny(22) |
|   4332  |   0    | 16386  | implicit |    uni-dir     | enabled | 2097154 |      |  permit  |
any_dest_any(16)   |
|   4342  | 32771  | 32773  |    5     | uni-dir-ignore | enabled | 2097154 | ICMP |  permit  |
fully_qual(7)     |
|   4343  | 32773  | 32771  |    5     |     bi-dir     | enabled | 2097154 | ICMP |  permit  |
fully_qual(7)     |
|   4340  | 32770  | 32772  |    38    |    uni-dir     | enabled | 2097154 | HTTP |  permit  |
fully_qual(7)     |
|   4338  | 32772  | 32770  |    37    |    uni-dir     | enabled | 2097154 | HTTP |  permit  |
fully_qual(7)     |
+---------+--------+--------+----------+----------------+---------+---------+------+----------+-
---------------------+
```

## 使用fTriage的数据路径验证 — 策略允许的流量

如果外部网络中的ICMP流量介于172.16.2.1和EPG2中的192.168.3.1之间，则可以使用fTriage捕获和分析该流量。在这种情况下，在枝叶交换机103和104上启动fTriage，因为流量可以进入其中之一：

```
admin@apic1:~> ftriage route -ii LEAF:103,104 -sip 172.16.2.1 -dip 192.168.3.1
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "InProgress", "pid": "14454",
"apicId": "1", "id": "0"}}}
Starting ftriage
Log file name for the current run is: ftlog_2019-10-02-22-30-41-871.txt
2019-10-02 22:30:41,874 INFO     /controller/bin/ftriage route -ii LEAF:103,104 -sip 172.16.2.1
-dip 192.168.3.1
2019-10-02 22:31:28,868 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-02 22:32:15,076 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-leaf3
Ingress: Eth1/12 (Po1) Egress: Eth1/12 (Po1) Vnid: 11365
2019-10-02 22:32:15,295 INFO     ftriage:     main:242  ingress encap string vlan-2551
2019-10-02 22:32:17,839 INFO     ftriage:     main:271  Building ingress BD(s), Ctx
2019-10-02 22:32:20,583 INFO     ftriage:     main:294  Ingress BD(s) Prod:VRF1:l3out-BGP:vlan-
2551
2019-10-02 22:32:20,584 INFO     ftriage:     main:301  Ingress Ctx: Prod:VRF1
2019-10-02 22:32:20,693 INFO     ftriage:   pktrec:490  bdsol-aci32-leaf3: Collecting transient
losses snapshot for LC module: 1
2019-10-02 22:32:38,933 INFO     ftriage:     nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4343 scope:34 filter:5
```

```
2019-10-02 22:32:39,931 INFO     ftriage:     main:522  Computed egress encap string vlan-2502
2019-10-02 22:32:39,933 INFO     ftriage:     main:313  Building egress BD(s), Ctx
2019-10-02 22:32:41,796 INFO     ftriage:     main:331  Egress Ctx Prod:VRF1
2019-10-02 22:32:41,796 INFO     ftriage:     main:332  Egress BD(s): Prod:BD2
2019-10-02 22:32:48,636 INFO     ftriage:     main:933  SIP 172.16.2.1 DIP 192.168.3.1
2019-10-02 22:32:48,637 INFO     ftriage:  unicast:973  bdsol-aci32-leaf3: <- is ingress node
2019-10-02 22:32:51,257 INFO     ftriage:  unicast:1202 bdsol-aci32-leaf3: Dst EP is local
2019-10-02 22:32:54,129 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: EP if(Po1) same as
egr if(Po1)
2019-10-02 22:32:55,348 INFO     ftriage:     misc:657  bdsol-aci32-leaf3:
DMAC(00:22:BD:F8:19:FF) same as RMAC(00:22:BD:F8:19:FF)
2019-10-02 22:32:55,349 INFO     ftriage:     misc:659  bdsol-aci32-leaf3: L3 packet getting
routed/bounced in SUG
2019-10-02 22:32:55,596 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: Dst IP is present in
SUG L3 tbl
2019-10-02 22:32:55,896 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: RW seg_id:11365 in
SUG same as EP segid:11365
2019-10-02 22:33:02,150 INFO     ftriage:     main:961  Packet is Exiting fabric with peer-
device: bdsol-aci32-n3k-3 and peer-port: Ethernet1/16
```

fTriage确认从L3OUT_EPG2到EPG的ICMP规则所匹配的分区规则：

```
2019-10-02 22:32:38,933 INFO     ftriage:     nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4343 scope:34 filter:5
```

### 使用fTriage的数据路径验证 — 策略不允许的流

对于从172.16.1.1(L3OUT-EPG)到192.168.3.1(EPG2)的ICMP流量，预计策略丢弃。

```
admin@apic1:~> ftriage route -ii LEAF:103,104 -sip 172.16.1.1 -dip 192.168.3.1
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "InProgress", "pid": "15139",
"apicId": "1", "id": "0"}}}
Starting ftriage
Log file name for the current run is: ftlog_2019-10-02-22-39-15-050.txt
2019-10-02 22:39:15,056 INFO     /controller/bin/ftriage route -ii LEAF:103,104 -sip 172.16.1.1
-dip 192.168.3.1
2019-10-02 22:40:03,523 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-02 22:40:43,338 ERROR    ftriage:  unicast:234  bdsol-aci32-leaf3: L3 packet getting fwd
dropped, checking drop reason
2019-10-02 22:40:43,339 ERROR    ftriage:  unicast:234  bdsol-aci32-leaf3: L3 packet getting fwd
dropped, checking drop reason
SECURITY_GROUP_DENY              condition setcast:236  bdsol-aci32-leaf3: Drop reason -
SECURITY_GROUP_DENY              condition set
2019-10-02 22:40:43,340 INFO     ftriage:  unicast:252  bdsol-aci32-leaf3: policy drop flow
sclass:32772 dclass:32771 sg_label:34 proto:1
2019-10-02 22:40:43,340 INFO     ftriage:     main:681  : Ftriage Completed with hunch: None
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "Idle", "pid": "0", "apicId": "0",
"id": "0"}}}
```

分类确认数据包已因SECURITY_GROUP_DENY(policy drop)原因而丢弃，并且派生的源pcTag为32772，目标pcTag为32771。根据分区规则检查此情况，这些EPG之间显然没有条目。

```
bdsol-aci32-leaf3# show zoning-rule  scope 2097154 src-epg 32772 dst-epg 32771
+---------+--------+--------+----------+-----+--------+-------+------+--------+----------+
| Rule ID | SrcEPG | DstEPG | FilterID | Dir | operSt | Scope | Name | Action | Priority |
+---------+--------+--------+----------+-----+--------+-------+------+--------+----------+
+---------+--------+--------+----------+-----+--------+-------+------+--------+----------+
```

## 示例 4：多个带有多个前缀的L3Outs

场景设置类似于示例3（L3Out和L3Out EPG定义），但两个L3Out EPG上定义的网络是0.0.0.0/0。

合同配置如下：

- ICMP1合同允许ICMP。
- 允许ICMP的ICMP2合同。
- EPG1(pcTag 32770)提供L3OUT-EPG(pcTag 32772)使用的ICMP1合同。
- EPG2(pcTag 32771)提供L3OUT-EPG2(pcTag 32773)使用的ICMP2合同。

在外部网络通告许多前缀的情况下，此配置可能看起来比较理想，但至少有两个遵循不同允许流模式的前缀块。在本示例中，一个前缀应仅允许ICMP1，另一个前缀应仅允许ICMP2。

尽管在同一VRF中使用'0.0.0.0/0'两次，但只有一个prefix在policy-mgr前缀表中编程：

```
bdsol-aci32-leaf3# vsh -c ' show system internal  policy-mgr prefix ' |  egrep "Prod|==|Addr"
Vrf-Vni VRF-Id Table-Id Table-State  VRF-Name                    Addr
Class Shared Remote Complete
======= ======  =========== =======  ============================
================================ ====== ====== ====== ========
2097154 35     0x23          Up     Prod:VRF1
```

下面重新研究了两个流。根据以上合同配置，预期如下：

1. ICMP2应允许172.16.2.1(L3OUT-EPG2)到192.168.3.1(EPG2)
2. 不应允许172.16.2.1(L3OUT-EPG2)到192.168.1.1(EPG1)，因为EPG1和L3OUT-EPG2之间没有合同

### 使用fTriage的数据路径验证 — 策略允许的流

使用从172.16.2.1(L3OUT-EPG2)到192.168.3.1(EPG2 — pcTag 32771)的ICMP流运行fTriage。

```
Starting ftriage
Log file name for the current run is: ftlog_2019-10-02-23-11-14-298.txt
2019-10-02 23:11:14,302 INFO     /controller/bin/ftriage route -ii LEAF:103,104 -sip 172.16.2.1
-dip 192.168.3.1
2019-10-02 23:12:00,887 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-02 23:12:44,565 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-leaf3
Ingress: Eth1/12 (Po1) Egress: Eth1/12 (Po1) Vnid: 11365
2019-10-02 23:12:44,782 INFO     ftriage:     main:242  ingress encap string vlan-2551
2019-10-02 23:12:47,260 INFO     ftriage:     main:271  Building ingress BD(s), Ctx
2019-10-02 23:12:50,041 INFO     ftriage:     main:294  Ingress BD(s) Prod:VRF1:l3out-BGP:vlan-
2551
2019-10-02 23:12:50,042 INFO     ftriage:     main:301  Ingress Ctx: Prod:VRF1
2019-10-02 23:12:50,151 INFO     ftriage:   pktrec:490  bdsol-aci32-leaf3: Collecting transient
losses snapshot for LC module: 1
2019-10-02 23:13:08,595 INFO     ftriage:     nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4336 scope:34 filter:5
2019-10-02 23:13:09,608 INFO     ftriage:     main:522  Computed egress encap string vlan-2502
2019-10-02 23:13:09,609 INFO     ftriage:     main:313  Building egress BD(s), Ctx
2019-10-02 23:13:11,449 INFO     ftriage:     main:331  Egress Ctx Prod:VRF1
```

```
2019-10-02 23:13:11,449 INFO     ftriage:     main:332  Egress BD(s): Prod:BD2
2019-10-02 23:13:18,383 INFO     ftriage:     main:933  SIP 172.16.2.1 DIP 192.168.3.1
2019-10-02 23:13:18,384 INFO     ftriage:  unicast:973  bdsol-aci32-leaf3: <- is ingress node
2019-10-02 23:13:21,078 INFO     ftriage:  unicast:1202 bdsol-aci32-leaf3: Dst EP is local
2019-10-02 23:13:23,926 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: EP if(Po1) same as
egr if(Po1)
2019-10-02 23:13:25,216 INFO     ftriage:     misc:657  bdsol-aci32-leaf3:
DMAC(00:22:BD:F8:19:FF) same as RMAC(00:22:BD:F8:19:FF)
2019-10-02 23:13:25,217 INFO     ftriage:     misc:659  bdsol-aci32-leaf3: L3 packet getting
routed/bounced in SUG
2019-10-02 23:13:25,465 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: Dst IP is present in
SUG L3 tbl
2019-10-02 23:13:25,757 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: RW seg_id:11365 in
SUG same as EP segid:11365
2019-10-02 23:13:32,235 INFO     ftriage:     main:961  Packet is Exiting fabric with peer-
device: bdsol-aci32-n3k-3 and peer-port: Ethernet1/16
```

分区规则4336允许（按照预期）此流。

### 使用fTriage的数据路径验证 — 策略不允许的流

使用从172.16.2.1(L3OUT-EPG2)到192.168.1.1(EPG1 — pcTag 32770)的ICMP流运行fTriage:

```
admin@apic1:~> ftriage route -ii LEAF:103,104 -sip 172.16.2.1 -dip 192.168.1.1
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "InProgress", "pid": "31500",
"apicId": "1", "id": "0"}}}
Starting ftriage
Log file name for the current run is: ftlog_2019-10-02-23-53-03-478.txt
2019-10-02 23:53:03,482 INFO     /controller/bin/ftriage route -ii LEAF:103,104 -sip 172.16.2.1
-dip 192.168.1.1
2019-10-02 23:53:50,014 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-02 23:54:39,199 INFO     ftriage:     main:839  L3 packet Seen on bdsol-aci32-leaf3
Ingress: Eth1/12 (Po1) Egress: Eth1/12 (Po1) Vnid: 11364
2019-10-02 23:54:39,417 INFO     ftriage:     main:242  ingress encap string vlan-2551
2019-10-02 23:54:41,962 INFO     ftriage:     main:271  Building ingress BD(s), Ctx
2019-10-02 23:54:44,765 INFO     ftriage:     main:294  Ingress BD(s) Prod:VRF1:l3out-BGP:vlan-
2551
2019-10-02 23:54:44,766 INFO     ftriage:     main:301  Ingress Ctx: Prod:VRF1
2019-10-02 23:54:44,875 INFO     ftriage:   pktrec:490  bdsol-aci32-leaf3: Collecting transient
losses snapshot for LC module: 1
2019-10-02 23:55:02,905 INFO     ftriage:     nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4341 scope:34 filter:5
2019-10-02 23:55:04,525 INFO     ftriage:     main:522  Computed egress encap string vlan-2501
2019-10-02 23:55:04,526 INFO     ftriage:     main:313  Building egress BD(s), Ctx
2019-10-02 23:55:06,390 INFO     ftriage:     main:331  Egress Ctx Prod:VRF1
2019-10-02 23:55:06,390 INFO     ftriage:     main:332  Egress BD(s): Prod:BD1
2019-10-02 23:55:13,571 INFO     ftriage:     main:933  SIP 172.16.2.1 DIP 192.168.1.1
2019-10-02 23:55:13,572 INFO     ftriage:  unicast:973  bdsol-aci32-leaf3: <- is ingress node
2019-10-02 23:55:16,159 INFO     ftriage:  unicast:1202 bdsol-aci32-leaf3: Dst EP is local
2019-10-02 23:55:18,949 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: EP if(Po1) same as
egr if(Po1)
2019-10-02 23:55:20,126 INFO     ftriage:     misc:657  bdsol-aci32-leaf3:
DMAC(00:22:BD:F8:19:FF) same as RMAC(00:22:BD:F8:19:FF)
2019-10-02 23:55:20,126 INFO     ftriage:     misc:659  bdsol-aci32-leaf3: L3 packet getting
routed/bounced in SUG
2019-10-02 23:55:20,395 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: Dst IP is present in
SUG L3 tbl
2019-10-02 23:55:20,687 INFO     ftriage:     misc:657  bdsol-aci32-leaf3: RW seg_id:11364 in
SUG same as EP segid:11364
```

```
2019-10-02 23:55:26,982 INFO     ftriage:     main:961  Packet is Exiting fabric with peer-
device: bdsol-aci32-n3k-3 and peer-port: Ethernet1/16
```

通过zoning-rule 4341允许（意外）此流。现在必须分析分区规则才能理解原因。

### 数据路径验证 — zoning-rules

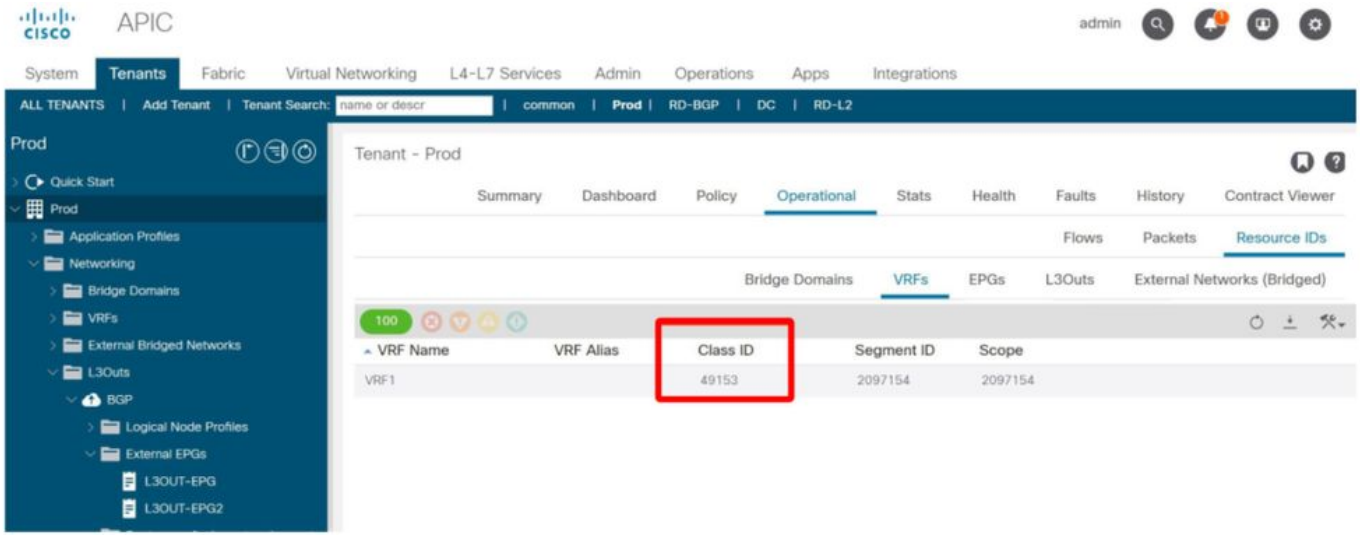
与最后两项测试对应的分区规则如下：

- 预期 — 流命中分区规则行4336（ICMP2合同）。
- 意外 — 流命中分区规则行4341（ICMP1合同）。

```
+---------+--------+--------+----------+---------+---------+---------+-------+----------+-------
---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |   Dir   |  operSt |  Scope  |  Name |  Action  |
Priority       |
+---------+--------+--------+----------+---------+---------+---------+-------+----------+-------
---------------+
|   4326  |   0    |   0    | implicit | uni-dir | enabled | 2097154 |       | deny,log |
any_any_any(21)    |
|   4335  |   0    | 16387  | implicit | uni-dir | enabled | 2097154 |       |  permit  |
any_dest_any(16)   |
|   4334  |   0    |   0    | implarp  | uni-dir | enabled | 2097154 |       |  permit  |
any_any_filter(17)  |
|   4333  |   0    |   15   | implicit | uni-dir | enabled | 2097154 |       | deny,log |
any_vrf_any_deny(22) |
|   4332  |   0    | 16386  | implicit | uni-dir | enabled | 2097154 |       |  permit  |
any_dest_any(16)   |
|   4339  | 32770  |   15   |    5     | uni-dir | enabled | 2097154 | ICMP2 |  permit  |
fully_qual(7)     |
|   4341  | 49153  | 32770  |    5     | uni-dir | enabled | 2097154 | ICMP2 |  permit  |
fully_qual(7)     |
|   4337  | 32771  |   15   |    5     | uni-dir | enabled | 2097154 | ICMP1 |  permit  |
fully_qual(7)     |
|   4336  | 49153  | 32771  |    5     | uni-dir | enabled | 2097154 | ICMP1 |  permit  |
fully_qual(7)     |
+---------+--------+--------+----------+---------+---------+---------+-------+----------+-------
---------------+
```

两个流都派生出src pcTag 49153。这是VRF的pcTag。这可以在UI中验证：

## 验证VRF的pcTag




当0.0.0.0/0前缀与L3Out一起使用时，会发生以下情况：

- 从内部EPG到0.0.0.0/0的L3Out EPG的流量将派生目标pcTag 15。
- 从0.0.0.0/0的L3Out EPG到ACI内部EPG的流量将生成VRF(49153)的源pcTag。

contract_parser脚本提供了分区规则的整体视图：

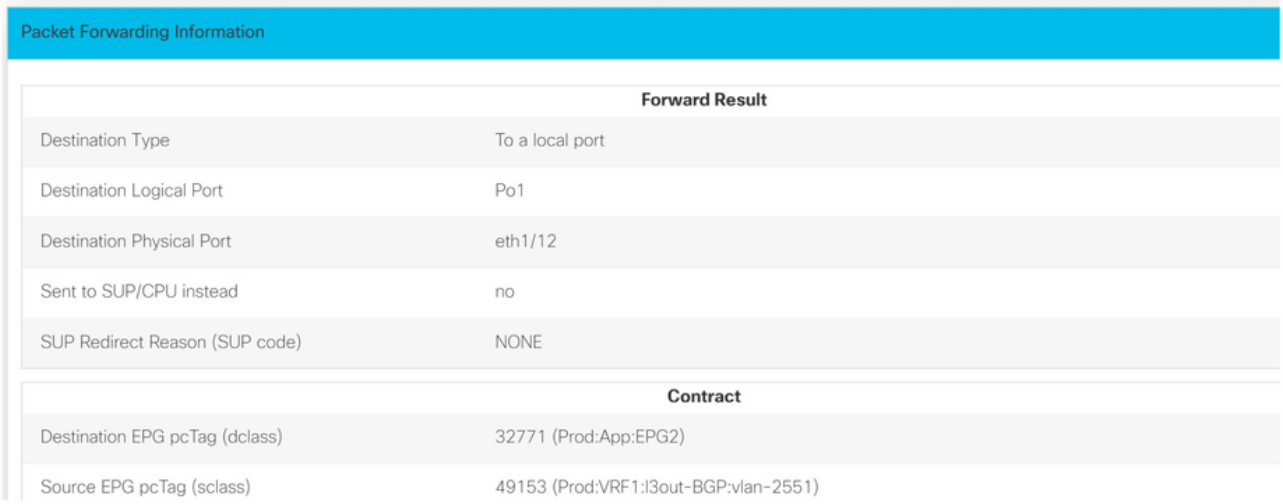
```
bdsol-aci32-leaf3# contract_parser.py --vrf Prod:VRF1
Key:
[prio:RuleId] [vrf:{str}] action protocol src-epg [src-l4] dst-epg [dst-l4]
[flags][contract:{str}] [hit=count]
[7:4339] [vrf:Prod:VRF1] permit ip icmp tn-Prod/ap-App/epg-EPG1(32770) pfx-0.0.0.0/0(15)
[contract:uni/tn-Prod/brc-ICMP2] [hit=0]
[7:4337] [vrf:Prod:VRF1] permit ip icmp tn-Prod/ap-App/epg-EPG2(32771) pfx-0.0.0.0/0(15)
[contract:uni/tn-Prod/brc-ICMP] [hit=0]
[7:4341] [vrf:Prod:VRF1] permit ip icmp tn-Prod/vrf-VRF1(49153) tn-Prod/ap-App/epg-EPG1(32770)
[contract:uni/tn-Prod/brc-ICMP2] [hit=270]
[7:4336] [vrf:Prod:VRF1] permit ip icmp tn-Prod/vrf-VRF1(49153) tn-Prod/ap-App/epg-EPG2(32771)
[contract:uni/tn-Prod/brc-ICMP] [hit=0]
```

### 使用ELAM Assistant应用确认数据包使用的pcTag

ELAM Assistant应用提供了另一种方法来确认实时流量的源和目标pcTag。

以下屏幕截图显示从pcTag到pcTag的流量32771ELAM结49153。

## 从src到dst的ELAM助理应32771输出更49153



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 (Prod:App:EPG2) |
| Source EPG pcTag (sclass) | 49153 (Prod:VRF1:l3out-BGP:vlan-2551) |

### 结论

必须在VRF中仔细跟踪0.0.0.0/0的使用情况，因为使用该子网的每个L3Out都将继承应用于使用该子网的其他L3Out的合同。这可能会导致计划外的许可流。

# 共享L3Out

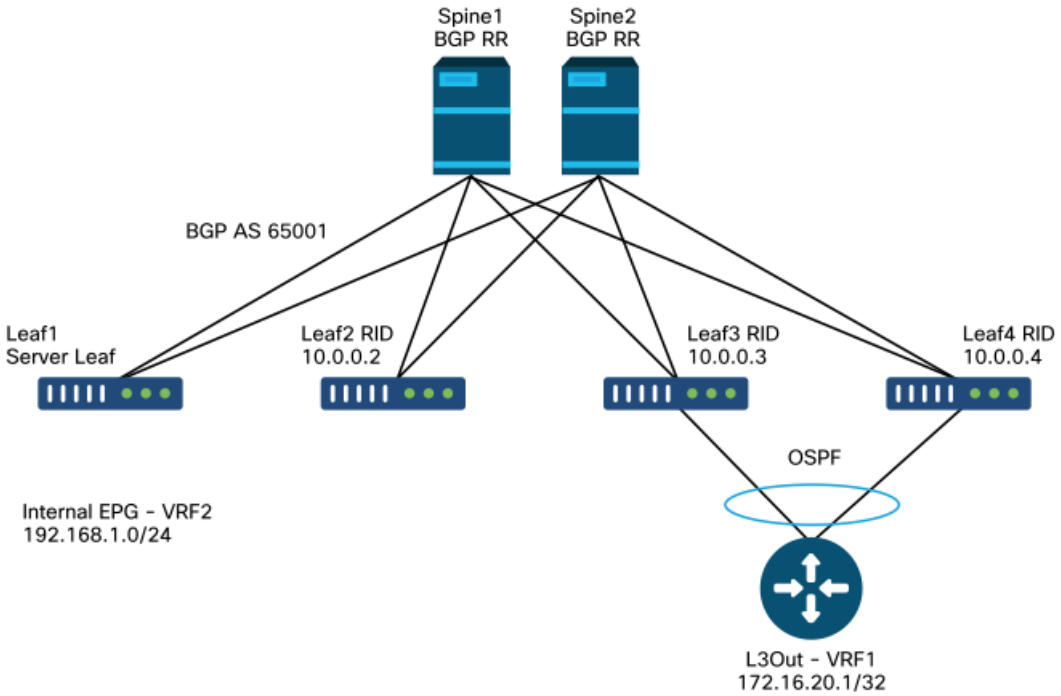
## 概述

本节将讨论如何排除共享L3Out配置中的路由通告故障。术语“共享L3Out”是指以下场景：L3Out位于一个VRF中，但与L3Out有合同的内部EPG位于另一个VRF中。使用共享L3Outs时，路由泄漏在内部对ACI交换矩阵执行。

本节不会深入介绍有关安全策略故障排除的详细信息。有关信息，请参阅本书的“安全策略”一章。出于安全考虑，本节还将不详细介绍外部策略前缀分类。请参阅“外部转发”一章中的“合同和L3Out”一节。

本部分使用以下拓扑作为示例。

## 共享L3Out拓扑




在较高层面上，必须配置以下配置才能使共享L3Out正常工作：

- L3Out子网必须配置有“共享路由控制子网”范围，以便将外部路由泄漏到内部VRF。也可以选择“Aggregate Shared”（聚合共享）选项，以泄漏比配置的子网更具体的全部路由。
- L3Out子网必须配置有“共享安全导入子网”(Shared Security Import Subnet)范围，以编程必要的安全策略，以允许通过此L3Out进行通信。
- 内部BD子网必须设置为“在VRF之间共享”和“对外通告”，以对外部VRF中的BD子网进行编程并通告该子网。
- 必须在共享L3Out的内部EPG和外部EPG之间配置“租户”或“全局”范围合同。

下一节将详细介绍如何在ACI中通告和获知泄漏的路由。

## 共享的L3Out工作流程 — 学习外部路由

本节将概述在将已了解的外部路由通告到交换矩阵时的路径。

### 在边界枝叶上看到的外部路由

此命令将显示从OSPF获知的外部路由：

```
leaf103# show ip route 172.16.20.1/32 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]
'%' in via output denotes VRF
```

```
172.16.20.1/32, ubest/mbest: 1/0
    *via 10.10.34.3, vlan347, [110/20], 03:59:59, ospf-default, type-2
```

接下来，必须将路由导入BGP。默认情况下，所有外部路由都应导入BGP。

### 边界枝叶上的BGP验证

路由必须位于BGP VPNv4 Address-family中，并且路由目标将分布在整个交换矩阵中。路由目标是由外部VRF导出并由需要接收路径的任何内部VRF导入的BGP扩展社区。

接下来，验证由BL上的外部VRF导出的路由目标。

```
leaf103# show bgp process vrf Prod:Vrf1

Information regarding configured VRFs:

BGP Information for VRF Prod:Vrf1
VRF Type                        : System
VRF Id                          : 85
VRF state                       : UP
VRF configured                  : yes
VRF refcount                    : 1
VRF VNID                        : 2392068
Router-ID                       : 10.0.0.3
Configured Router-ID            : 10.0.0.3
Confed-ID                       : 0
Cluster-ID                      : 0.0.0.0
MSITE Cluster-ID                : 0.0.0.0
No. of configured peers         : 1
No. of pending config peers     : 0
No. of established peers        : 0
VRF RD                          : 101:2392068
VRF EVPN RD                     : 101:2392068

...

    Wait for IGP convergence is not configured
    Export RT list:
        65001:2392068
    Import RT list:
        65001:2392068
    Label mode: per-prefix
```

上述输出显示，从外部VRF通告到VPNv4的所有路径都应收到路由目标65001:2392068。

接下来，检验bgp路径：

```
leaf103# show bgp ipv4 unicast 172.16.20.1/32 vrf Prod:Vrf1
BGP routing table information for VRF Prod:Vrf1, address family IPv4 Unicast
BGP routing table entry for 172.16.20.1/32, version 30 dest ptr 0xa6f25ad0
Paths: (2 available, best #1)
Flags: (0x80c0002 00000000) on xmit-list, is not in urib, exported
  vpn: version 17206, (0x100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type: redist 0x408 0x1 ref 0 adv path ref 2, path is valid, is best path
```

```
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (10.0.0.3)
      Origin incomplete, MED 20, localpref 100, weight 32768
      Extcommunity:
          RT:65001:2392068
          VNID:2392068
          COST:pre-bestpath:162:110

  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    10.0.64.64         10.0.72.66
  Path-id 2 not advertised to any peer
```

以上输出显示路径具有正确的路由目标。也可以使用“show bgp vpnv4 unicast 172.16.20.1 vrf overlay-1”命令验证VPNv4路径。

### 服务器枝叶上的验证

对于要安装BL通告路由的内部EPG枝叶，必须将路由目标（如上所述）导入内部VRF。可以检查内部VRF的BGP进程以验证以下内容：

```
leaf101# show bgp process vrf Prod:Vrf2

Information regarding configured VRFs:

BGP Information for VRF Prod:Vrf2
VRF Type                        : System
VRF Id                          : 54
VRF state                       : UP
VRF configured                  : yes
VRF refcount                    : 0
VRF VNID                        : 2916352
Router-ID                       : 192.168.1.1
Configured Router-ID            : 0.0.0.0
Confed-ID                       : 0
Cluster-ID                      : 0.0.0.0
MSITE Cluster-ID                : 0.0.0.0
No. of configured peers         : 0
No. of pending config peers     : 0
No. of established peers        : 0
VRF RD                          : 102:2916352
VRF EVPN RD                     : 102:2916352
...
    Wait for IGP convergence is not configured
    Import route-map 2916352-shared-svc-leak
    Export RT list:
        65001:2916352
    Import RT list:
        65001:2392068
        65001:2916352
```

以上输出显示导入由外部VRF导出的路由目标的内部VRF。此外，还引用了“导入路由映射”。导入路由映射包括在共享L3Out中定义的特定前缀，带有“共享路由控制子网”标志。

可以检查路由映射内容，以确保它包含外部前缀：

```
leaf101# show route-map 2916352-shared-svc-leak
route-map 2916352-shared-svc-leak, deny, sequence 1
 Match clauses:
   pervasive: 2
 Set clauses:
route-map 2916352-shared-svc-leak, permit, sequence 2
 Match clauses:
   extcommunity  (extcommunity-list filter): 2916352-shared-svc-leak
 Set clauses:
route-map 2916352-shared-svc-leak, permit, sequence 1000
 Match clauses:
   ip address prefix-lists: IPv4-2392068-16387-5511-2916352-shared-svc-leak
   ipv6 address prefix-lists: IPv6-deny-all
 Set clauses:
a-leaf101# show ip prefix-list IPv4-2392068-16387-5511-2916352-shared-svc-leak
ip prefix-list IPv4-2392068-16387-5511-2916352-shared-svc-leak: 1 entries
  seq 1 permit 172.16.20.1/32
```

以上输出显示导入路由映射，其中包括要导入的子网。

最终验证包括检查路由是否在BGP表中以及是否安装在路由表中。

服务器枝叶上的BGP表：

```
leaf101# show bgp ipv4 unicast 172.16.20.1/32 vrf Prod:Vrf2
BGP routing table information for VRF Prod:Vrf2, address family IPv4 Unicast
BGP routing table entry for 172.16.20.1/32, version 3 dest ptr 0xa763add0
Paths: (2 available, best #1)
Flags: (0x08001a 00000000) on xmit-list, is in urib, is best urib route, is in HW
  vpn: version 10987, (0x100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type: internal 0xc0000018 0x40 ref 56506 adv path ref 2, path is valid, is best path
             Imported from 10.0.72.64:5:172.16.20.1/32
  AS-Path: NONE, path sourced internal to AS
    10.0.72.64 (metric 3) from 10.0.64.64 (192.168.1.102)
      Origin incomplete, MED 20, localpref 100, weight 0
      Received label 0
      Received path-id 1
      Extcommunity:
          RT:65001:2392068
          VNID:2392068
          COST:pre-bestpath:162:110
      Originator: 10.0.72.64 Cluster list: 192.168.1.102
```

路由会导入到内部VRF BGP表，并具有预期的路由目标。

可以验证安装的路由：

```
leaf101# vsh -c "show ip route 172.16.20.1/32 detail vrf Prod:Vrf2"
IP Route Table for VRF "Prod:Vrf2"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%' in via output denotes VRF
172.16.20.1/32, ubest/mbest: 2/0
```

```
   *via 10.0.72.64%overlay-1, [200/20], 01:00:51, bgp-65001, internal, tag 65001 (mpls-vpn)
        MPLS[0]: Label=0 E=0 TTL=0 S=0 (VPN)
        client-specific data: 548
        recursive next hop: 10.0.72.64/32%overlay-1
        extended route information: BGP origin AS 65001 BGP peer AS 65001 rw-vnid: 0x248004
table-id: 0x36 rw-mac: 0
   *via 10.0.72.67%overlay-1, [200/20], 01:00:51, bgp-65001, internal, tag 65001 (mpls-vpn)
        MPLS[0]: Label=0 E=0 TTL=0 S=0 (VPN)
        client-specific data: 54a
        recursive next hop: 10.0.72.67/32%overlay-1
        extended route information: BGP origin AS 65001 BGP peer AS 65001 rw-vnid: 0x248004
table-id: 0x36 rw-mac: 0
```

上述输出使用特定的“vsh -c”命令获取“detail”输出。“detail”标志包括重写VXLAN VNID。这是外部VRF的VXLAN VNID。当BL收到具有此VNID的数据平面流量时，它知道在外部VRF中进行转发决策。

rw-vnid值以十六进制表示，因此转换为十进制会获取2392068的VRF VNID。使用“show system internal epm vrf all”搜索相应的VRF 枝叶上的| grep 2392068'。可以使用'moquery -c fvCtx -f 'fv.Ctx.seg=="2392068"命令对APIC执行全局搜索。

下一跳的IP也应指向BL PTEP，并且“%overlay-1”表示下一跳的路由查找在重叠VRF中。

## 共享的L3Out工作流 — 通告内部路由

与前面部分一样，将内部BD子网从共享L3Out中通告出去的工作方式如下：

- BD子网（内部VRF）作为静态路由安装在BL（外部VRF）上。此静态路由部署是内部EPG和L3Out之间合同关系的结果。
- 当BD子网上设置了“Advertised External”范围时，静态路由会重分布到外部协议中。

### 检验BL上的BD静态路由

```
leaf103# vsh -c "show ip route 192.168.1.0 detail 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]
'%' in via output denotes VRF

192.168.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive
    *via 10.0.120.34%overlay-1, [1/0], 00:55:27, static, tag 4294967292
         recursive next hop: 10.0.120.34/32%overlay-1
         vrf crossing information:  VNID:0x2c8000 ClassId:0 Flush#:0
```

请注意，在上述输出中，内部VRF的VNID已设置为重写。下一跳也设置为代理v4任播地址。

以上路由通过“路由通告”部分所示的相同路由映射在外部进行通告。

如果将BD子网设置为“外部通告”，则会将其重新分发到内部EPG与其有合同关系的每个**L3Out**的外部协议中。

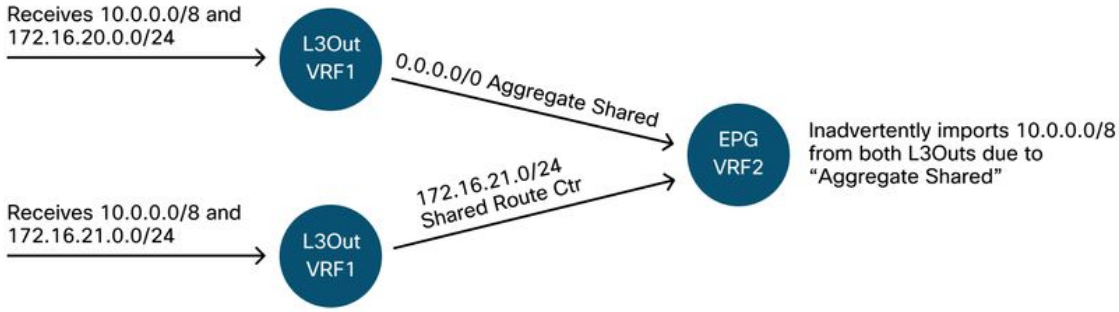
## 共享L3Out故障排除场景 — 意外的路由泄漏

此场景在外部VRF中有多个L3Out，并且内部EPG正在接收来自L3Out的路**由，其中网络未使用“共享”**范围选项进行定义。

**“聚合共享”的使用**

请看下图：

## 意外的路由泄漏




BGP导入映射在VRF级别应用，其中前缀列表是通过**“共享路由控制子**网”标志编程的。如果VRF1中的一个L3Out具有包含“共享路由控制子网”的子网，则VRF1内L3Out上接收的所有匹配此共享路由控制子网的路由都将导入VRF2。

上述设计可能导致意外的流量。如果内部EPG与意外的通告L3Out EPG之间没有合同，则会发生流量丢弃。

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）。