

# 排除ACI外部轉發故障

## 目錄

[簡介](#)

[背景資訊](#)

[概觀](#)

[L3Out元件](#)

[L3Out的主要元件](#)

[外部路由](#)

[高級外部路由流](#)

[L3Out EPG配置選項](#)

[正在定義的L3Out子網包括「scope」定義](#)

[本節中使用的L3Out拓撲](#)

[L3Out拓撲](#)

[相鄰關係](#)

[BGP](#)

[對等連線配置檔案 — Local-AS](#)

[對等連線配置檔案 — 遠端AS](#)

[L3Out — BGP對等連線配置檔案](#)

[邏輯節點配置檔案 — 節點關聯](#)

[BGP CLI驗證 \( 使用回送的eBGP範例 \)](#)

[OSPF](#)

[L3Out — OSPF介面配置檔案 — 區域ID和型別](#)

[邏輯介面配置檔案 — SVI](#)

[OSPF介面配置檔案](#)

[OSPF介面配置檔案 — Hello/Dead計時器和網路型別](#)

[OSPF介面策略詳細資訊](#)

[OSPF CLI驗證](#)

[EIGRP](#)

[EIGRP介面配置檔案](#)

[EIGRP CLI驗證](#)

[路由通告](#)

[網橋域路由通告工作流程](#)

[在應用L3Out和內部EPG之間的合約之前](#)

[在應用L3Out和內部EPG之間的合約之後](#)

[在BD子網中選擇「外部通告」後](#)

[將L3Out關聯到BD之後](#)

[BGP路由通告](#)

[EIGRP路由通告](#)

[網橋域L3配置](#)

[網橋域路由通告故障排除場景](#)

[預設匯出拒絕路由配置檔案](#)

[外部路由匯入工作流程](#)

[路由安裝在BL路由表中](#)

[驗證內部枝葉上的路由](#)

[外部路由故障排除場景](#)

[運輸路線廣告 workflow](#)

[傳輸路由拓撲](#)

[路由標籤策略](#)

[匯出路由控制](#)

[接收和通告BL時的傳輸路由是相同的](#)

[傳輸路由故障排除#1案：未通告傳輸路由](#)

[傳輸路由故障排除#2案：未收到傳輸路由](#)

[具有單個VRF的外部路由器 — 未收到傳輸路由](#)

[傳輸路由疑難解答方#37 — 意外通告的傳輸路由](#)

[合約和L3Out](#)

[L3Out上基於字首的EPG](#)

[L3Out的pcTag的位置](#)

[範例 1：具有特定首碼的單個L3Out](#)

[具有「外部EPG的外部子網」範圍的子網](#)

[範例 2：具有多個字首的單個L3Out](#)

[範例3a:VRF中的多個L3Out EPG](#)

[驗證L3Out pcTag](#)

[範例3b:具有不同合約的多個L3Out EPG](#)

[使用fTriage驗證資料路徑 — 策略允許的流](#)

[使用fTriage的資料路徑驗證 — 策略不允許的流](#)

[範例 4:有多個帶有多個字首的L3Outs](#)

[使用fTriage的資料路徑驗證 — 策略允許的流](#)

[使用fTriage的資料路徑驗證 — 策略不允許的流](#)

[資料路徑驗證 — zoning-rules](#)

[驗證VRF的pcTag](#)

[使用ELAM Assistant應用確認資料包使用的pcTag](#)

[用於從src到dst的ELAM助理應32771程式輸49153](#)

[結論](#)

[共用L3Out](#)

[概觀](#)

[共用L3Out拓撲](#)

[共用的L3Out workflow — 學習外部路由](#)

[在邊界枝葉上顯示的外部路由](#)

[邊界枝葉上的BGP驗證](#)

[伺服器枝葉上的驗證](#)

[共用的L3Out workflow — 通告內部路由](#)

[檢驗BL上的BD靜態路由](#)

[共用的L3Out故障排除場景 — 意外的路由洩漏](#)

[「聚合共用」的使用](#)

[意外的路由洩漏](#)

## 簡介

本文檔介紹瞭解ACI中的L3out並對其進行故障排除的步驟

## 背景資訊

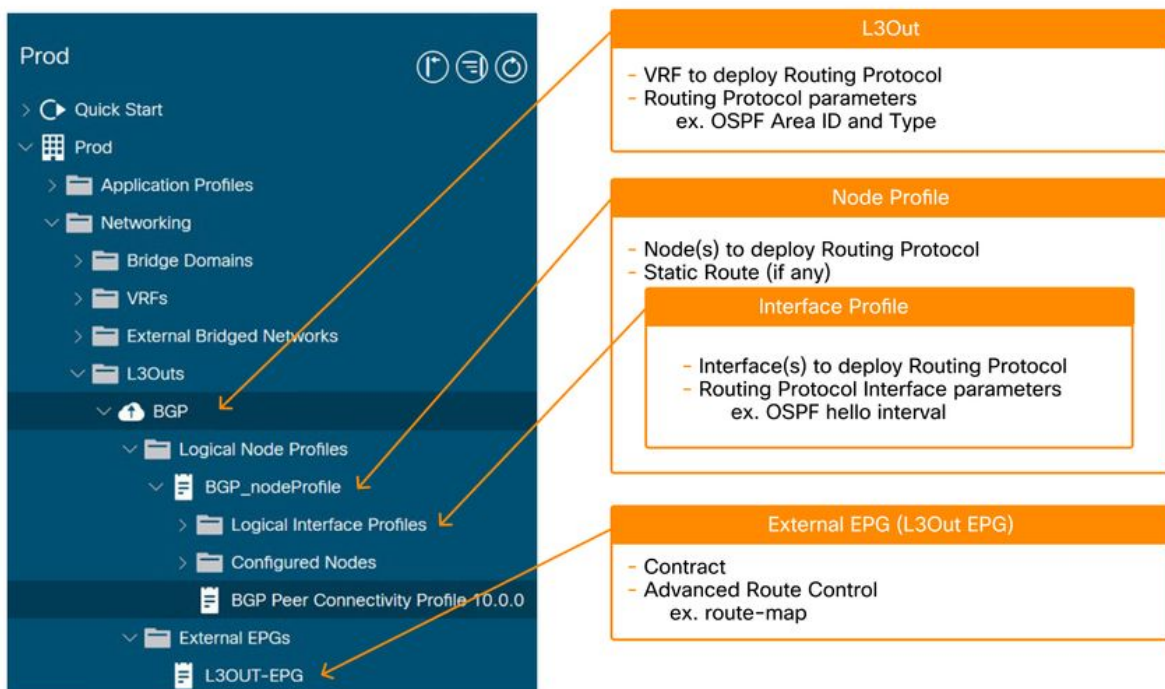
本文檔中的資料摘自[Troubleshooting Cisco Application Centric Infrastructure, Second Edition](#)書籍，具體來說是External Forwarding - Overview， External Forwarding - Adjacency， External Forwarding - Route advertisement， External Forwarding - Contract和L3out 和External Forwarding - Share L3out章。

## 概觀

### L3Out元件

下圖說明了配置L3 Outside(L3Out)所需的主要構建塊。

### L3Out的主要元件

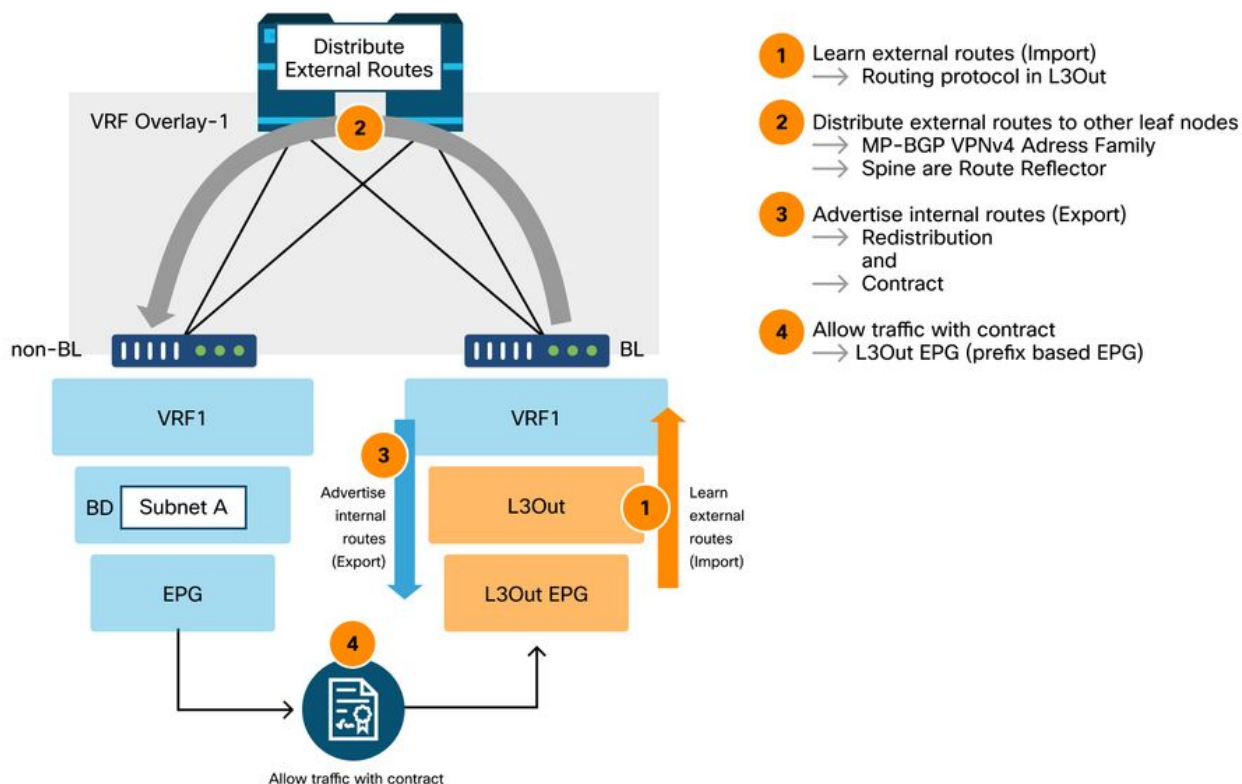


1. L3Out的根：選擇要部署的路由協定（例如OSPF、BGP）。選擇VRF以部署路由協定。選擇L3Out域以定義L3Out的可用枝葉介面和VLAN。
2. 節點配置檔案：選擇枝葉交換機以部署路由協定。這些交換機通常稱為「邊界枝葉交換機」(BL)。為每個邊界枝葉上的路由協定配置Router-ID(RID)。與普通路由器不同，ACI不會根據交換機上的IP地址自動分配路由器ID。配置靜態路由。
3. 介面配置檔案：配置枝葉介面以運行路由協定。  
即介面型別（SVI、路由埠、子介面）、介面ID和IP地址等。為介面級路由協定引數（如hello間隔）選擇策略。
4. 外部EPG(L3Out EPG): 「外部EPG」是部署與L3Out關聯的所有策略（例如IP地址或SVI）以建立鄰居的硬性要求。稍後將詳細介紹如何使用外部EPG。

## 外部路由

下圖顯示外部工藝路線所涉及的高級工序。

## 高級外部路由流



1. BL將與外部路由器建立路由協定鄰接關係。
2. 路由字首從外部路由器接收並作為VPNv4地址系列路徑注入到MP-BGP中。至少必須將兩個主幹節點配置為BGP路由反射器，以便將外部路由反射到所有枝葉節點。
3. 從其他L3Out收到的內部字首（BD子網）和/或字首必須顯式重新分發到路由協定中，以通告給外部路由器。
4. 安全實施：L3Out至少包含一個L3Out EPG。必須在L3Out EPG上使用或提供協定（從類名稱也稱為I3extInstP），以允許流量進/出L3Out。

## L3Out EPG配置選項

在L3Out EPG部分中，可以使用一系列「範圍」和「聚合」選項定義子網，如下所示：

正在定義的L3Out子網包括「scope」定義

# Create Subnet



IP Address:   
address/mask

Name:

scope:  Export Route Control Subnet  
 Import Route Control Subnet  
 External Subnets for the External EPG  
 Shared Route Control Subnet  
 Shared Security Import Subnet

BGP Route Summarization Policy:

aggregate:  Aggregate Export  
 Aggregate Import  
 Aggregate Shared Routes

Route Control Profile:

Name	Direction
------	-----------

「範圍」選項：

- **匯出路由控制子網**：此範圍是通過L3Out將子網從ACI通告（匯出）到外部。雖然這主要用於傳輸路由，但也可以用於通告BD子網，如「ACI BD子網通告」一節所述。
- **匯入路由控制子網**：此範圍用於從L3Out學習（匯入）外部子網。預設情況下，此作用域被禁用，因此呈灰色顯示，並且邊界枝葉(BL)從路由協定獲取任何路由。當需要限制通過OSPF和BGP獲知的外部路由時，可以啟用此範圍。EIGRP不支援此功能。要使用此範圍，需要在給定L3Out上首先啟用「匯入路由控制實施」。
- **外部EPG的外部子網（匯入安全）**：此作用域用於允許具有已配置子網的資料包通過合約從或到L3Out。它根據子網將資料包分類到已配置的L3Out EPG中，以便可以將L3Out EPG上的合約應用於該資料包。此範圍是最長字首匹配，而不是像路由表的其他範圍那樣完全匹配。如果10.0.0.0/16在L3Out EPG A中配置了「外部EPG的外部子網」，則在該子網中具有IP的任何資料包（如10.0.1.1）都將分類到L3Out EPG A中，以在其上使用合約。這並不是說「外部EPG的外部子網」作用域將路由10.0.0.0/16安裝在路由表中。它將建立一個不同的內部表，以便僅根據合約將子網對映到EPG(pcTag)。它對路由協定行為沒有任何影響。此作用域將在學習子網的L3Out上配置。
- **共用路由控制子網**：此作用域是將外部子網洩漏到另一個VRF。ACI使用MP-BGP和路由目標將外部路由從一個VRF洩漏到另一個VRF。此作用域建立帶有子網的字首清單，該字首清單用作過濾器，用於匯出/匯入MP-BGP中具有路由目標的路由。此作用域將在學習原始VRF中的子網的L3Out上配置。
- **共用安全匯入子網**：此作用域用於當資料包通過L3Out的VRF移動時，允許具有已配置子網的資料包。路由表中的路由會洩漏給另一個具有上述「共用路由控制子網」的VRF。但是，另一個VRF尚未知道洩漏的路由應屬於哪個EPG。「共用安全匯入子網」將洩漏路由所屬的L3Out EPG通知另一個VRF。因此，僅當還使用「外部EPG的外部子網」時，才能使用此範圍，否則原始VRF不知道該子網屬於哪個L3Out EPG。此範圍也是最長字首匹配。

「聚合」選項：

- **彙總匯出**：此選項只能用於0.0.0.0/0和「匯出路由控制子網」。為0.0.0.0/0同時啟用「匯出路

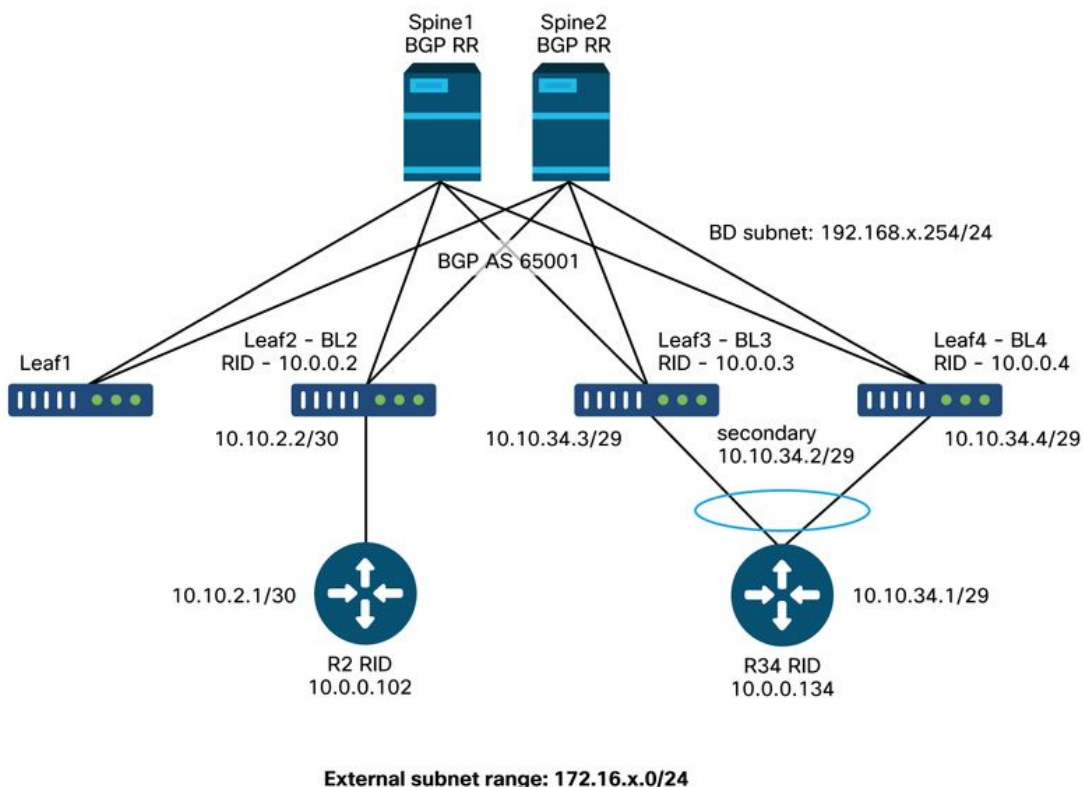
由控制子網」和「聚合匯出」時；它會建立一個字首清單，其中包含「0.0.0.0/0 le 32」，該字首清單匹配任何子網。因此，當L3Out需要向外部通告（匯出）任何路由時，可以使用此選項。當需要更精細的聚合時，可以使用帶有明確字首清單的路由對映/配置檔案。

- **聚合匯入**：此選項只能用於具有「匯入路由控制子網」的0.0.0.0/0。當為0.0.0.0/0同時啟用「匯入路由控制子網」和「聚合匯入」時，它會建立字首清單，該字首清單的「0.0.0.0/0 le 32」匹配任何子網。因此，當L3Out需要從外部學習（匯入）任何路由時，可以使用此選項。但是，通過禁用預設的「匯入路由控制實施」，也可以完成相同的事情。當需要更精細的聚合時，可以使用帶有明確字首清單的路由對映/配置檔案。
- **聚合共用路由**：此選項可用於具有「共用路由控制子網」的任何子網。例如，當為10.0.0.0/8同時啟用「共用路由控制子網」和「聚合共用路由」時，它會建立字首清單，該字首清單的「10.0.0.0/8 le 32」與10.0.0.0/8、10.1.0.0/16等匹配。

## 本節中使用的L3Out拓撲

本節將使用以下拓撲：

### L3Out拓撲



## 相鄰關係

本節介紹如何對L3Out介面上的路由協定鄰接關係進行故障排除和驗證。

下面是幾個要檢查的引數，適用於所有ACI外部路由協定：

- **路由器ID**：在ACI中，即使路由協定不同，每個L3Out也需要在同一枝葉上的相同VRF中使用相

同的路由器ID。此外，同一枝葉上只有一個L3Outs可以使用路由器ID（通常為BGP）建立環回。

- **MTU**：雖然MTU的硬性要求僅用於OSPF鄰接關係，但建議匹配所有路由協定的MTU，以確保用於路由交換/更新任何巨型資料包無需分段即可傳輸，因為大多數控制平面幀將使用DF（不分段）位集傳送，如果幀大小超過介面的最大MTU，該幀將丟棄該幀。
- **MP-BGP路由器反射器**：如果沒有此功能，BGP進程不會在葉節點上啟動。雖然OSPF或EIGRP僅建立鄰居並不需要這樣做，但BL仍需要將外部路由分發到其他枝葉節點。
- **故障**：請務必在配置完成期間和之後檢查故障。

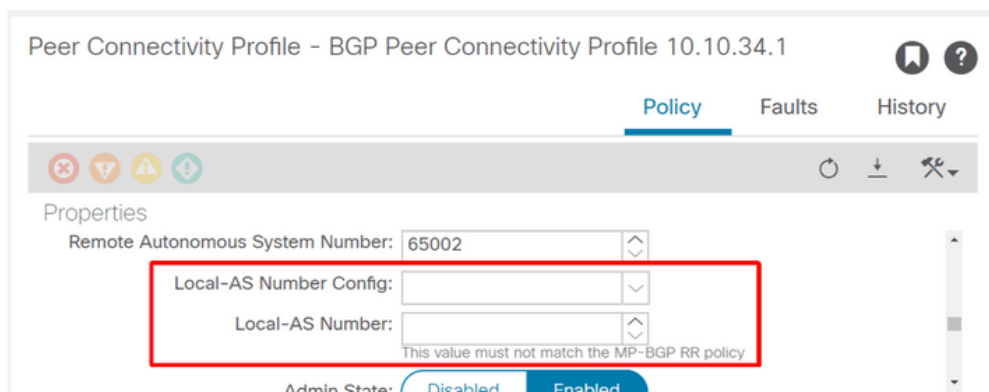
## BGP

本節使用「概述」部分中拓撲中BL3、BL4和R34上的環回之間的eBGP對等示例。R34上的BGP AS已啟65002。

建立BGP鄰接關係時，請驗證以下條件。

- 本地BGP AS編號（ACI BL端）。

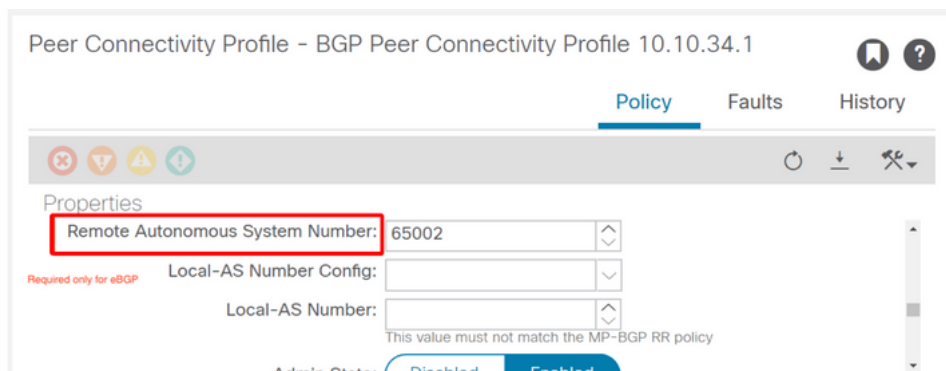
### 對等連線配置檔案 — Local-AS



使用者L3Out的BGP AS編號將自動與BGP路由反射器策略中配置的infra-MP-BGP的BGP AS相同。除非需要將ACI BGP AS偽裝到外部世界，否則不需要在BGP對等連線配置檔案中配置「本地AS」。這表示外部路由器應指向在BGP路由反射器中設定的BGP AS。

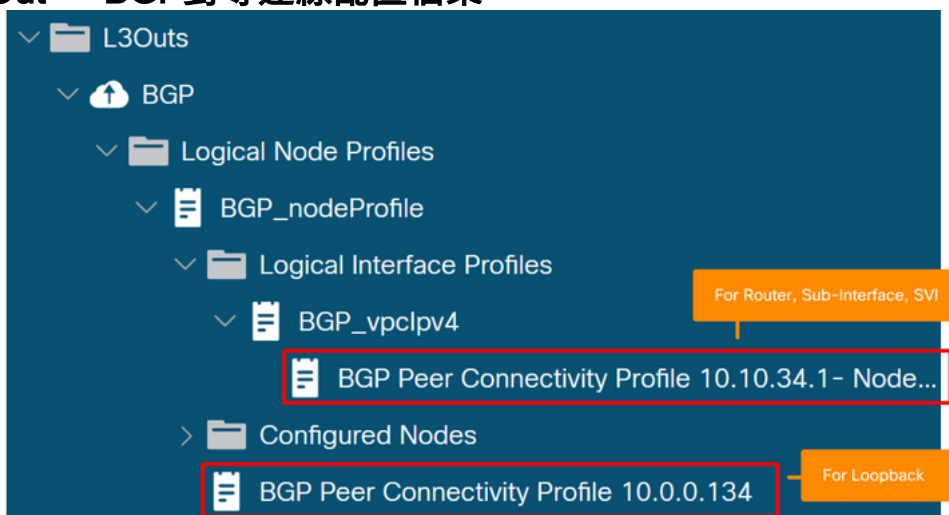
注意 — 需要本地AS配置的場景與獨立NX-OS「local-as」命令相同。

- 遠端BGP AS編號（外部端） **對等連線配置檔案 — 遠端AS**



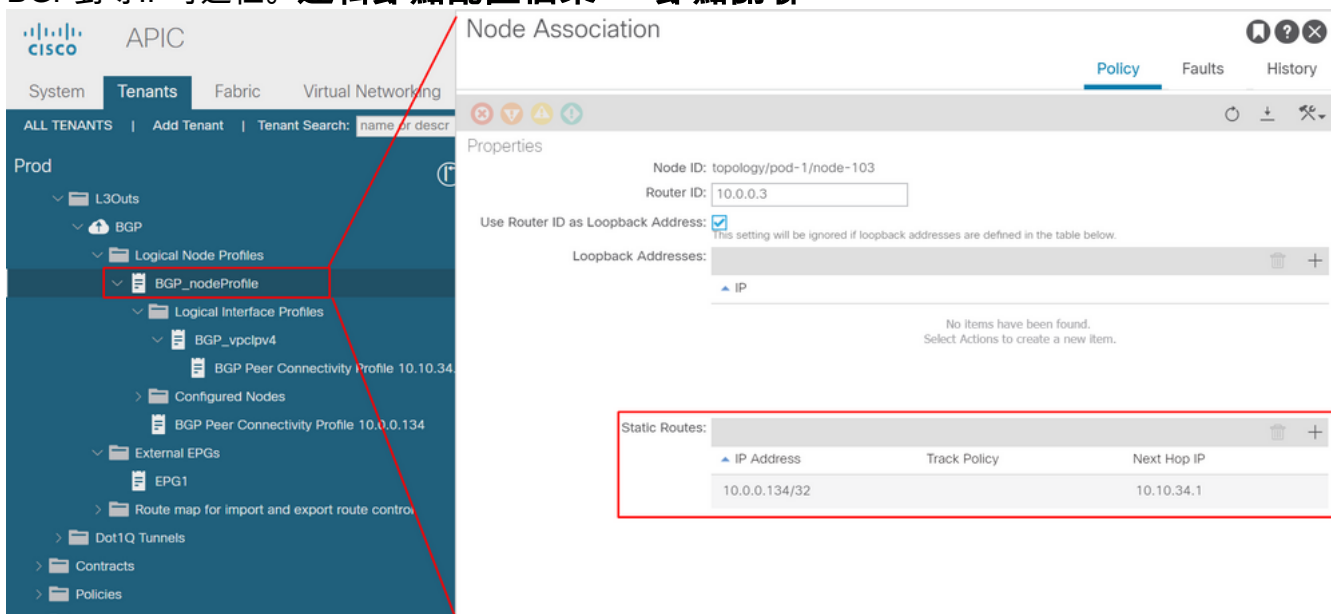
只有在鄰居的BGP AS不同於ACI BGP AS的eBGP中才需要遠端BGP AS編號。BGP對等會話

## 的源IP。L3Out — BGP對等連線配置檔案



ACI支援從典型ACI L3Out介面型別 ( 路由、子介面、SVI ) 頂部的環回介面獲取BGP會話。當BGP會話需要來自環回時，在Logical Node Profile下配置BGP對等連線配置檔案。當BGP會話需要來自路由/子介面/SVI時，在Logical Interface Profile下配置BGP對等連線配置檔案。

## BGP對等IP可達性。邏輯節點配置檔案 — 節點關聯



當BGP對等IP是環回時，請確保BL和外部路由器可以到達對等體的IP地址。靜態路由或OSPF可用於獲得對等IP的可達性。BGP CLI驗證 ( 使用回送的eBGP範例 ) 以下步驟的CLI輸出是從拓撲結構中的BL3的Overview部分收集的。1.檢查BGP會話是否已建立以下CLI輸出中的「State/PfxRcd」表示已建立BGP會話。

```
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」顯示空閒或活動，則尚未與對等裝置交換BGP資料包。在這種情況下，請檢查以下各項並進入下一步。



- 確保外部路由器正確指向ACI BGP AS(本地AS編號65001)。
- 確保ACI BGP對等體連線配置檔案指定了外部路由器從中獲取BGP會話(10.0.0.134)的正確鄰居IP。
- 確保ACI BGP對等體連線配置檔案指定外部路由器的正確鄰居AS(GUI中的遠端自治系統編號，在CLI中顯示為AS 65002)。

## 2.檢查BGP鄰居詳細資訊 ( BGP對等體連線配置檔案 )

以下命令顯示BGP鄰居建立的關鍵引數。

- 鄰居IP:10.0.0.134 .
- 鄰居BGP AS:遠端AS 65002。
- 來源 IP:使用loopback3作為更新源。
- eBGP多重躍點：外部BGP對等體可能最多相距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對等體正確建立，「本地主機」和「外部主機」將顯示在輸出的底部。

## 3.檢查BGP對等體的IP可達性

```
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
```

確保從ACI BGP的源IP對鄰居IP執行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.在外部路由器上檢查相同內容

以下是在外部路由器 ( 獨立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.額外步驟 — tcpdump

在ACI枝葉節點上，tcpdump工具可以嗅探「kpm\_inb」CPU介面，以確認協定資料包是否到達枝葉CPU。使用L4連線埠179(BGP)作為過濾器。

```
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 AreaID 1(NSSA)的「概述」部分中的拓撲中BL3、BL4和R34之間的OSPF鄰居關係的示例。

以下是檢查OSPF鄰接關係的常見標準。

- OSPF區域ID和型別

## L3Out — OSPF介面配置檔案 — 區域ID和型別



與任何路由裝置一樣，OSPF區域ID和型別需要在兩個鄰居上匹配。OSPF區域ID配置的一些特定於ACI的限制包括：

- 一個L3Out只能有一個OSPF區域ID。
- 只有兩個L3Outs位於兩個不同的枝葉節點上時，它們才能在同一VRF中使用相同的OSPF區域ID。

雖然OSPF ID不需要是主幹0，但在傳輸路由的情況下，同一枝葉上的兩個OSPF L3Outs之間需要該主幹0；其中一個必須使用OSPF區域0，因為OSPF區域之間的任何路由交換必須通過OSPF區域0。

- MTU

## 邏輯介面配置檔案 — SVI

Logical Interface Profile - OSPF\_vpclpv4

Policy | Faults | History

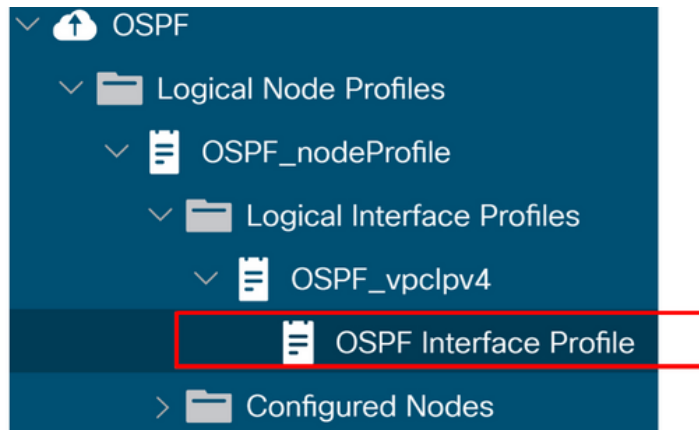
General | Routed Sub-Interfaces | Routed Interfaces | **SVI** | Floating SVI

Path	Side A IP	Side B IP	Secondary IP Address	IP Address	MAC Address	MTU (bytes)	Encap	Encap Scope
Pod-1/Node-103-104/N9K_VPC_3-4_13	10.10.34.3/29	10.10.34.4/29	10.10.34.2/29	0.0.0.0	00:22:BD:F8:19:FF	9000	vlan-2502	Local

ACI上的預設MTU為9000位元組，而不是1500位元組，後者通常用於傳統路由裝置。確保MTU與外部裝置匹配。當OSPF鄰居建立由於MTU而失敗時，它將停滯在EXCHANGE/DROTHER。

- IP子網掩碼。OSPF要求鄰居IP使用相同的子網掩碼。
- OSPF介面配置檔案。

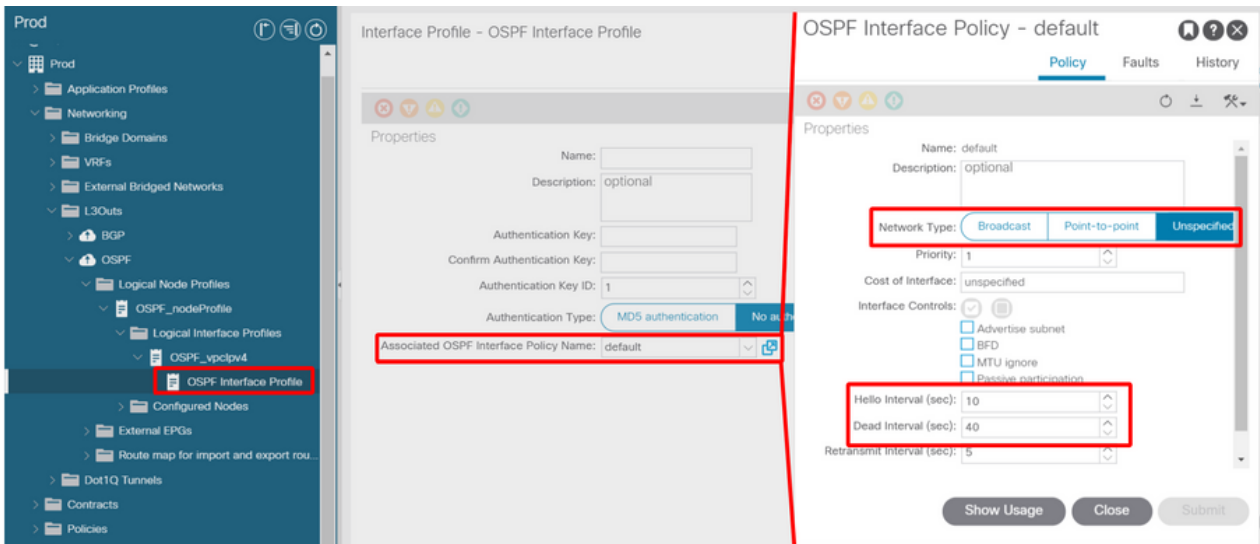
## OSPF介面配置檔案



這相當於在獨立NX-OS配置上啟用OSPF的「ip router ospf <tag> area <area id>」。否則，枝葉介面將不會加入OSPF。

- OSPF Hello/Dead計時器，網路型別

## OSPF介面配置檔案 — Hello/Dead計時器和網路型別



## OSPF介面策略詳細資訊

# Create OSPF Interface Policy



Name: OSPFIntPolicy

Description: optional

Network Type:  Broadcast  Point-to-point  Unspecified

Priority: 1

Cost of Interface: unspecified

Interface Controls:

- Advertise subnet
- BFD
- MTU ignore
- Passive participation

Hello Interval (sec): 10

Dead Interval (sec): 40

Retransmit Interval (sec): 5

Transmit Delay (sec): 1

OSPF要求每個鄰居裝置上的Hello計時器和失效計時器匹配。這些在OSPF介面配置檔案中配置。

確保OSPF介面網路型別與外部裝置匹配。當外部裝置使用點對點型別時，ACI端也需要顯式配置點對點。這些也在OSPF介面配置檔案中配置。

## OSPF CLI驗證

以下步驟中的CLI輸出是從拓撲中的BL3的「概述」部分收集的。

### 1.檢查OSPF鄰居狀態

如果以下CLI中的「State」為「FULL」，則會正確建立OSPF鄰居。否則，請繼續執行下一步以檢查引數。

```
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，當對SVI使用相同的VLAN ID時，BL將在外部路由器之上彼此形成OSPF鄰居。這是因為ACI具有稱為L3Out BD（或外部BD）的內部泛洪域，用於L3Out SVI中的每個VLAN ID。請注意，VLAN ID 28是一個稱為PI-VLAN（平台獨立VLAN）的內部VLAN，而不是線路上使用的實際VLAN（接入封裝VLAN）。使用以下命令驗證存取封裝VLAN('vlan-2502')。

```
f2-leaf3# show vlan id 28 extended
```

VLAN Name	Encap	Ports
28	vxlan-14942176, vlan-2502	Eth1/13, Po1

通過訪問封裝VLAN ID也可以獲得相同的輸出。

```
f2-leaf3# show vlan encap-id 2502 extended
```

VLAN Name	Encap	Ports
28	vxlan-14942176, vlan-2502	Eth1/13, Po1

## 2.檢查OSPF區域

確保OSPF區域ID和型別與鄰居相同。如果OSPF介面配置檔案丟失，該介面將不會加入OSPF，並且不會顯示在OSPF CLI輸出中。

```
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.檢查OSPF介面詳細資訊

確保介面級別引數滿足OSPF鄰居建立要求，例如IP子網、網路型別、Hello/Dead計時器。請注意VLAN ID以指定SVI為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.檢查與鄰居的IP連通性

雖然OSPF Hello資料包是鏈路本地組播資料包，但第一個OSPF LSDB交換所需的OSPF DBD資料包是單播。因此，對於OSPF鄰居建立，還需要驗證單播可達性。

```
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.在外部路由器上檢查相同的配置

以下是外部路由器（獨立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
```

請確保在物理介面上驗證MTU。

#### 6.額外步驟 — tcpdump

在ACI枝葉節點上，使用者可以在「kpm\_inb」CPU介面上執行tcpdump，以驗證協定資料包是否到達枝葉CPU。儘管OSPF有多個過濾器，但IP協定號是最全面的過濾器。

- IP協定號：proto 89(IPv4)或ip6 proto 0x59(IPv6)
- 鄰居的IP地址：主機<ip>
- OSPF連結本地模組IP:主機224.0.0.5或主機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

```

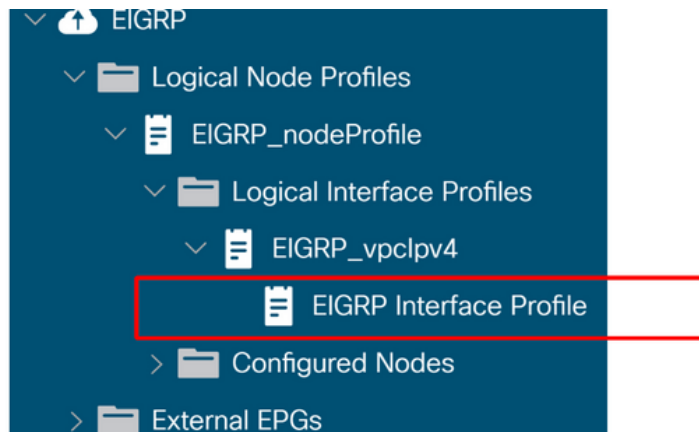
## EIGRP

本部分使用EIGRP AS 10中的「概述」部分拓撲中的BL3、BL4和R34之間的EIGRP鄰居關係示例。

以下是建立EIGRP鄰接關係的常見標準。

- EIGRP AS:為L3Out分配了一個EIGRP AS。這必須與外部裝置匹配。
- EIGRP介面配置檔案。

## EIGRP介面配置檔案



這等效於獨立NX-OS裝置上的「ip router eigrp <as>」配置。否則，枝葉介面將不會加入EIGRP。

- MTU

雖然這不必為簡單建立EIGRP鄰居關係而匹配，但EIGRP拓撲交換資料包可能會大於對等體之間的介面上允許的最大MTU，而且由於這些資料包不允許分段，因此它們將被丟棄，因此EIGRP鄰居關係將會翻動。

## EIGRP CLI驗證

以下步驟中的CLI輸出是從拓撲中的BL3的「概述」部分收集的。

### 1.檢查EIGRP鄰居狀態

```

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，當外部路由器使用SVI的相同VLAN ID時，BL將在外部路由器之上彼此形成EIGRP鄰居關



係。這是因為ACI具有稱為L3Out BD ( 或外部BD ) 的內部泛洪域，用於L3Out SVI中的每個VLAN ID。

請注意，VLAN ID 29是稱為PI-VLAN ( 平台獨立VLAN ) 的內部VLAN，而不是有線上使用的實際VLAN ( 接入封裝VLAN )。使用以下命令驗證存取封裝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
```

通過訪問封裝VLAN ID也可以獲得相同的輸出。

```
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.檢查EIGRP介面詳細資訊

確保EIGRP在預期介面上運行。如果不是，請檢查Logical Interface Profile和EIGRP Interface Profile。

```
f2-leaf3# show ip eigrp interfaces vrf Prod:VRF3
EIGRP interfaces for process 10 VRF Prod:VRF3
Interface      Peers  Xmit Queue  Mean   Pacing Time  Multicast  Pending
              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.在外部路由器上檢查相同的配置

以下為外部路由器 ( 獨立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.額外步驟 — tcpdump

在ACI枝葉節點上，使用者可以對「kpm\_inb」CPU介面執行tcpdump，以確認協定資料包是否到達枝葉的CPU。使用IP協定號88(EIGRP)作為過濾器。

```
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

本節重點介紹ACI中路由通告的驗證和故障排除。具體來說，它檢視的示例包括：

- 網橋域子網通告。
- 運輸路線通告。
- 匯入和匯出路由控制。

本節將討論路由洩漏，因為路由洩漏與後面各節中的共用L3Outs有關。

### 網橋域路由通告 workflow

在檢視常規故障排除之前，使用者應熟悉橋域通告的運作原理。

當BD和L3Out位於同一個VRF中時，BD通告涉及：

- 在L3Out和內部EPG之間具有合約關係。
- 將L3Out關聯到橋接域。
- 選擇BD子網上的「外部通告」。

此外，還可以使用匯出路由配置檔案控制橋接域通告，從而防止需要關聯L3Out。但是，仍應選擇「外部廣告」。這是一個不太常見的用例，因此將不在此討論。

L3Out和EPG之間需要合約關係，才能使BD普及靜態路由推送到BL。實際的路由通告是通過將靜態路由重分發到外部協定來處理的。最後，重分發路由對映將僅安裝在與BD關聯的L3Outs中。這樣，路由不會通告出所有L3Outs。

在這種情況下，BD子網為192.168.1.0/24，應該通過OSPF L3Out通告該子網。

### 在應用L3Out和內部EPG之間的合約之前

```
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
```

請注意，BD路由尚未出現在BL上。

## 在應用L3Out和內部EPG之間的合約之後

此時尚未進行其他配置。L3Out尚未與BD關聯，並且未設定「外部廣告」標誌。

```
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
```

請注意，BD子網路路由（由沈浸式標誌指示）現在部署在BL上。但請注意，該路由已標籤。此標籤值是分配給BD路由的隱式值，之後才配置為「外部通告」。所有外部協定都拒絕重新分發此標籤。

。

## 在BD子網中選擇「外部通告」後

L3Out尚未與BD關聯。但是請注意，標籤已清除。

```
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
```

此時，路由仍然沒有對外通告，因為沒有與此字首匹配的路由對映和字首清單來重分發到外部協定。可以使用以下命令驗證這一點：

```
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路由被程式設計為靜態路由，因此請通過運行「show route-map <route-map name>」，然後在

路由對映中存在的任何字首清單上運行「show ip prefix-list <name>」來檢查靜態重分發路由對映。在下一步中執行此操作。

## 將L3Out關聯到BD之後

如前所述，此步驟導致字首清單與靜態到外部協定重分發路由對映中安裝的BD子網匹配。

```
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
```

驗證首碼清單：

```
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子網以重新分發到OSPF。

此時，從L3Out通告BD子網的配置和驗證工作流程已完成。在此之後，核查將是特定協定的。例如：

- 對於EIGRP，驗證該路由是否正在拓撲表中安裝為「show ip eigrp topology vrf <name>」
- 對於OSPF，驗證該路由是否作為外部LSA安裝在資料庫表中，並顯示「show ip ospf database vrf <name>」
- 若是BGP，請使用「show bgp ipv4 unicast vrf <name>」驗證路由是否在BGP RIB中

## BGP路由通告

對於BGP，隱式允許所有靜態路由進行重分發。與BD子網匹配的路由對映應用於BGP鄰居級別。

```
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
```

在上方範例中，10.0.0.134是在L3Out中設定的BGP鄰居。

## EIGRP路由通告

與OSPF一樣，路由對映用於控制靜態到EIGRP的重分發。這樣，只應重新分發與L3Out關聯並設定為「外部通告」的子網。可以使用以下命令驗證這一點：

```
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
```

最終工作BD配置如下所示。

## 網橋域L3配置

The screenshot displays the Cisco APIC interface for configuring a Bridge Domain (BD1). The left sidebar shows the navigation tree with 'Networking' > 'Bridge Domains' > 'BD1' selected. The main content area shows the 'Policy' tab for 'Bridge Domain - BD1'. Under the 'L3 Configurations' sub-tab, a table lists the subnets:

Gateway Address	Scope	Primary IP Address	Virtual IP	Subnet Control
192.168.1.1/24	Advertised Externally	False	False	

Below the table, the 'Associated L3 Outs' section shows 'L3 Out' with 'OSPF' selected. The interface also includes a 'Properties' section with 'EP Move Detection Mode' set to 'GARP based detection' and buttons for 'Show Usage', 'Reset', and 'Submit'.

## 網橋域路由通告故障排除場景

在這種情況下，典型的症狀通常是配置的BD子網不會從L3Out中通告。按照上一個工作流程瞭解哪個元件已損壞。

在配置級別過低之前先驗證以下內容：

- EPG和L3Out之間是否存在合約？
- L3Out是否與BD關聯？
- BD子網是否設定為向外部通告？
- 外部協定鄰接是否已開啟？

**可能的原因：BD未部署**

此案例適用於多種不同情況，例如：

- 內部EPG使用VMM與按需選項整合，並且尚未將VM終端連線到EPG的埠組。
- 已建立內部EPG，但尚未配置靜態路徑繫結，或配置靜態路徑的介面已關閉。

在這兩種情況下，都不會部署BD，因此，不會將BD靜態路由推送到BL。此處的解決方案是在連結到此BD的EPG中部署一些活動資源，以便部署子網。

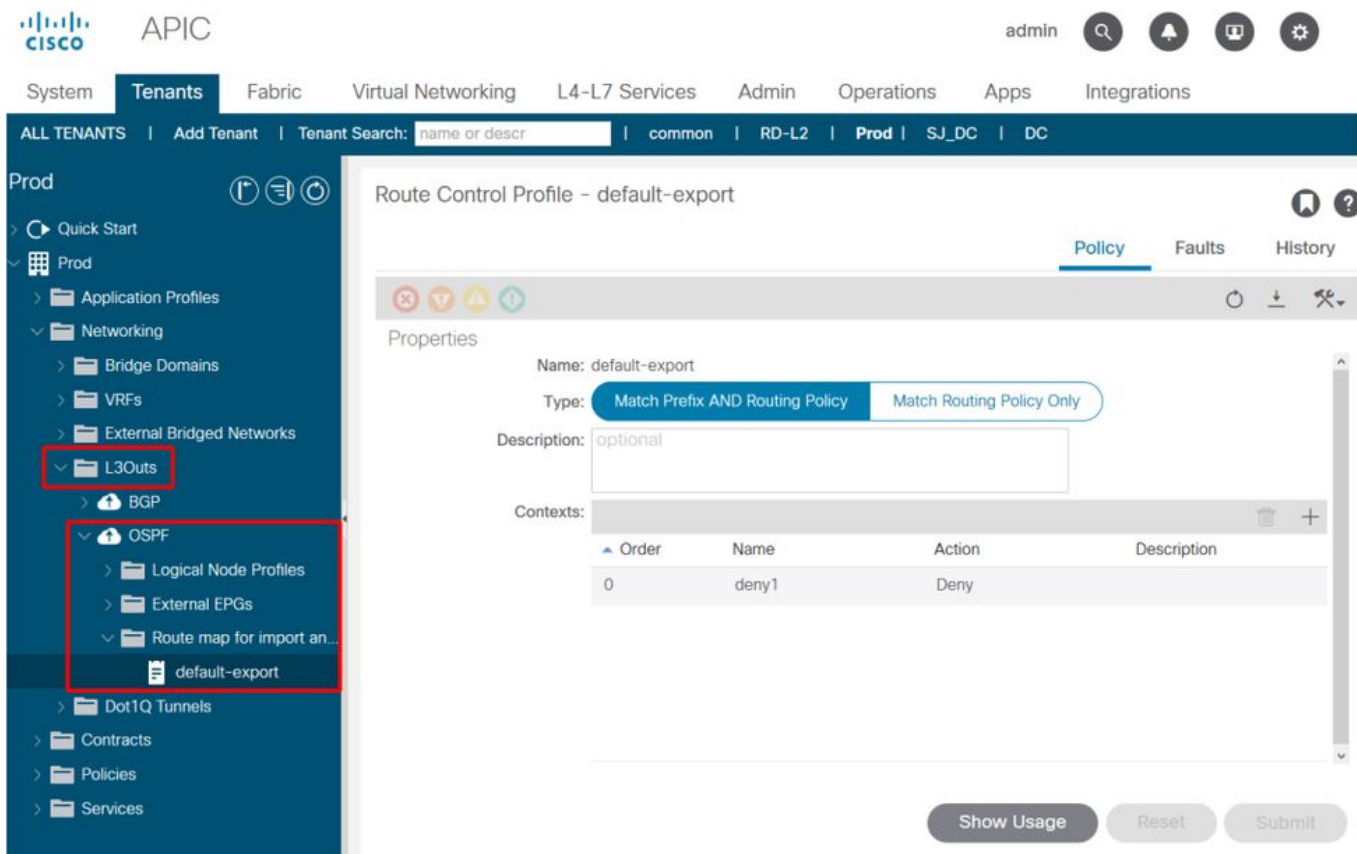
**可能的原因：OSPF L3Out配置為「Stub」或「NSSA」，且無重分發**

將OSPF用作L3Out協定時，仍然必須遵循基本OSPF規則。末節區域不允許重分佈的LSA，但可以通告預設路由。NSSA區域允許重分發路徑，但是必須在L3Out上選擇「將重分發的LSA傳送到NSSA區域」。或者NSSA也可以通過禁用「Originate Summary LSA」來通告預設路由，這也是禁用「Send Redistributed LSA's into NSSA Area」的典型情況。

**可能的原因：在L3Out下配置「Deny」操作的「Default-Export」路由配置檔案**

在L3Out下配置路由配置檔案時，如果使用的名稱是「default-export」或「default-import」，則路由配置檔案會隱式應用於L3Out。此外，如果將default-export route-profile設定為deny操作並配置為「Match Prefix and Routing Policy」，則應該從此L3Out中通告BD子網並隱式拒絕：

**預設匯出拒絕路由配置檔案**



如果選擇了「僅匹配路由策略」選項，則default-export route-profile中的字首匹配不會隱式包括BD子網。

## 外部路由匯入工作流程

本節討論ACI如何通過L3Out獲知外部路由並將其分配給內部枝葉節點。在後面的章節中，還包括傳輸和路由洩漏使用案例

與上一節一樣，使用者應瞭解更高級別的情況。

預設情況下，通過外部協定獲知的所有路由都重新分發到內部交換矩陣BGP進程中。無論在外部EPG下配置了哪些子網以及選擇了哪些標誌，此情況均成立。有兩個例子說明這不是真的。

- 如果頂級L3Out策略的「路由控制實施」選項設定為「匯入」。在這種情況下，路由匯入模型將從阻止清單模型（僅指定不應允許的內容）轉到允許清單模型（除非另有配置，否則所有內容都隱式拒絕）。
- 如果外部協定是EIGRP或OSPF，並且使用的Interleak Route-Profile與外部路由不匹配。

要將外部路由分發到內部枝葉，必須發生以下情況：

- 該路由必須在BL上從外部路由器獲知。要成為重新分發到交換矩陣MP-BGP進程的候選者，必須將路由安裝在路由表中，而不是僅安裝在協定RIB中。
- 必須允許將路由重新分發或通告到內部BGP進程中。除非使用匯入路由控制實施或Interleak路由配置檔案，否則應始終發生這種情況。
- 必須配置BGP路由反射器策略並將其應用於應用於Pod配置檔案的Pod策略組。如果未應用該設定，則不會在交換機上初始化BGP進程。

如果內部EPG/BD與L3Out位於同一VRF中，則內部EPG/BD使用外部路由只需上述三個步驟。

## 路由安裝在BL路由表中

在這種情況下，應該在BL 103和104上獲知的外部路由是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
```

很明顯，它通過OSPF獲知後會安裝到路由表中。如果在此處未看到該協定，請檢查單個協定並確保鄰接關係已啟動。將路由重新分發到BGP在檢查未使用「Import」實施或Interleak路由配置檔案後，可以通過檢視用於BGP重分發的外部協定的路由對映來驗證重分發路由對映。請參閱以下命令：

```
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
```

這裡很明顯，「permit-all」路由對映用於OSPF到BGP的重分發。這是預設設定。從這裡可以檢驗BL並檢查源自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
```

在上面的輸出中，0.0.0.0/0表示它源自本地。通告的對等體清單是交換矩陣中充當路由反射器的主幹節點。

## 驗證內部枝葉上的路由

BL應通過VPNv4 BGP地址系列將其通告給主幹節點。主幹節點應將其通告給任何已部署VRF的枝葉節點（非路由洩漏示例中為true）。在這些枝葉節點中的任何一個上，運行「show bgp vpnv4 unicast <route> vrf overlay-1」以驗證它是否在VPNv4中

使用以下命令驗證內部枝葉上的路由。

```
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
```

在上述輸出中，路由是通過BGP得知的，下一跳應是這些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
```

## 外部路由故障排除場景

在這種情況下，內部枝葉(101)沒有收到外部路由。

與往常一樣，首先檢查基本知識。請確保：

- BL上的路由協定鄰接關係已啟動。
- BGP路由反射器策略應用於Pod策略組和Pod配置檔案。

如果上述標準正確，下面是一些可能導致問題的更高級示例。

### 可能的原因：未在內部枝葉上部署VRF

在這種情況下，問題將是，在預期外部路由的內部枝葉上沒有部署資源的EPG。這可能是由於靜態路徑繫結僅在關閉介面上配置或僅存在請求模式VMM整合EPG且未檢測到動態附件。

由於L3Out VRF未部署在內部枝葉上（在內部枝葉上使用「show vrf」進行驗證），因此內部枝葉不會從VPNv4匯入BGP路由。

要解決此問題，使用者應在內部枝葉上的L3Out VRF中部署資源。

### 可能的原因：正在使用匯入路由實施

如前所述，當啟用匯入路由控制實施時，L3Out只接受明確允許的外部路由。通常，該功能以表對映的形式實現。表對映位於協定RIB和實際路由表之間，因此它只影響路由表中的內容。

在下面的輸出中，Import Route-Control已啟用，但沒有任何明確允許的路由。請注意，LSA位於OSPF資料庫中，但不位於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
```

以下是現在安裝的導致此行為的表對映：

```
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:
```

區域100 ( 在此L3Out上配置的區域 ) 中的任何學習內容都將被此表對映隱式拒絕，以便它不會安裝到路由表中。

要解決此問題，使用者應使用「Import Route Control Subnet」標誌在外部EPG上定義子網，或建立與要安裝的字首匹配的匯入路由配置檔案。

- 請注意，EIGRP不支援匯入實施。
- 另請注意，對於BGP，匯入實施將作為應用於BGP鄰居的入站路由對映實施。檢查「BGP路由通告」子部分，獲取有關如何檢查此通告的詳細資訊。

**可能的原因：正在使用Interleak配置檔案**

Interleak Route-Profiles用於EIGRP和OSPF L3Outs，用於控制從IGP重分發到BGP的內容，並允許應用設定BGP屬性等策略。

如果沒有interleak Route-Profile，所有路由都會隱式匯入到BGP。

沒有互漏路由配置檔案：

```
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interleak路由配置檔案：

```
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
```

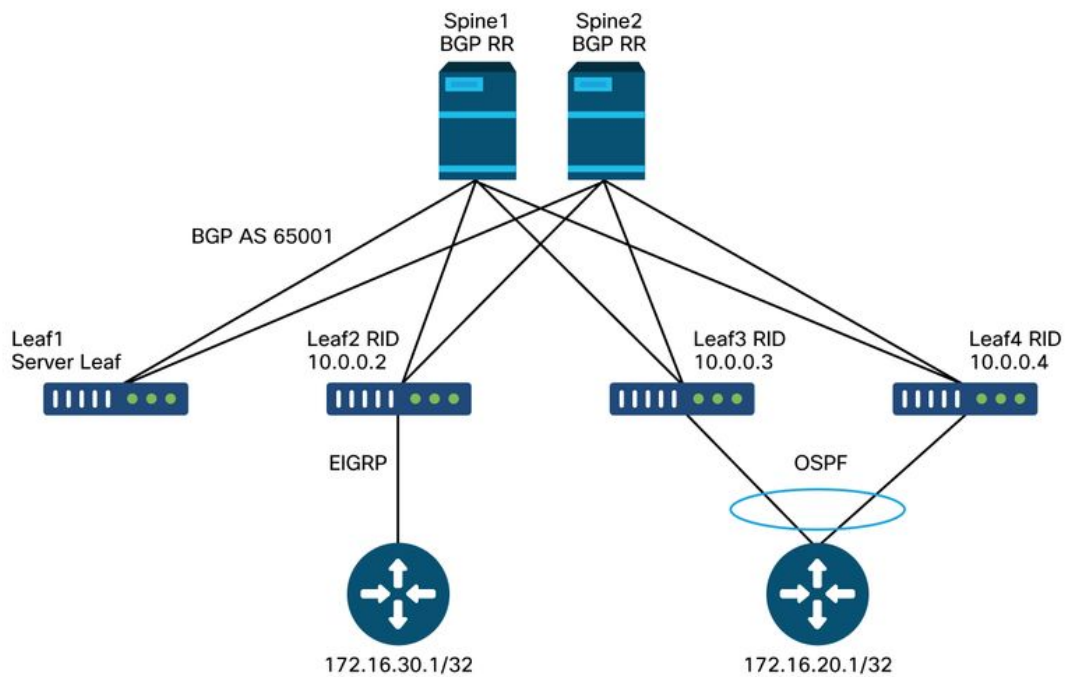
上面突出顯示的路由對映將僅允許已配置的Interleak配置檔案中明確匹配的內容。如果外部路由不匹配，將不會將其重新分發到BGP。

## 運輸路線廣告 workflow

本節將討論如何從一個L3Out路由通告出另一個L3Out。這也包括需要通告L3Out上直接配置的靜態路由的情況。它不會涉及每個具體的協定考慮事項，而是會涉及如何在ACI中實施該事項。此時不會進入VRF間傳輸路由。

此方案將使用以下拓撲：

## 傳輸路由拓撲



如何從OSPF獲知172.16.20.1並將其通告到EIGRP的高級流程，以及整個流程和故障排除方案的驗證將在下面討論。

要將172.16.20.1路由通告到EIGRP，必須配置以下其中一項：

- 可以在EIGRP L3Out上使用「Export Route-Control Subnet」標籤定義要通告的子網。如概述部分所述，此標籤主要用於傳輸路由，並定義應從該L3Out通告的子網。
- 配置0.0.0.0/0並選擇「聚合匯出」和「匯出路由控制子網」。這將建立一個路由對映，以便重新分發到與0.0.0.0/0及所有更特定字首（這是有效的匹配any）匹配的外部協定。請注意，將0.0.0.0/0與「Aggregate Export」一起使用時，將不會為重分發匹配靜態路由。這是為了防止無意中通告不應通告的BD路由。
- 最後，可以建立與要通告的字首匹配的匯出路由配置檔案。使用此方法可以配置「Aggregate」選項，該選項帶有除0.0.0.0/0之外的字首。

上述配置將導致傳輸路由被通告，但它仍然需要設定安全策略以允許資料平面流量通過。與任何EPG到EPG通訊一樣，在允許流量之前必須存在合約。

請注意，不能在同一VRF中配置具有「外部EPG的外部子網」的重複外部子網。配置時，子網需要比0.0.0.0更加具體。只為接收路由的L3Out配置「外部EPG的外部子網」非常重要。請勿在本應通告此路由的L3Out上配置此命令。

此外，還必須瞭解所有傳輸路由都使用特定的VRF標籤進行標籤。預設情況下，此標籤為4294967295。路由標籤策略配置在「租戶>網路>協定>路由標籤」：

## 路由標籤策略

The screenshot shows the Cisco APIC interface. The top navigation bar includes 'System', 'Tenants', 'Fabric', 'Virtual Networking', 'L4-L7 Services', 'Admin', 'Operations', 'Apps', and 'Integrations'. The 'Tenants' tab is active, and the 'Prod' tenant is selected. The left sidebar shows a tree view of configuration objects, with 'Route Tag' selected and 'nonDefaultName' highlighted. The main panel displays the configuration for 'Protocol - Route Tag' with a table:

Name	Tag	Description
nonDefaultName	11111	

The bottom of the panel shows pagination information: Page 1 of 1, Objects Per Page: 15, and Displaying Objects 1 - 1 Of 1.

然後，此路由標籤策略應用於VRF。此標籤的用途主要是防止回圈。當從L3Out通告回中轉路由時，將應用此路由標籤。如果收到這些路由時帶有相同的路由標籤，則丟棄該路由。

### 檢驗通過OSPF的接收BL上是否存在路由

與最後一節一樣，首先檢驗最初應接收正確路由的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
```

現在，假設廣告L3Out位於不同的BL上（如拓撲所示）（後面的場景將討論它在同一個BL上的位置）。

### 驗證接收OSPF BL上的BGP中是否存在路由

對於要通告到外部EIGRP路由器的OSPF路由，需要將該路由通告到接收OSPF BL上的BGP中。

```
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

```

路由在BGP中。

### 在EIGRP BL上檢驗應通告其已安裝的路由

```

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

```

它安裝在路由表中，帶有指向原始邊界枝葉節點的重疊下一跳。

```

leaf102# acidiag fmvread

      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

```

### 驗證是否在BL上通告路由

由於在配置的子網上設定了「匯出路由控制子網」標誌，BL 102將通告路由：

### 匯出路由控制

External EPG Instance Profile - instP

Policy Operational Stats Health Faults History

General Contracts Subject Labels EPG Labels

100

Properties

Configuration Status: applied

Configuration Issues:

Preferred Group Member: Exclude Include

Subnets:

IP Address	Scope	Name	Aggregate	Route Control Profile	Route Summarization Policy
0.0.0.0/0		External Subnets for the External EPG			
172.16.20.1/32		Export Route Control Subnet			

Show Usage Reset Submit

Current System Time: 2019-10-27 12:24 UTC-04:00

使用以下命令檢視由於此「匯出路由控制」標誌而建立的路由對映：

```
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
```

要查詢「BGP > EIGRP重分發」，請檢視路由對映。但是，無論源協定是OSPF、EIGRP還是BGP，路由對映本身都應相同。將使用不同的路由對映控制靜態路由。

```
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
```

在上面的輸出中，在此字首上設定VRF標籤用於環路預防，並且使用「匯出路由控制」配置的子網被顯式匹配。

## 接收和通告BL時的傳輸路由是相同的

如前所述，當接收和通告BL不同時，必須使用BGP通過交換矩陣通告路由。如果BL相同，則可以在枝葉上的協定之間直接執行重分發或通告。

以下簡要介紹如何實施此功能：

- 在同一枝葉上的兩個OSPF L3Outs之間傳輸路由：路由通告通過應用於OSPF進程級別的「area-filter」進行控制。區域0中的L3Out必須部署在枝葉上，因為路由是在區域之間通告的，而不是通過重分發。使用「show ip ospf vrf <name>」檢視過濾器清單。使用「show route-map <filter name>」顯示過濾器的內容。
- 在同一枝葉上的OSPF和EIGRP L3Outs之間傳輸路由：路由通告通過重分發路由對映控制，通過「show ip ospf」和「show ip eigrp」可以看到。請注意，如果同一BL上存在多個OSPF L3Outs，則僅將另一個OSPF L3Outs重分佈到其中一個OSPF L3Outs的唯一方法是，另一個是已禁用「將重分佈的LSA傳送到NSSA區域」的Stub或NSSA，以便它不允許任何外部LSA。
- 在同一枝葉上的OSPF或EIGRP與BGP之間傳輸路由：通過重分佈路由對映控制到IGP的路由通告。通過直接應用於應傳送路由的bgp鄰居的出站路由對映，控制到BGP的路由通告。可以使用「show bgp ipv4 unicast neighbor <neighbor address> vrf <name>」驗證這一點 | grep Outbound'。
- 在同一枝葉上的兩個BGP I3Outs之間傳輸路由：所有通告通過直接應用於應向其傳送路由的bgp鄰居的路由對映來控制。可以使用「show bgp ipv4 unicast neighbor <neighbor address> vrf <name>」驗證這一點 | grep Outbound'。

## 傳輸路由故障排除#1案：未通告傳輸路由

此疑難排解情況涉及應該通過一個L3Out獲知的路由，而不是從另一個L3Out發出。

與往常一樣，在檢視ACI特定內容之前先檢查基本知識。

- 協定鄰接關係是否已啟動？
- ACI應通告的路由是否首先從外部協定獲知？
- 對於BGP，路徑是否由於某個BGP屬性而被丟棄？（as-path等）。
- 接收的L3Out是否包含在OSPF資料庫、EIGRP拓撲表或BGP表中？
- BGP路由反射器策略是否應用於應用於Pod配置檔案的Pod策略組？

如果所有基本協定驗證都配置正確，下面是未通告的傳輸路由的一些其他常見原因。

## 可能的原因：無OSPF區域0

如果受影響的拓撲涉及同一邊界枝葉上的兩個OSP L3Outs，則必須有一個區域0用於從一個區域通告到另一個區域的路由。有關更多詳細資訊，請檢視上面的「同一枝葉上兩個OSPF L3Outs之間的傳輸路由」專案符號。

## 可能的原因：OSPF區域是末節或NSSA

如果OSPF L3Out配置了末節或NSSA區域，但未配置為通告外部LSA，則會出現這種情況。使用OSPF時，外部LSA永遠不會通告到末節區域。如果選擇「將重分發的LSA傳送到NSSA區域」，則會將這些埠通告到NSSA區域。

## 傳輸路由故障排除#2案：未收到傳輸路由

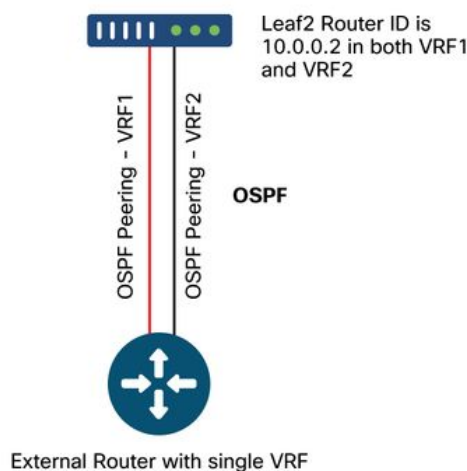
在此案例中，問題在於ACI L3Out通告的某些路由沒有在另一個L3Out中接收回來。如果L3Outs位於兩個獨立的結構中，並且由外部路由器連線，或者L3Outs位於不同的VRF中，並且路由正通過外部路由器在VRF之間傳遞，則此方案可能適用。

## 可能的原因：多個VRF中的BL配置了相同的路由器ID

從配置角度來看，路由器ID不能在同一VRF中重複。但是，只要兩個VRF沒有連線到相同的路由協定域，通常可以在不同的VRF中使用相同的路由器ID。

請考慮以下拓撲：

## 具有單個VRF的外部路由器 — 未收到傳輸路由



這裡的問題在於，ACI枝葉會看到收到具有其自己的Router-ID的LSA，從而導致這些未安裝在OSPF資料庫中。

此外，如果VPC對出現相同的設定，則會在某些路由器上持續新增和刪除LSA。例如，路由器會看到來自其VPC對等體的LSA和VRF以及來自其他VRF中發起的同一節點（具有相同路由器ID）的LSA。

要解決此問題，使用者應確保節點在具有L3Out的每個VRF內具有不同的唯一路由器ID。

## 可能的原因：使用相同VRF標籤從一個ACI交換矩陣中接收的一個L3Out路由

除非更改，否則ACI中的預設路由標籤始終相同。如果在不更改預設VRF標籤的情況下，從一個VRF或ACI交換矩陣中的一個L3Out向另一個VRF或ACI交換矩陣中的另一個L3Out通告路由，則接收BL將丟棄路由。

此方案的解決方案只是為ACI中的每個VRF使用唯一的路由標籤策略。

## 傳輸路由疑難解答方#37 — 意外通告的傳輸路由

當傳輸路由被通告出並不打算通告它們的L3Out時，將會出現此情況。

### 可能的原因：0.0.0.0/0與「聚合匯出」的用法

當外部子網配置為0.0.0.0/0並帶有「Export Route Control Subnet」和「Aggregate Export」時，結果是安裝了匹配的所有重分發路由對映。在這種情況下，通過OSPF、EIGRP或BGP獲取的路由上的所有路由都會從配置該路由的L3Out中通告。

以下是因彙總匯出而部署到枝葉的路由對映：

```
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
```

這是涉及ACI環境的路由環路的第一大原因。

## 合約和L3Out

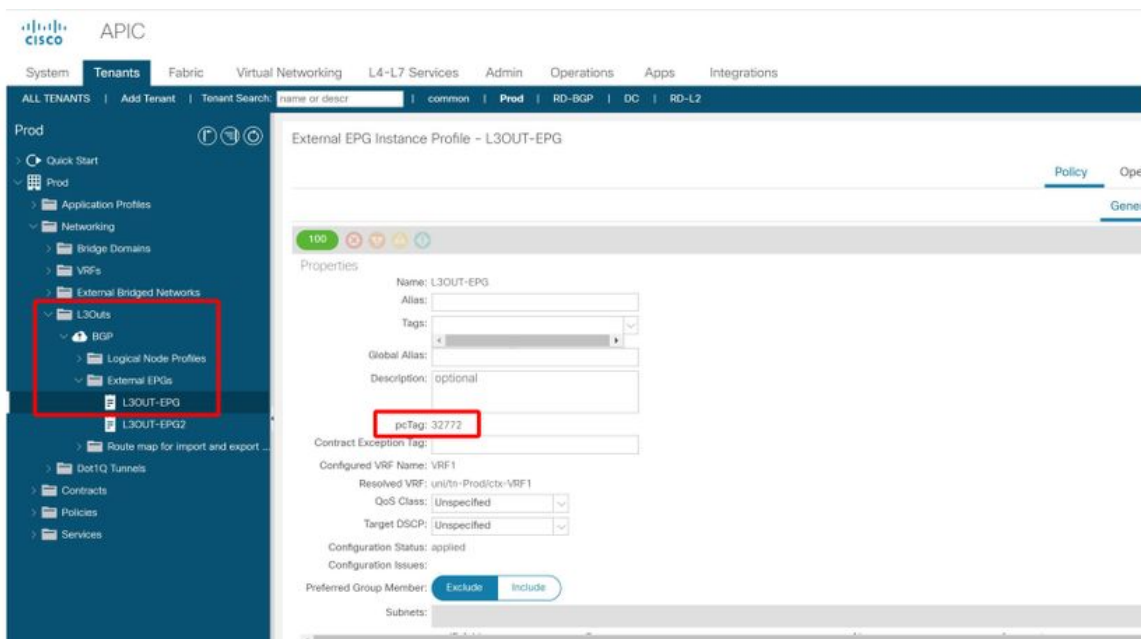
### L3Out上基於字首的EPG

在內部EPG (非L3Out) 中，在派生源的pcTag和目標EPG的pcTag後實施合約。在下行鏈路埠上收到的資料包的封裝VLAN/VXLAN通過將資料包分類到EPG來驅動此pcTag。在學習MAC地址或IP地址時，會學習該地址及其訪問封裝和關聯的EPG pcTag。有關pcTag和合約實施的更多詳細資訊，請參閱「安全策略」一章。

L3Out還使用位於「Tenant > Networking > L3OUT > Networks > L3OUT-EPG」下的L3Out EPG (外部EPG) 驅動pcTag。但是，L3Outs不依賴VLAN和介面對資料包進行此類分類。而是基於「最長字首匹配」方式的源字首/子網。因此，L3Out EPG可以稱為基於字首的EPG。根據子網將資料包分類到L3Out後，它遵循與常規EPG類似的策略實施模式。

下圖概述了可在GUI中找到的給定L3Out EPG的pcTag。

### L3Out的pcTag的位置



使用者負責定義基於字首的EPG表。這是使用「外部EPG的外部子網」子網範圍完成的。使用該範圍的每個子網集都將在靜態最長字首匹配(LPM)表中新增一個條目。此子網將指向用於該字首內的任何IP地址的pcTag值。

可使用以下命令在枝葉交換機上驗證基於字首的EPG子網的LPM表：

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

備註：

- LPM表條目範圍設為VRF VNID。依照vrf\_vnid/src pcTag/dst pcTag完成查詢。
- 每個條目都指向一個pcTag。因此，兩個L3Out EPG不能在同一VRF中使用具有相同掩碼長度的同一子網。

- 子網0.0.0.0/0始終使用特殊的pcTag 15。因此，可以複製該子網，但只有在完全瞭解策略實施影響的情況下才能複製。
- 此表在兩個方向上都使用。從L3Out到Leaf Local Endpoint，源pcTag是使用此表派生的。從枝葉本地端點到L3Out，目標pcTag是使用此表派生的。
- 如果VRF具有「策略控制實施方向」的「輸入」實施設定，則LPM字首表將顯示在L3Out BL以及VRF中具有與L3Out合約的所有枝葉交換機上。

## 範例 1：具有特定首碼的單個L3Out

**案例:**vrf Prod:VRF1中的單個BGP L3Out和一個L3Out EPG。字首172.16.1.0/24是從外部源接收的，因此必須將其分類到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，將子網新增到字首表中。

## 具有「外部EPG的外部子網」範圍的子網

## Create Subnet

IP Address:   
address/mask

Name:

scope:  Export Route Control Subnet  
 Import Route Control Subnet  
 External Subnets for the External EPG  
 Shared Route Control Subnet  
 Shared Security Import Subnet

BGP Route Summarization Policy:

aggregate:  Aggregate Export  
 Aggregate Import  
 Aggregate Shared Routes

Route Control Profile:

Name	Direction

驗證具有L3Out VRF的枝葉交換機上的字首清單的程式設計：

```
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位於vrf32772圍範圍2097154。

## 範例 2：具有多個字首的單個L3Out

在上一個範例中展開後，在此情境中，L3Out接收多個字首。當輸入每個字首在功能上合理時，另一個選項（取決於預期設計）是接受L3Out上接收的所有字首。

可以使用「0.0.0.0/0」字首來完成此操作。

# 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
  - 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:

Route Control Profile:

Name ▲ Direction

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

這將產生以下policy-mgr字首表條目：

```
bdso1-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
```

請注意，分配給0.0.0.0/0的pcTag使用值15，而非32772。pcTag 15是保留的系統pcTag，僅與0.0.0.0/0一起使用，後者充當萬用字元以匹配L3Out上的所有字首。

如果VRF具有使用0.0.0.0/0的單一L3Out和單一L3Out EPG，則策略字首保持唯一，並且是捕獲所有內容的最簡單方法。

### 範例3a:VRF中的多個L3Out EPG

在此案例中，同一VRF中有多個L3Out EPG。

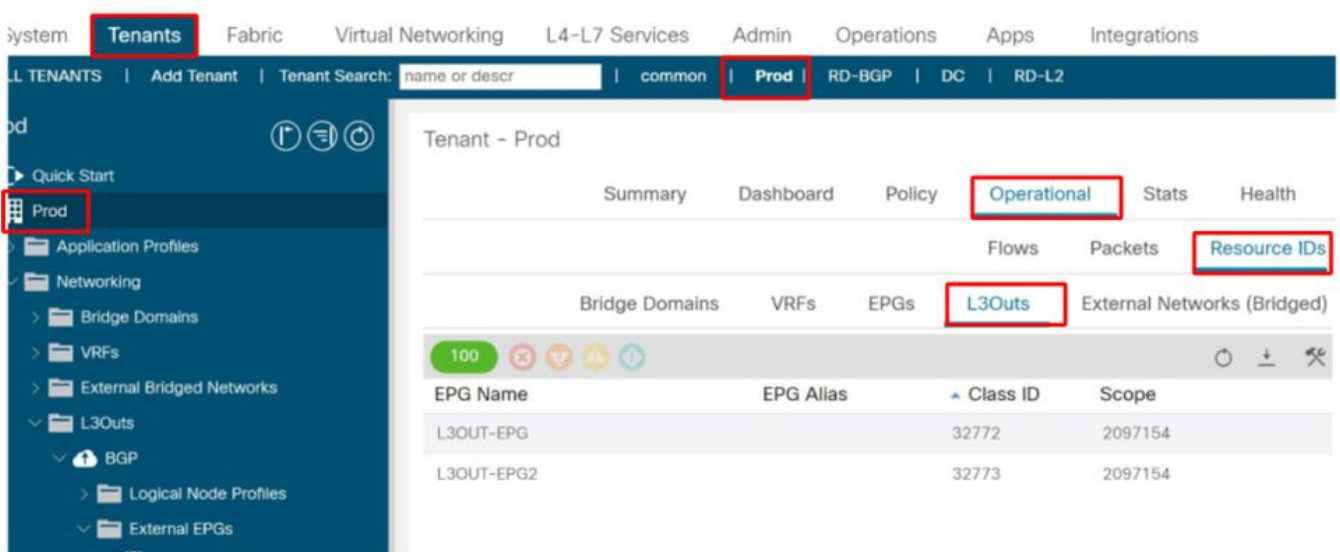
附註：從基於字首的EPG角度來看，以下兩個配置將產生等效的LPM policy-mgr字首表條目：

1. 兩個L3Out，各一個L3Out EPG。
2. 一個L3Out和兩個L3Out EPG

在這兩種情況下，L3Out EPG的總數都是2。這意味著每個子網都有自己的pcTag和相關子網。

給定的L3Out EPG的所有pcTags都可以在GUI的「Tenant > Operational > Resource id > L3Outs」中檢視

### 驗證L3Out pcTag



在此方案中，ACI交換矩陣從外部路由器接收多個字首，L3Out EPG定義如下：



- 172.16.1.0/24已分配給L3OUT-EPG。
- 172.16.2.0/24已分配給L3OUT-EPG2。
- 172.16.0.0/16已分配給L3OUT-EPG(以捕獲172.16.3.0/24字首)。

要與此匹配，配置將定義如下：

- L3OUT-EPG的子網172.16.1.0/24和172.16.0.0/16的範圍均為「外部EPG的外部子網」。
- L3OUT-EPG2的子網172.16.2.0/24的範圍為「外部EPG的外部子網」。

生成的字首表條目將為：

```
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分配給pcTag 32773(L3OUT-EPG2),172.16.0.0/16分配給32772(L3OUT-EPG)。

在此方案中，172.16.1.0/24的條目是冗餘的，因為/16超網被分配到同一個EPG。

當目標是將不同的合約應用於單個L3Out中的字首組時，多個L3Out EPG非常有用。下一個示例將說明合約如何與多個L3Out EPG配合使用。

### 範例3b:具有不同合約的多個L3Out EPG

此方案包含以下設定：

- ICMP合約僅允許ICMP。
- HTTP合約僅允許tcp目標埠80。
- EPG1(pcTag 32770)提供L3OUT-EPG(pcTag 32772)使用的HTTP合約。
- EPG2(pcTag 32771)提供L3OUT-EPG2(pcTag 32773)使用的ICMP合約。

將使用以上示例中的相同policymgr字首：

- L3OUT-EPG中的172.16.1.0/24應允許HTTP到EPG1
- L3OUT-EPG2中的172.16.2.0/24應允許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
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驗證資料路徑 — 策略允許的流

對於外部網路上的172.16.2.1和EPG2中的192.168.3.1之間的ICMP流，可以使用fTriage捕獲和分析該流。在這種情況下，在枝葉交換機103和104上啟動fTriage，因為流量可能進入其中任一交換機：

```

admin@apic1:~> ftrriage route -ii LEAF:103,104 -sip 172.16.2.1 -dip 192.168.3.1
fTriage Status: {"dbgFtrriage": {"attributes": {"operState": "InProgress", "pid": "14454",
"apicId": "1", "id": "0"}}}
Starting ftrriage
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/ftrriage route -ii LEAF:103,104 -sip 172.16.2.1
-dip 192.168.3.1
2019-10-02 22:31:28,868 INFO ftrriage: main:1165 Invoking ftrriage with default password
and default username: apic#fallback\admin
2019-10-02 22:32:15,076 INFO ftrriage: 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 ftrriage: main:242 ingress encap string vlan-2551
2019-10-02 22:32:17,839 INFO ftrriage: main:271 Building ingress BD(s), Ctx
2019-10-02 22:32:20,583 INFO ftrriage: main:294 Ingress BD(s) Prod:VRF1:l3out-BGP:vlan-
2551
2019-10-02 22:32:20,584 INFO ftrriage: main:301 Ingress Ctx: Prod:VRF1
2019-10-02 22:32:20,693 INFO ftrriage: pktrec:490 bdsol-aci32-leaf3: Collecting transient
losses snapshot for LC module: 1
2019-10-02 22:32:38,933 INFO ftrriage: nxos:1404 bdsol-aci32-leaf3: nxos matching rule
id:4343 scope:34 filter:5

```

```

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

```

**FTriage確認從L3OUT\_EPG2到EPG的ICMP規則所命中的分割槽規則：**

```

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

```

### 使用fTriage的資料路徑驗證 — 策略不允許的流

對於來源為172.16.1.1(L3OUT-EPG)且目的地為192.168.3.1(EPG2)的ICMP流量，預期原則捨棄。

```

admin@apic1:~> ftrriage route -ii LEAF:103,104 -sip 172.16.1.1 -dip 192.168.3.1
fTriage Status: {"dbgFtrriage": {"attributes": {"operState": "InProgress", "pid": "15139",
"apicId": "1", "id": "0"}}}
Starting ftrriage
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/ftrriage route -ii LEAF:103,104 -sip 172.16.1.1
-dip 192.168.3.1
2019-10-02 22:40:03,523 INFO      ftrriage:      main:1165     Invoking ftrriage with default password
and default username: apic#fallback\admin
2019-10-02 22:40:43,338 ERROR      ftrriage:      unicast:234   bdsol-aci32-leaf3: L3 packet getting fwd
dropped, checking drop reason
2019-10-02 22:40:43,339 ERROR      ftrriage:      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      ftrriage:      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      ftrriage:      main:681      : Ftrriage Completed with hunch: None
fTriage Status: {"dbgFtrriage": {"attributes": {"operState": "Idle", "pid": "0", "apicId": "0",
"id": "0"}}}

```

分類確認資料包已因SECURITY\_GROUP\_DENY (策略丟棄) 原因被丟棄，並且派生的源pcTag為32772，目標pcTag為32771。根據分割槽規則檢查此項時，這些EPG之間顯然沒有條目。

```

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 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

## 範例 4:有多個帶有多個字首的L3Outs

此案例設定與範例3 ( L3Out和L3Out EPG定義 ) 類似，但兩個L3Out EPG上定義的網路為0.0.0.0/0。

合約配置如下：

- ICMP1合約允許ICMP。
- ICMP2合約允許ICMP。
- EPG1(pcTag 32770)提供L3OUT-EPG(pcTag 32772)使用的ICMP1合約。
- EPG2(pcTag 32771)提供L3OUT-EPG2(pcTag 32773)使用的ICMP2合約。

在外部網路通告許多字首的情況下，此配置可能看起來比較理想，但至少要有兩個字首塊遵循不同的允許流模式。在本範例中，一個首碼應僅允許ICMP1，而另一個首碼應僅允許ICMP2。

儘管在同一VRF中使用了'0.0.0.0/0'兩次，但在policy-mgr字首表中只程式設計了一個字首：

```
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
```

下面重新審查了兩個流量。根據上面的合約配置，預期如下：

1. ICMP2應允許172.16.2.1(L3OUT-EPG2)到192.168.3.1(EPG2)
2. 不應允許172.16.2.1(L3OUT-EPG2)到192.168.1.1(EPG1)，因為EPG1和L3OUT-EPG2之間沒有合約

### 使用fTriage的資料路徑驗證 — 策略允許的流

使用ICMP流量從172.16.2.1(L3OUT-EPG2)到192.168.3.1(EPG2 — pcTag 32771)運行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

```

分割槽規則4336允許 ( 如預期的 ) 此流。

## 使用fTriage的資料路徑驗證 — 策略不允許的流

使用ICMP流從172.16.2.1(L3OUT-EPG2)到192.168.1.1(EPG1 — pcTag 32770)運行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

分割槽規則4341允許 ( 意外地 ) 此流。現在必須分析分割槽規則以瞭解原因。

### 資料路徑驗證 — zoning-rules

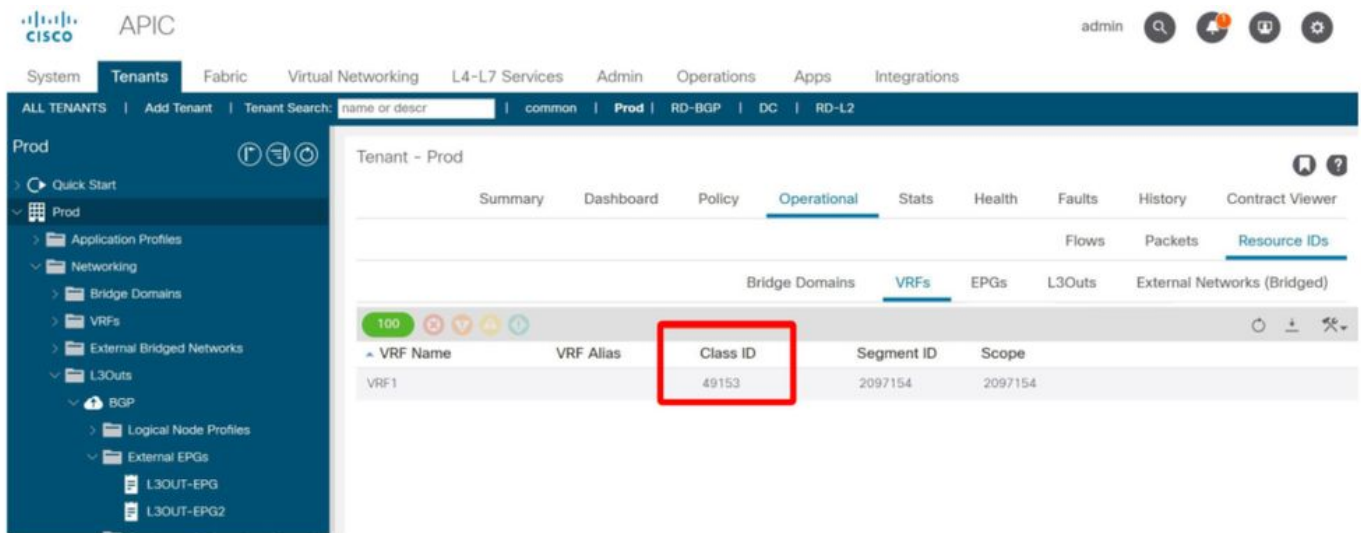
與最後2個測試相對應的分割槽規則如下：

- 預期 — 流點選分割槽規則行4336 ( ICMP2合約 ) 。
- 意外 — 流命中分割槽規則行4341 ( ICMP1合約 ) 。

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 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) |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+
```

兩個流都派生出49153的src pcTag。這是VRF的pcTag。這可在UI中驗證：

### 驗證VRF的pcTag



將0.0.0.0/0字首與L3Out一起使用時，會發生以下情況：

- 從內部EPG到0.0.0.0/0的L3Out EPG的流量將派生目標pcTag 15。
- 從0.0.0.0/0的L3Out EPG到ACI內部EPG的流量將派生VRF(49153)的源pcTag。

contract\_parser指令碼提供了分割槽規則的整體檢視：

```

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應用確認資料包使用的pcTag

ELAM助理應用提供了另一種方法來確認即時流量的源和目標pcTag。

以下螢幕截圖顯示了從pcTag到pcTag的流量32771ELAM結49153。

## 用於從src到dst的ELAM助理應32771程式輸49153

Packet Forwarding Information	
<b>Forward Result</b>	
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
<b>Contract</b>	
Destination EPG pcTag (dclass)	32771 (Prod:App:EPG2)
Source EPG pcTag (sclass)	49153 (Prod:VRF1:l3out-BGP:vlan-2551)

## 結論

在VRF中，必須仔細跟蹤0.0.0.0/0的使用情況，因為使用該子網的每個L3Out都將繼承應用於使用該子網的其他每個L3Out的合約。這可能會導致計畫外的許可流。

## 共用L3Out

### 概觀

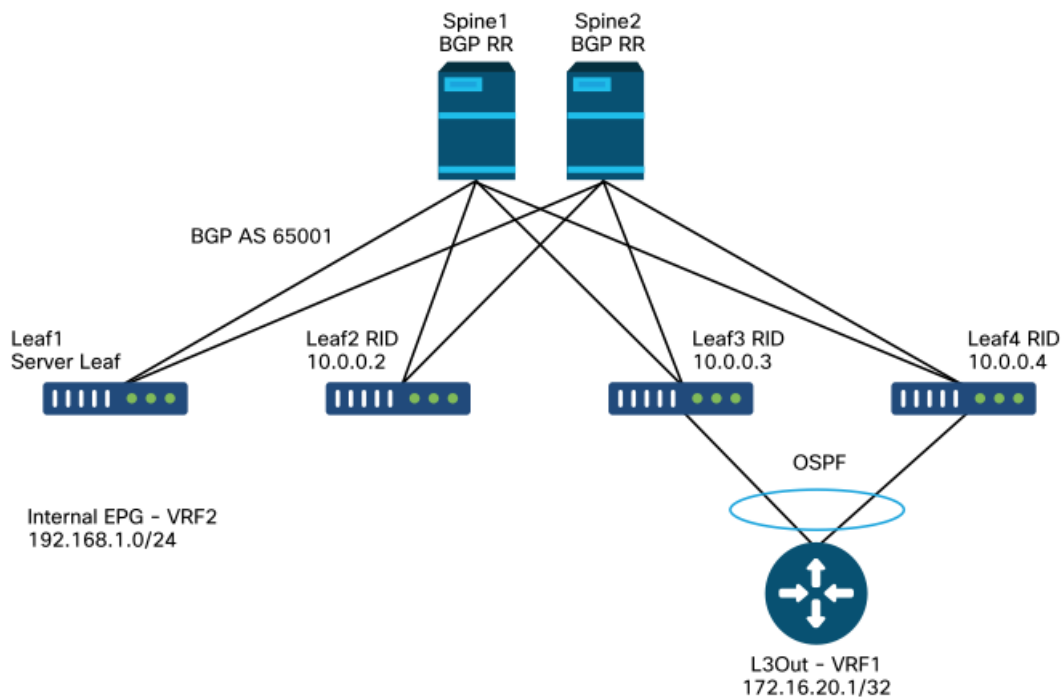
本節將討論如何在共用L3Out配置中對路由通告進行故障排除。術語「共用L3Out」是指以下情況：L3Out位於一個VRF中，但與L3Out有合約的內部EPG位於另一個VRF中。使用共用L3Outs時，路由洩漏在內部對ACI交換矩陣執行。

本節不會深入介紹有關安全策略故障排除的詳細資訊。有關資訊，請參閱本書的「安全策略」一章。出於安全考慮，本節還將不再詳細介紹外部策略字首分類。請參閱「外部轉發」一章中的「合約和L3Out」一節。

本節使用以下拓撲作為示例。

### 共用L3Out拓撲





在高級別上，必須準備好以下配置才能使共用L3Out正常工作：

- L3Out子網必須配置有「共用路由控制子網」範圍，以便將外部路由洩漏到內部VRF。也可以選擇「Aggregate Shared」（聚合共用）選項，以洩漏比配置的子網更為具體的所有路由。
- 必須使用「共用安全匯入子網」範圍配置L3Out子網，以規劃通過此L3Out進行通訊所必需的安全策略。
- 內部BD子網必須設定為「在VRF之間共用」和「向外部通告」，以在外部VRF中對BD子網進行程式設計並進行通告。
- 必須在共用L3Out的內部EPG和外部EPG之間配置「租戶」或「全域性」範圍合約。

下一節將詳細介紹如何在ACI中通告和學習洩漏的路由。

## 共用的L3Out工作流程 — 學習外部路由

本節將概述在將已學習外部路由通告到交換矩陣時的路徑。

### 在邊界枝葉上顯示的外部路由

此命令將顯示從OSPF獲知的外部路由：

```
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
```

接下來，必須將路由匯入BGP。預設情況下，所有外部路由都應匯入到BGP中。

## 邊界枝葉上的BGP驗證

該路由必須位於BGP VPNv4 Address-family中，並且路由目標將分佈在整個交換矩陣中。route-target是由外部VRF匯出並由需要接收路徑的任何內部VRF匯入的BGP擴展社群。

接下來，檢驗由BL上的外部VRF匯出的路由目標。

```
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
```

上面的輸出顯示，從外部VRF通告到VPNv4的任何路徑都應收到路由目標65001:2392068。

接下來，驗證bgp路徑：

```
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

```

上面的輸出顯示，路徑具有正確的路由目標。也可使用「show bgp vpnv4 unicast 172.16.20.1 vrf overlay-1」命令驗證VPNv4路徑。

## 伺服器枝葉上的驗證

對於內部EPG枝葉要安裝BL通告的路由，它必須將路由目標（如上所述）匯入內部VRF。可以檢查內部VRF的BGP進程以驗證這一點：

```

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

```

上面的輸出顯示了匯入由外部VRF匯出的路由目標的內部VRF。此外，還引用了「匯入路由對映」。匯入路由對映包括在共用L3Out中用「共用路由控制子網」標誌定義的特定字首。

可以檢查路由對映內容，以確保它包含外部字首：

```

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

```

上面的輸出顯示了包含要匯入的子網的匯入路由對映。

最終驗證包括檢查該路由是否在BGP表中，以及它是否安裝在路由表中。

伺服器枝葉上的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

```

該路由被匯入到內部VRF BGP表中，並具有期望的路由目標。

可以驗證安裝的路由：

```

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

```

上述輸出使用特定的「vsh -c」命令獲取「detail」輸出。「detail」標誌包括重寫VXLAN VNID。這是外部VRF的VXLAN VNID。當BL接收到具有此VNID的資料平面流量時，它知道在外部VRF中做出轉發決策。

rw-vnid值以十六進位制表示，因此轉換為十進位制形式將獲得2392068的VRF VNID。使用「show system internal epm vrf all」搜尋相應的VRF 枝葉上的| grep 2392068'。可以使用'moquery -c fvCtx -f 'fv.Ctx.seg="2392068"'命令對APIC==行全域性搜尋。

下一躍點的IP也應指向BL PTEP，而「%overlay-1」表示下一躍點的路由查詢在重疊VRF中。

## 共用的L3Out工作流 — 通告內部路由

與前面幾節一樣，將內部BD子網從共用L3Out中通告出去的工作由下列各項處理：

- BD子網（內部VRF）作為靜態路由安裝在BL（外部VRF）上。此靜態路由部署是內部EPG和L3Out之間合約關係的結果。
- 當BD子網上設定了「Advertised External」範圍時，靜態路由會重新分發到外部協定中。

## 檢驗BL上的BD靜態路由

```

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

```

請注意，在上述輸出中，內部VRF的VNID已設定為重寫。下一跳也設定為proxy-v4-anycast地址。

以上路由是通過「路由通告」部分中顯示的相同路由對映在外部進行通告。

如果將BD子網設定為「向外部通告」，則將其重新分發到內部EPG與其有合約關係的每個L3Out的外部協定。

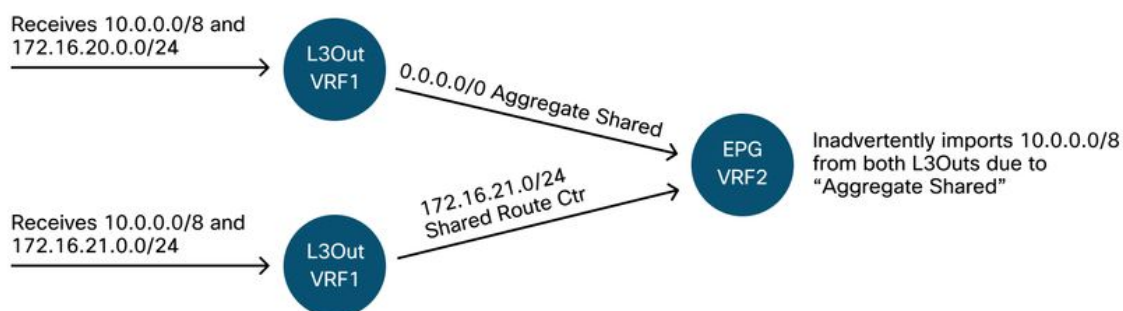
## 共用的L3Out故障排除場景 — 意外的路由洩漏

此方案在外部VRF中有多個L3Out，並且內部EPG正在從L3Out接收路由，其中網路未使用「共用」範圍選項進行定義。

## 「聚合共用」的使用

請考慮下圖：

### 意外的路由洩漏



具有從「共用路由控制子網」標誌程式設計的字首清單的BGP匯入映射在VRF級別應用。如果VRF1中的一個L3Out具有「共用路由控制子網」的子網，則在VRF1內L3Out上接收的所有與此共用路由控制子網匹配的路由都將匯入到VRF2。

上述設計可能導致意外的流量。如果內部EPG與意外的廣告L3Out EPG之間沒有合約，則會發生流量丟棄。

## 關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。