排除ACI中的L3Out子網分類故障

目錄

簡介

本文檔介紹思科ACI的L3Out EPG中的外部子網分類，

縮寫

- BD:網橋域
- EPG:終端組
- ExEPG:外部終結點組
- RIB:路由資訊庫
- VRF:虛擬路由和轉送
- 類ID:標識EPG的標籤

外部EPG分類

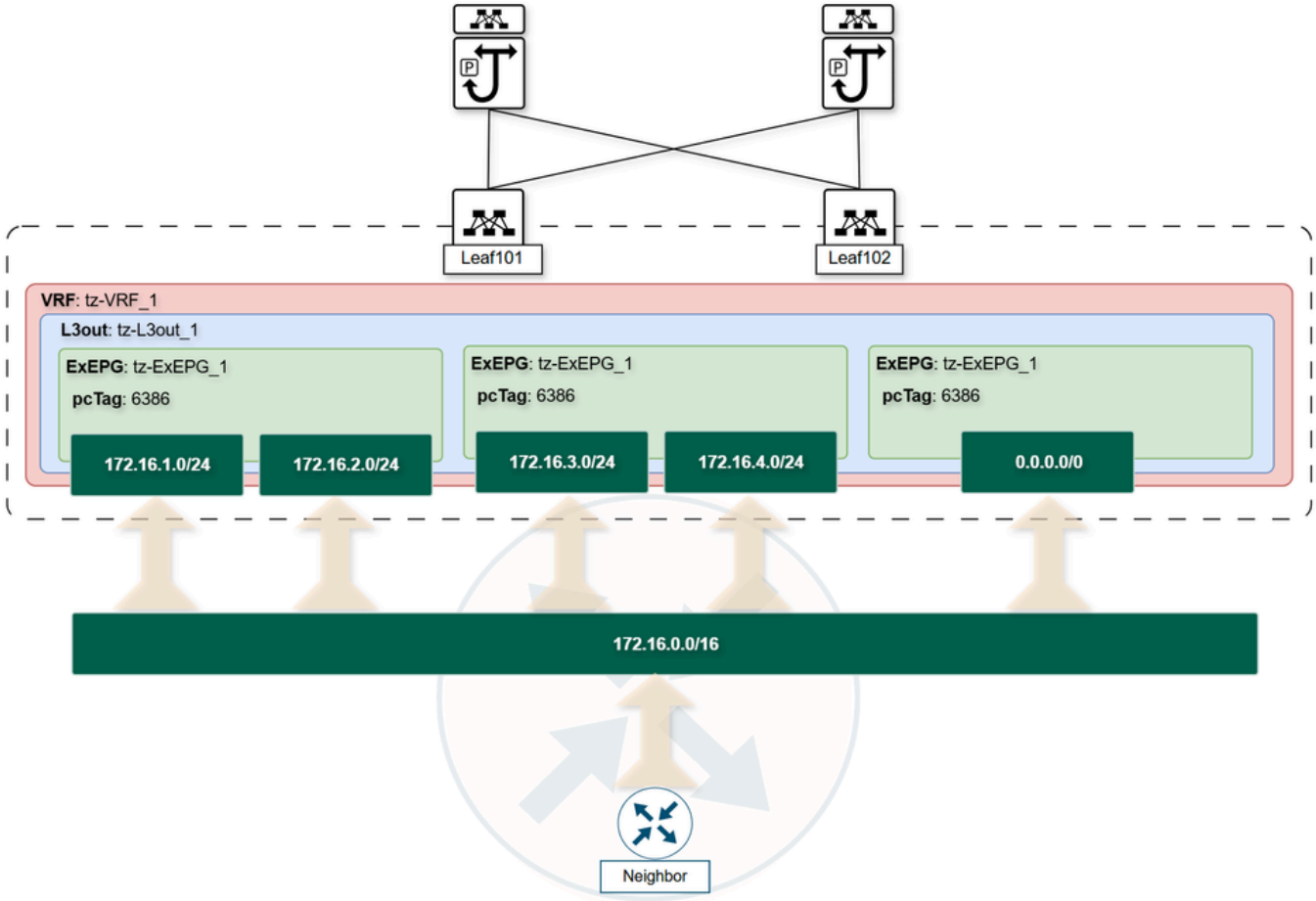
思科ACI中的外部EPG表示通過L3Outs連線的外部路由網路。與常規EPG對端點進行分類的方式類似，外部EPG對外部子網按每個VRF進行分類，這意味著每個子網在其VRF上下文中必須是唯一的。

一個常見的誤解是外部EPG子網僅包含通過動態路由協定接受的字首。但是，建立L3Out時，有一個預設路由對映過濾傳入通告；因此，預設情況下會接受由動態路由協定通告的所有字首。在ExEPG下定義子網的主要用途是分類，僅將唯一的pcTag分配給ExEPG中包含的子網，用於合約實施和策略應用。

此分類可實現精細的策略控制。例如，單個外部鄰居可以將超網通告給ACI，然後將其分段為多個

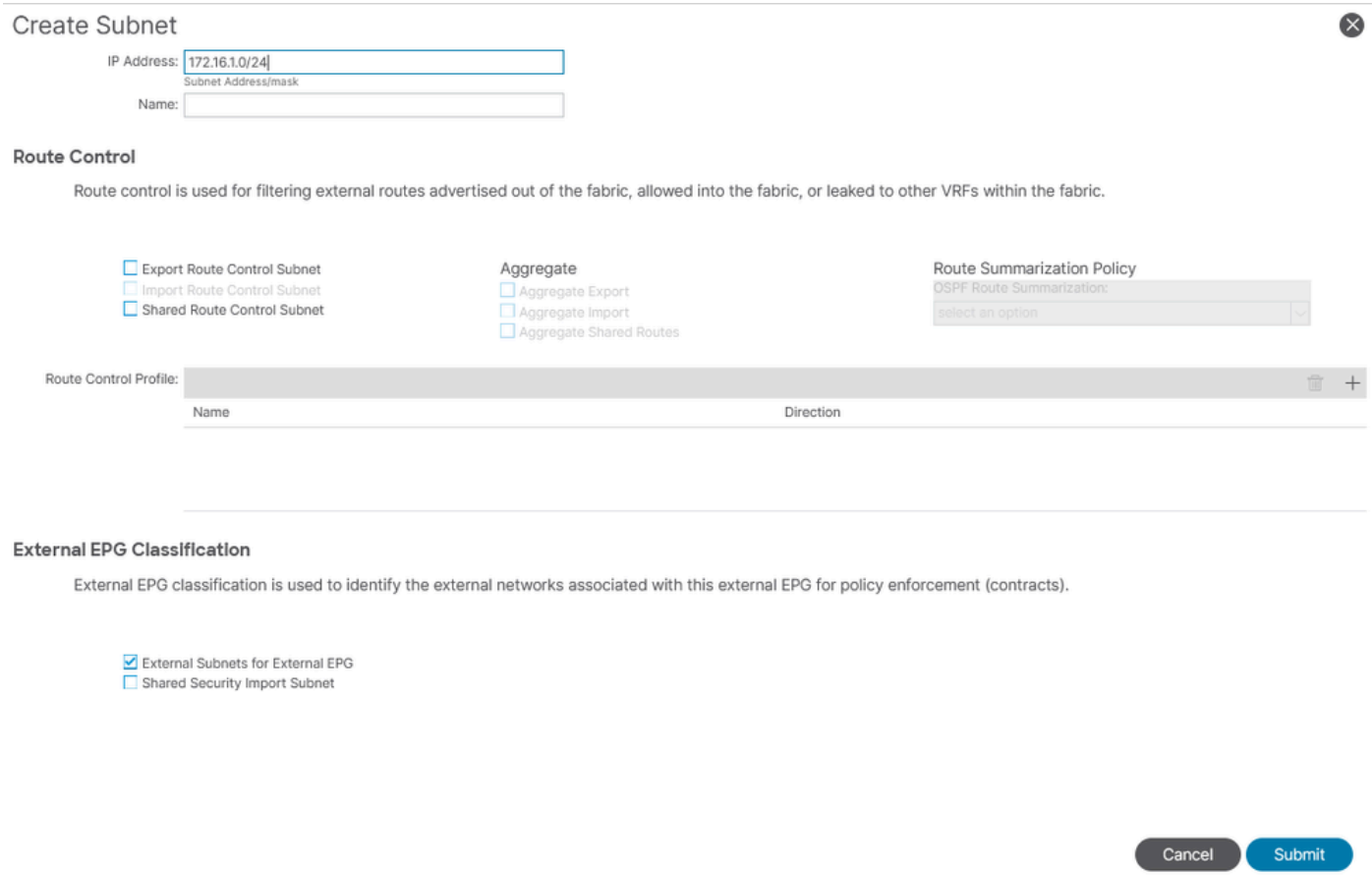
ExEPG。這允許將不同的約定操作應用於不同的子網，例如允許特定的內部EPG僅與指定的外部子網通訊，或在到達其最終目的地之前將目的地為特定字首的流量重定向到PBR節點。

此圖說明思科ACI如何根據外部EPG對外部子網進行分類，從而實現精確的流量分段和合約實施。



外部EPG子網標誌

為了在ACI中對ExEPG中的外部字首進行分類和管理，在ExEPG下建立子網字首時，會配置特定的子網標誌。本節詳細介紹每個標誌及其預期用途：



- 外部EPG的外部子網：
此標籤表示子網位於ACI交換矩陣之外，並且未在任何網橋域或EPG中配置。僅當字首由路由鄰居通告或靜態注入RIB時，才必須使用該字首。預設情況下啟用此標誌。

- 匯出路由控制子網：
此標籤指定通過動態路由協定將子網從ACI通告給路由鄰居。不能同時啟用外部子網的External Subnet for External EPG標誌，因為這樣做可能會導致第3層路由環路。由於ACI將子網分類為外部子網並通告給後端，因此儘管路由協定中存在環路避免機制，但這樣做可能導致路由不一致。

- 共用路由控制子網：
當子網字首要在多個VRF之間共用時，設定此標誌，從而啟用情景之間的路由洩漏。

- 共用安全匯入子網：
與共用路由控制子網標誌結合使用，這樣可以跨不同的VRF共用外部子網的安全pcTags，從而促進一致的策略實施。

- 匯入路由控制子網：
此標誌允許精細控制從路由鄰居接收的字首。預設情況下，ACI接受所有傳入路由通告；啟用此標誌需要啟用路由控制實施以過濾傳入字首。

- 聚合部分：
此部分僅適用於quad-0(0.0.0.0/0)子網，它總結了RIB中用於聚合匯出或匯入的所有字首。當子網洩漏到其他VRF時，它們將總結為聚合共用路由以最佳化路由表。

驗證和疑難排解指令

路由

首先，該路由必須存在於邊界枝葉交換機上的VRF路由表中。例如，此命令顯示VRF tz:tz-VRF_1中的BGP路由：

<#root>

Leaf101#

show ip route bgp vrf tz:tz-VRF_1

IP Route Table for VRF "tz:tz-VRF_1"

'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

172.16.1.0/24

, ubest/mbest: 1/0

***via 10.10.1.2**

%tz:tz-VRF_1, [20/0], 00:00:04, bgp-65002, external, tag 65003
Leaf101#

這確認該路由已安裝在VRF路由表中並可用於轉發決策。

分類

路由出現在路由表後，分類會根據策略確定處理流量的方式。在ACI中，分類與ExEPG及其關聯子網相關聯。

要驗證ExEPG下的子網分類，可以為l3extInstP類（代表外部EPG例項）查詢APIC。其子類l3extSubnet列出了在該ExEPG下配置的子網。舉例來說：

<#root>

moquery -c l3extInstP -f 'l3ext.InstP.dn*"[tenant name].*[l3out name]"' -x rsp-subtree=children rsp-

<#root>

APIC#

moquery -c l3extInstP -f 'l3ext.InstP.dn*"tz.*l3out"' -x rsp-subtree=children rsp-subtree-class=l3extSub

```
Total Objects shown: 1

# l3ext.InstP

name : tz-ExEPG_1


!-- cut for brevity --!
configSt : applied
descr :
dn : uni/tn-tz/out-l3out/instP-tz-ExEPG_1
!-- cut for brevity --!
floodOnEncap : disabled
isSharedSrvMsiteEPg : no
lcOwn : local
matchT : AtleastOne
mcast : no
modTs : 2025-09-10T00:36:49.239+00:00
monPolDn : uni/tn-common/monepg-default
nameAlias :
pcEnfPref : unenforced

pcTag : 32771


pcTagAllocSrc : idmanager
prefGrMemb : exclude
prio : unspecified
rn : instP-tz-ExEPG_1
scope : 3047430
status : modified
targetDscp : unspecified
triggerSt : triggerable
txId : 1152921504612318828
uid : 15374
userdom : :all:

# l3ext.Subnet

ip : 172.16.1.0/24


!-- cut for brevity --!
dn : uni/tn-tz/out-l3out/instP-tz-ExEPG_1/extsubnet-[172.16.1.0/24]
extMngdBy :
lcOwn : local
modTs : 2025-09-10T01:05:13.249+00:00
monPolDn : uni/tn-common/monepg-default
!-- cut for brevity --!
rn : extsubnet-[172.16.1.0/24]

scope : import-security


status :
uid : 15374
userdom : :all:

APIC#
```

如果未返回l3extSubnet類的輸出，則表示在外部EPG下未配置任何子網。如果沒有配置子網，ACI無法將pcTag與傳入流量子網相關聯，導致流量被丟棄，儘管路由表中存在路由。

要注意的另一個重要方面是子網範圍，它代表為相關子網設定的標誌：

- Import-security

該子網已標有外部EPG的外部子網。

- export-rtctrl

該子網已標籤了匯出路由控制。

- import-rtctrl

該子網已標籤有匯入路由控制。

- 共用安全

該子網已標籤有共用安全匯入子網。

- shared-rtctrl

該子網已標籤有共用路由控制。

路由協定和控制平面進程在收到來自所述鄰居的字首後更新路由表，然後將其程式設計到HAL L3轉發表中。HAL L3路由表示已程式設計到枝葉交換機上的硬體轉發表(ASIC)中的實際第3層路由。這些路由源自路由協定和路由表計算，用於轉發決策。

<#root>

<-- When the prefix is not configured under the External EPG, a classification of 0xf is seen -->
Leaf101#

vsh_lc -c 'show platform internal hal l3 routes vrf tz:tz-VRF_1' | egrep "Prefix/Len|172.16.1.0" | cut -

VRF | Prefix/Len | RT|CLSS| Flags

4675| 172.16.1.0/ 24| UC| f|spi,dpi

Leaf101#

<-- When the prefix is configured under the External EPG, a classification of the pcTag in hexadecimal
Leaf101#

vsh_lc -c 'show platform internal hal l3 routes vrf tz:tz-VRF_1' | egrep "Prefix/Len|172.16.1.0" | cut -

VRF | Prefix/Len | RT|CLSS| Flags

4675| 172.16.1.0/ 24| UC|8003|spi,dpi

Leaf101#

```
Leaf101#
vsh_lc -c '
dec 0x8003'

32771
Leaf101#
```

隨後，當在ExEPG中使用外部EPG的外部子網標誌配置子網時，一個名為Policy Manager(policy-mgr)的內部進程將使用此子網條目和關聯的pcTag更新其字首到pcTag的對映表。Policy Manager用作交換矩陣集中式策略協調引擎，將高級策略定義轉換為ACI交換矩陣中可操作的配置。這通過基於配置的外部子網執行正確的流量分類和轉發決策的pcTags ，確保一致且安全的應用程式連線和網路行為。

<#root>

```
Leaf101#
vsh -c 'show system internal policy-mgr prefix' | egrep "tz:tz-VRF_1"

3047430 36 0x80000024 Up tz:tz-VRF_1 ::/0 15 True True False False
3047430 36 0x24 Up tz:tz-VRF_1 0.0.0.0/0 15 True True False False
3047430 36 0x24 Up tz:tz-VRF_1 172.16.1.0/24 32771 True True False False

Leaf101#
```

這確認字首172.16.1.0/24正被鄰居通告給ACI邊界枝葉交換機，並且ACI已將字首分類到pcTag32771下

合約

分割槽規則是在交換矩陣內的EPG（包括ExEPG）之間實施合約策略的基本過程。外部EPG的VRF VNID（範圍）和pcTag可用於定義和驗證在源EPG和目標EPG之間應用的通訊規則。本質上，分割槽規則將高級合約關係轉換為在枝葉交換機上程式設計的特定的、可執行的規則。

要考慮的一個重要方面是合約在交換矩陣中的安裝位置。預設情況下，VRF配置策略控制實施方向設定為輸入。此設定確定源終端所在的枝葉交換機上安裝了給定合約的分割槽規則。

在本練習中，流量從L3Out傳入，分割槽規則安裝在連線到該L3Out的邊界枝葉上，因為該枝葉充當進入交換矩陣的流量的源枝葉。

<#root>

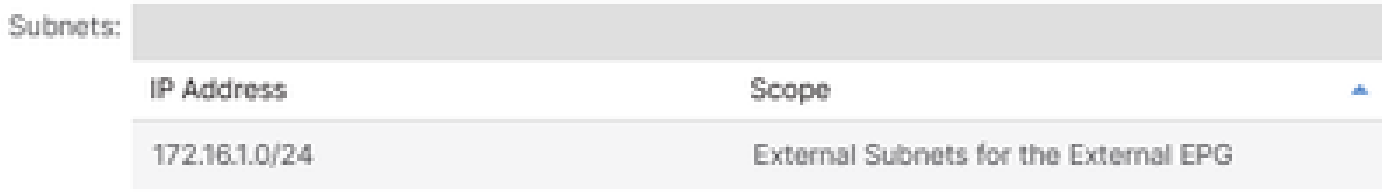
Leaf101#

show zoning-rule scope 3047430 | egrep "Rule|---|32771"

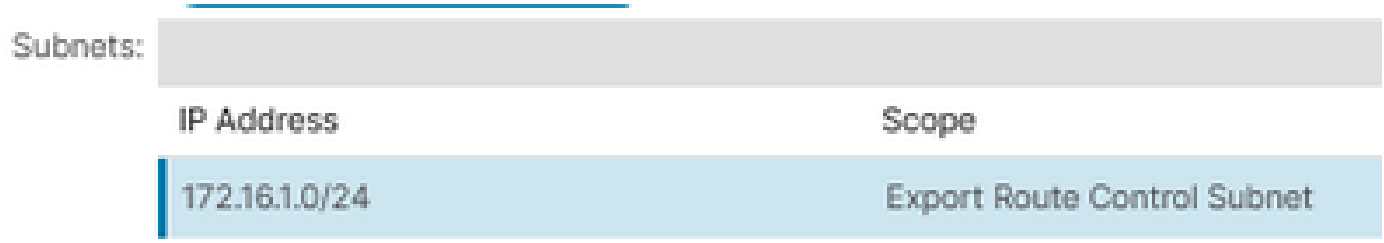
```
+---------+--------+--------+----------+----------------+---------+---------+--------------------+------
| Rule ID | SrcEPG | DstEPG | FilterID |      Dir       | operSt  |  Scope  |        Name        |
+---------+--------+--------+----------+----------------+---------+---------+--------------------+------
|  4441   | 49153  | 32771  |     5    |     bi-dir     | enabled | 3047430 |     tz:Contract    |
|  4500   | 32771  | 49153  |     5    | uni-dir-ignore | enabled | 3047430 |     tz:Contract    |
+---------+--------+--------+----------+----------------+---------+---------+--------------------+------
Leaf101#
```

傳輸路由

傳輸路由通過通告從一個L3Out獲知的外部路由到另一個L3Out，使交換矩陣能夠充當傳輸網路。要正確配置傳輸路由，傳入子網必須標有External Subnet for External EPG標誌。



同時，將此子網通告給其他外部對等體的L3Out必須在相應子網中啟用匯出路由控制子網標誌。此標籤允許通過該L3Out上配置的路由協定將子網重新分發並從交換矩陣通告。



最後，需要配置接收和匯出L3out之間的合約以完成路由分發過程。

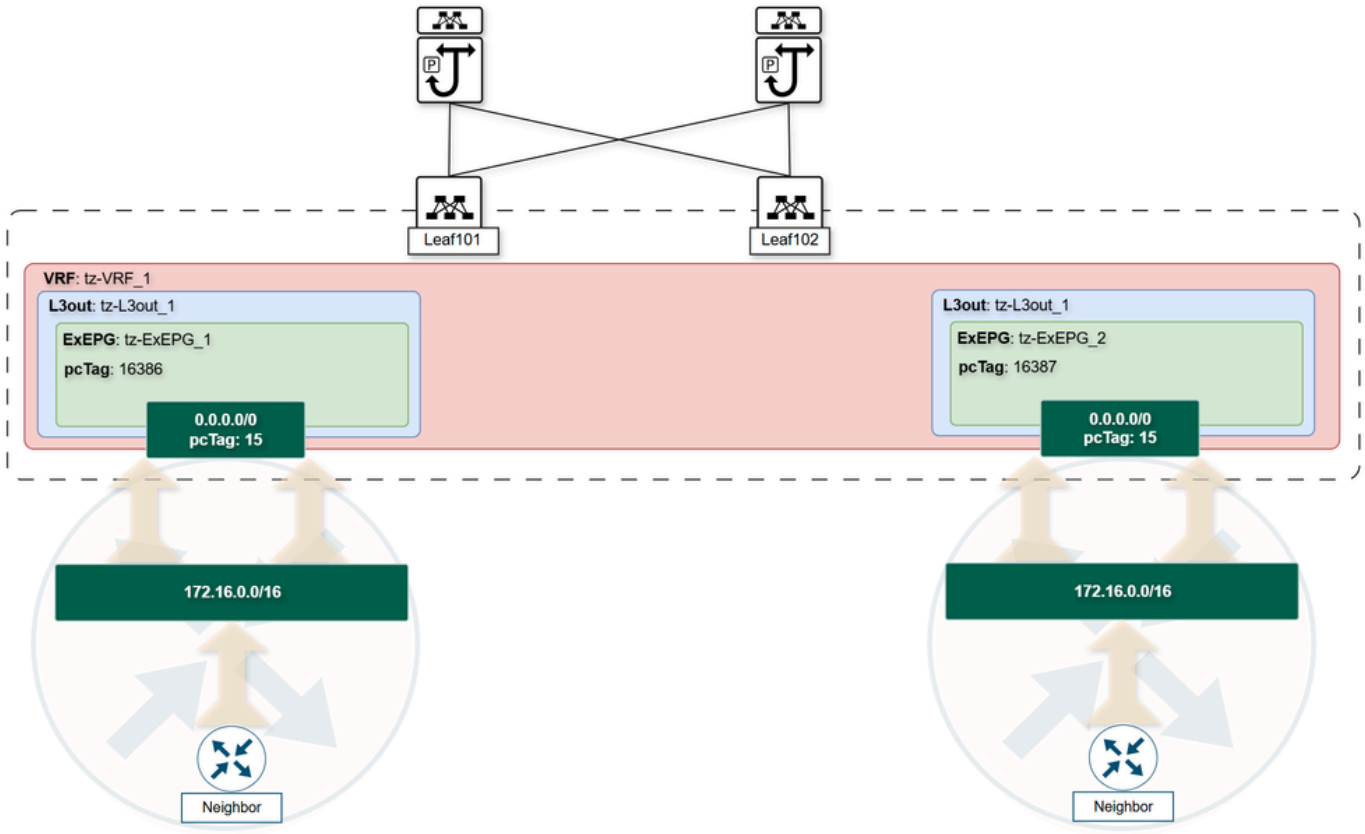
子網外部EPG分類中的常見問題

pcTag 15

在此之前，本文檔中曾指出，ExEPG子網可幫助您根據策略實施原因將子網分類到正確的pcTag中。此分類的一個重要例外是使用「外部EPG的外部子網」標誌配置的4-0子網(0.0.0.0/0)。此子網始

終分配保留的pcTag 15，有效地用作VRF中所有外部流量的萬用字元。

此圖顯示了在同一VRF中的多個ExEPG上為外部EPG配置外部子網的quad-0所面臨的問題：



- 四進位制0子網通常被誤認為是預設路由。雖然有時確實如此（例如動態路由鄰居僅向ACI L3Out通告預設路由時），但是quad-0子網在ACI中的角色卻更廣泛，即是一種全捕獲分類。
- 常見的做法是配置多個ExEPG和四進位制0子網以接受鄰居通告的所有字首。雖然這實現了廣泛接受的目標，但是當在同一個VRF中配置多個具有4-0的ExEPG時，可能導致意外的非對稱路由。當同一個VRF中的多個ExEPG配置了4-0作為外部子網時，ACI無法決定性地選擇將哪個L3Out用於特定目標子網。反之，任意選擇一個L3Out。
- 如果隨機選擇的L3Out沒有允許通訊所需的合約，則此行為可能導致非對稱路由、流量間歇性甚至流量丟棄。

重疊的子網

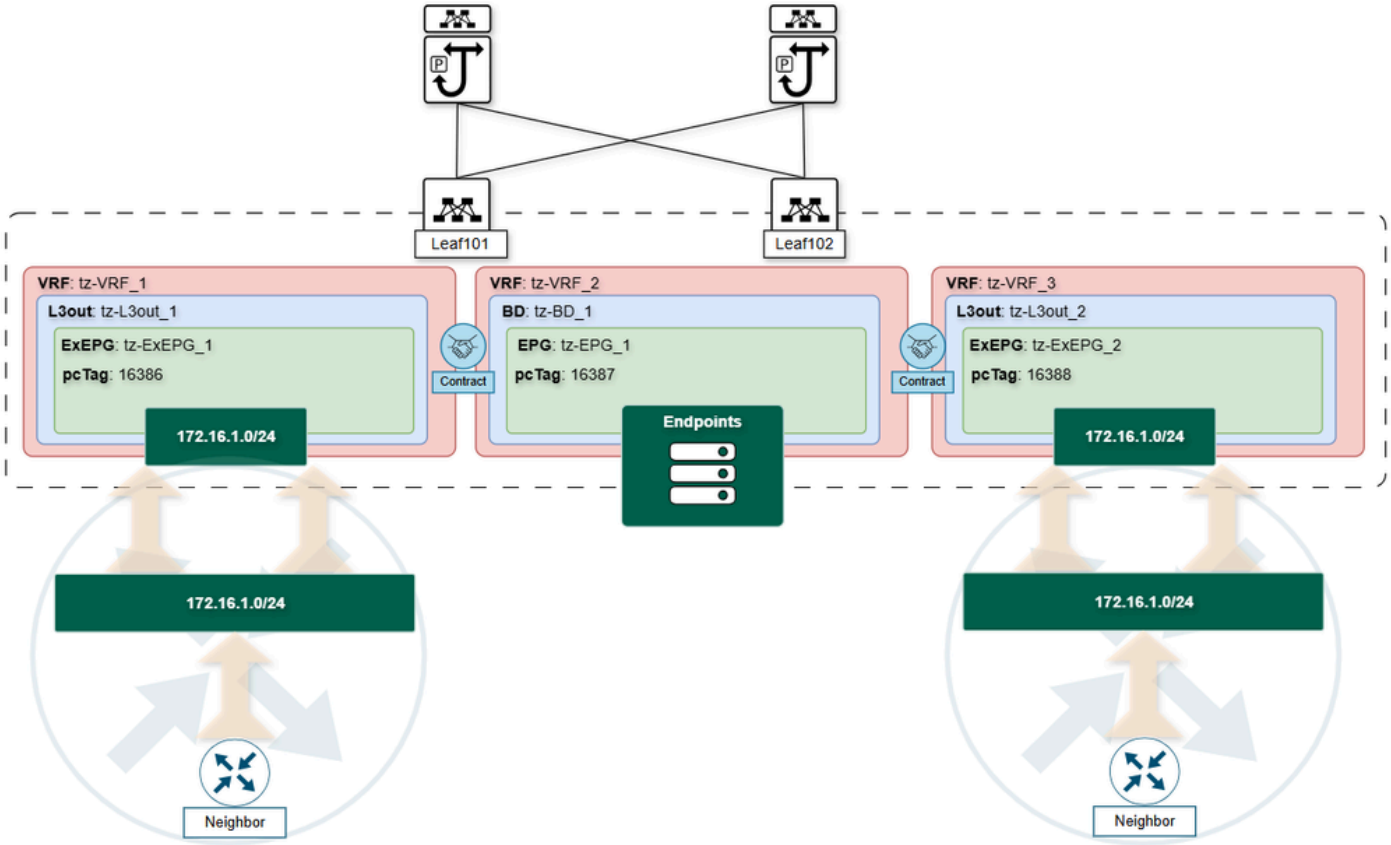
不允許在不同ExEPG之間配置相同的子網。嘗試這樣做會觸發故障「F0467:字首條目已在另一個EPG」中使用，可防止VRF中的子網重複。

但是，由於每個VRF都維護獨立的路由表上下文，因此不同VRF之間可能存在重疊的子網。這種分離允許在屬於不同VRF的ExEPG中配置相同的子網。儘管如此，在涉及這些重疊子網的VRF路由洩漏時，必須小心謹慎，因為子網分類(pcTag)與路由資訊(RIB)的衝突可能導致非對稱轉發決策。

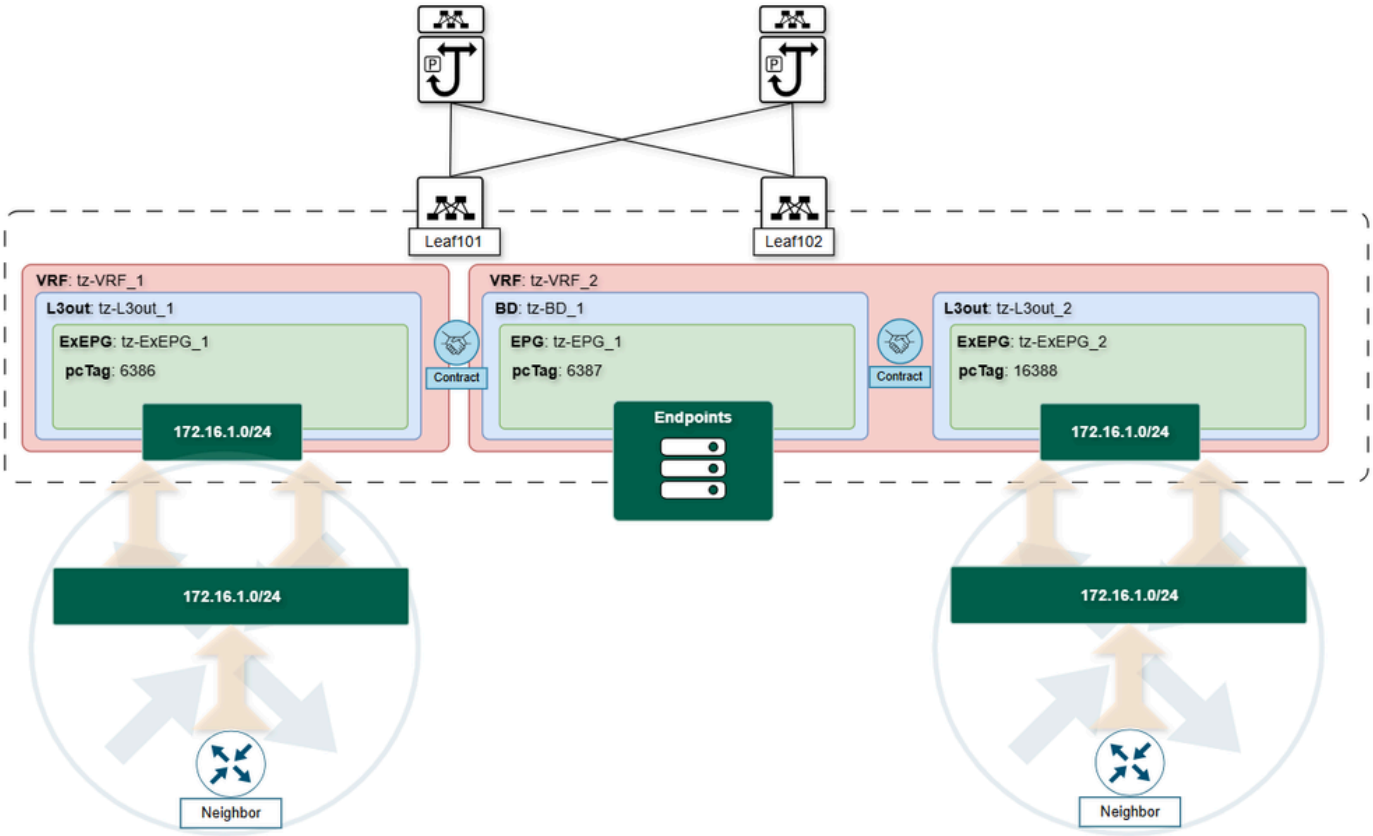
主要場景包括：

- 從兩個VRF到第三個VRF的路由洩漏：
  當兩個VRF將同一子網洩漏到第三個VRF時，接收VRF會根據來自APIC的共用策略安裝它收到的第一個子網。如果處理此VRF的枝葉交換機重新啟動或路由選擇更改，路由表可能會更新

到其他L3Out，從而導致轉發行為不一致。



- 與洩漏子網重疊的本地到VRF L3Out ExEPG:
  在使用路由洩漏的設計中，如果本地L3Out ExEPG配置了與洩漏的子網相同的子網，則本地路由條目始終優先於洩漏的路由。



這些情況突出表明，非對稱轉發問題源自分類和轉發決策層，而不是路由表本身。雖然子網分類將子網與用於策略實施的特定L3Out和ExEPG相關聯，但路由表可以指向不同的L3Out目標。這種不匹配會導致流量轉發不一致，從而導致潛在的連線問題或策略實施差距。

匯入路由控制預設行為更改

預設情況下，ACI接受來自鄰居的所有傳入路由通告。要控制接受哪些字首，必須啟用路由控制實施：L3Out根對象上的入站：

導航到Tenants > [tenant name] > Networking > L3outs > [L3out name]。

此操作在選定的路由協定下建立路由對映。

<#root>

```
Border Leaf#
```

show ip bgp neighbors vrf tz:tz-VRF1 | egrep route-map

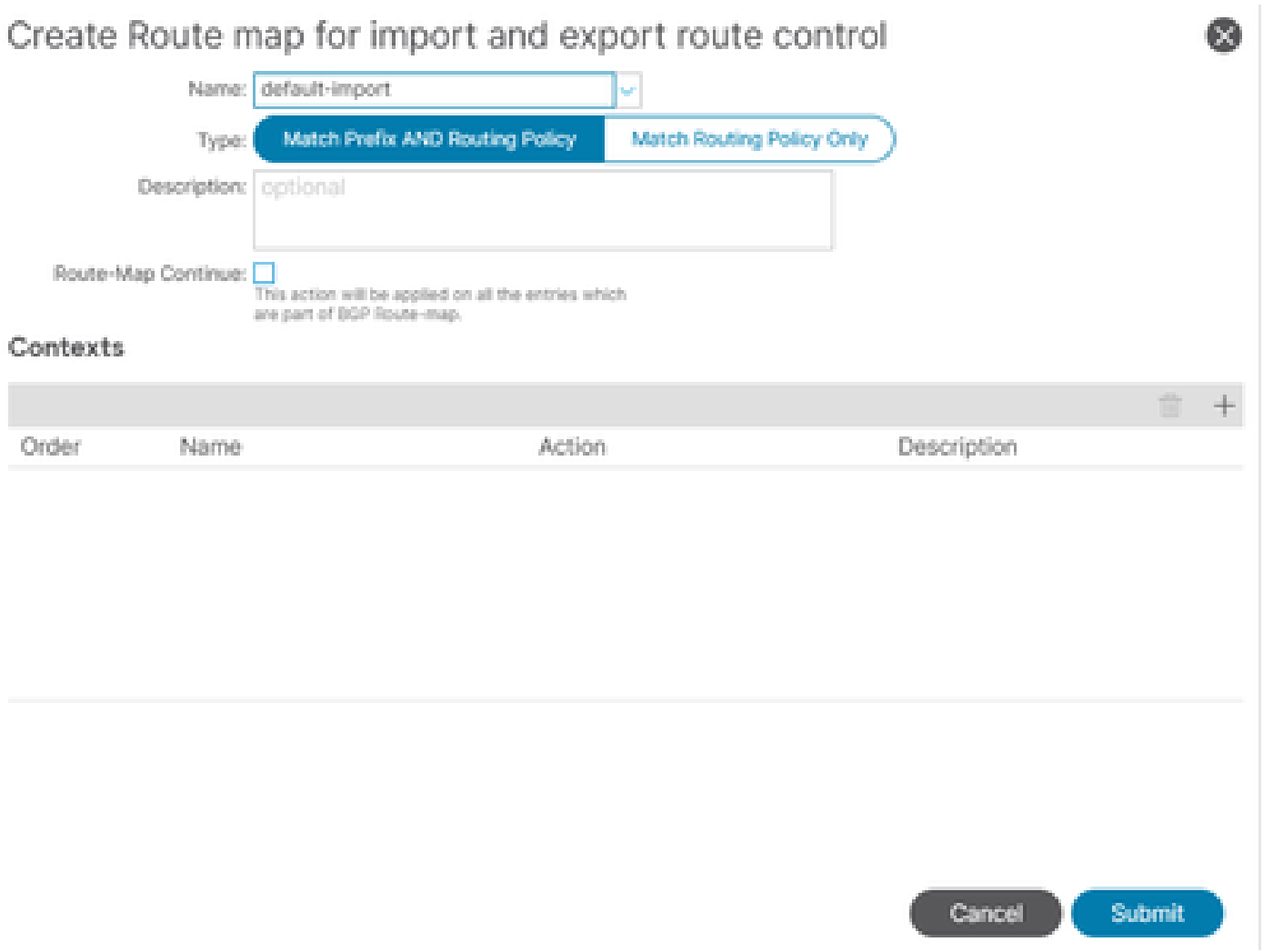
```
Outbound route-map configured is exp-l3out-ExEPG-peer-2981888, handle obtained
```

```
Inbound route-map configured is imp-l3out-ExEPG-peer-2981888, handle obtained


Border Leaf#
show route-map imp-l3out-ExEPG-peer-2981888


route-map imp-l3out-ExEPG-peer-2981888,
permit
, sequence 15801
Match clauses:
ip address prefix-lists: IPv4-peer49155-2981888-exc-ext-inferred-import-dst


ipv6 address prefix-lists: IPv6-deny-all
Set clauses:

Border Leaf#
show ip prefix-list IPv4-peer49155-2981888-exc-ext-inferred-import-dst


ip prefix-list IPv4-peer49155-2981888-exc-ext-inferred-import-dst: 1 entries
seq 1 permit 172.16.1.0/24
Border Leaf#
```

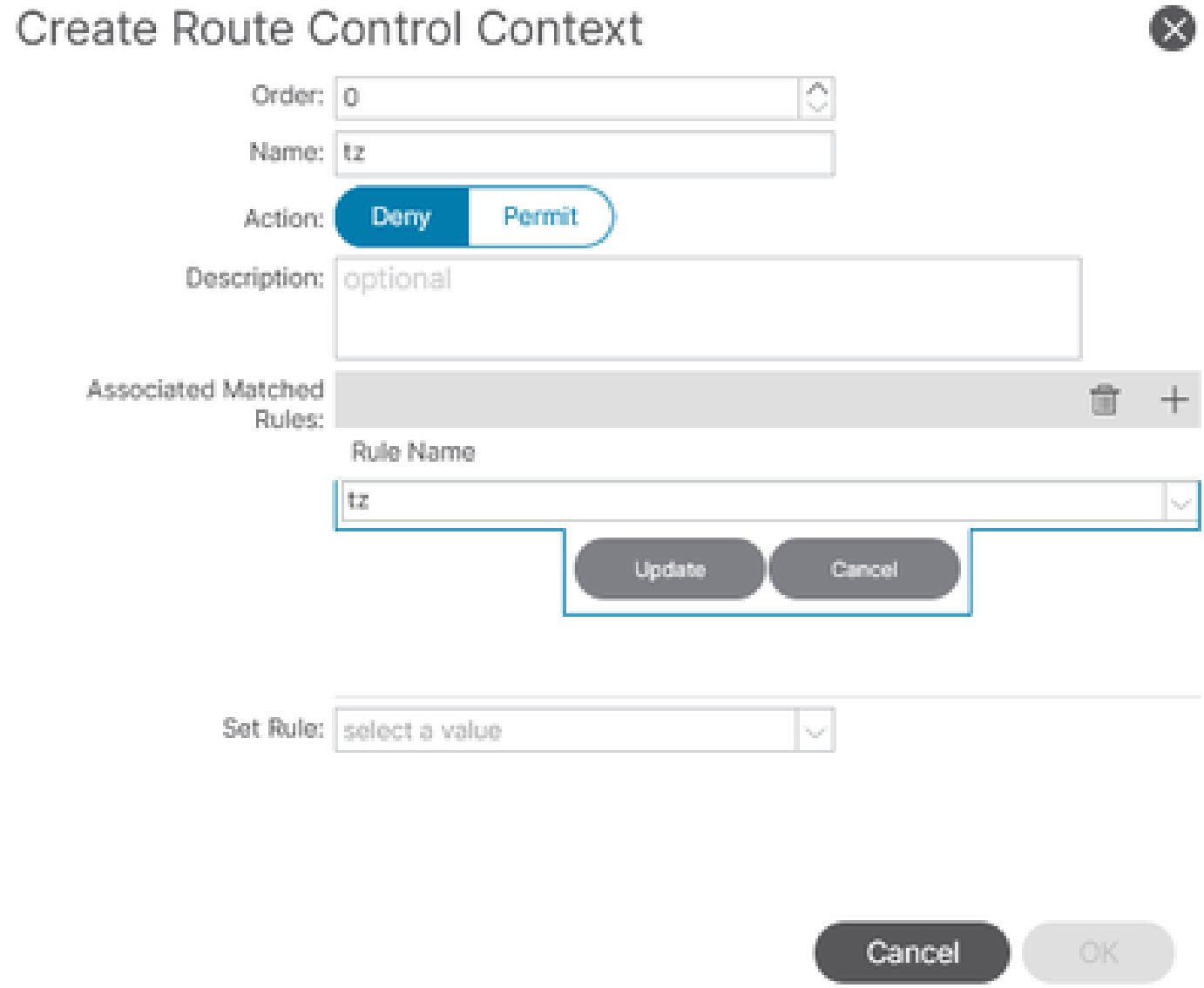
預設情況下，此匯入路由對映允許所有傳入字首。要修改此行為，請執行以下操作：

導航到Tenants > [tenant name] > Networking > L3outs > [L3out name] > Route map for import and export route control

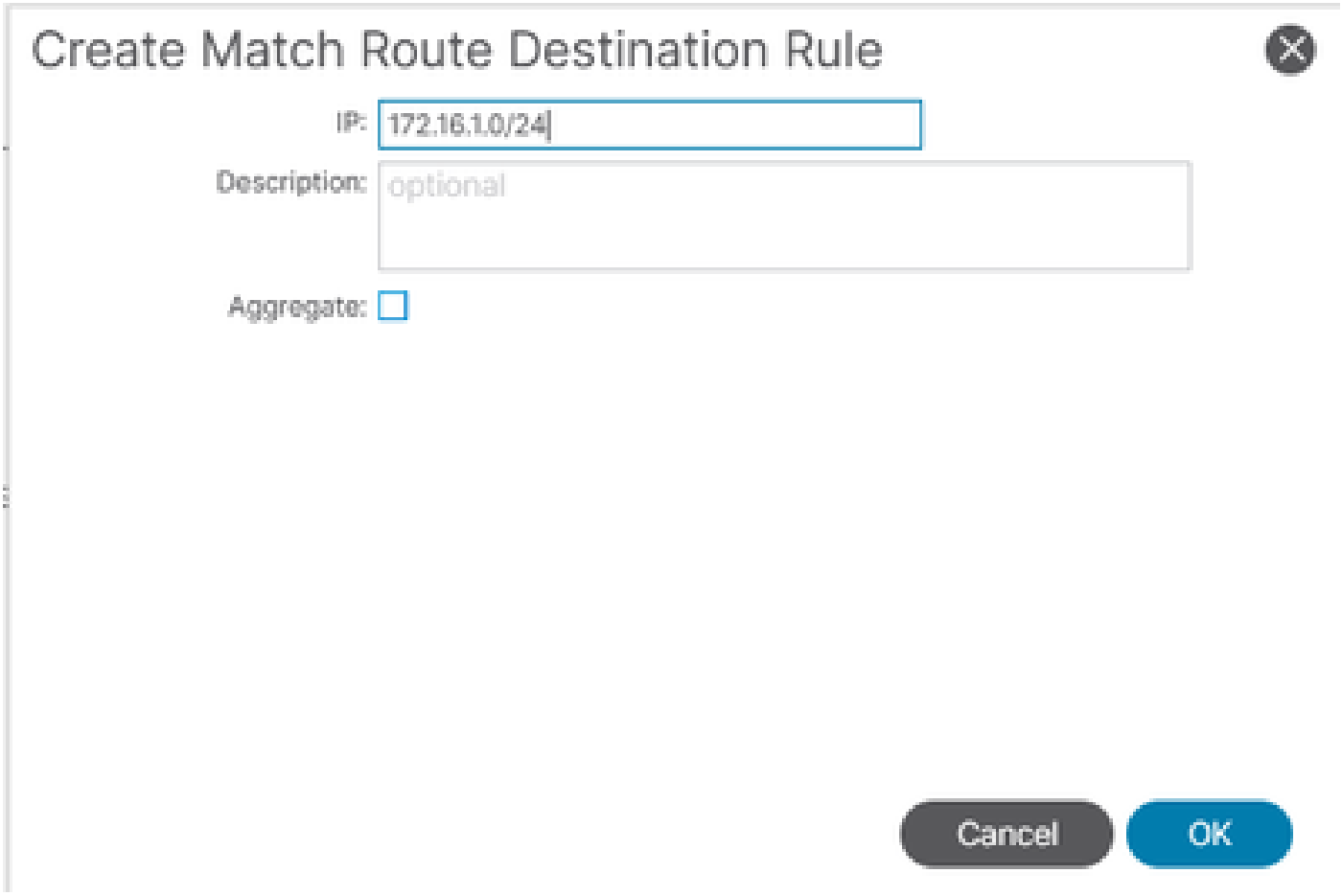
選擇預設的匯入路由對映，或使用右上角的gear圖示建立一個新的匯入路由對映。



在Context部分中，建立新的Associated Matched Rule。



在Match Rules部分，滾動到Match Prefix，然後新增要控制的特定子網。



提交策略後，匯入路由對映操作會相應地更改，從而強制實施所需的字首過濾。

<#root>

```
Border Leaf#
```

show route-map imp-l3out-ExEPG-peer-2981888

```
route-map imp-l3out-ExEPG-peer-2981888,
```

deny

```
, sequence 8001
Match clauses:
ip address prefix-lists: IPv4-peer49155-2981888-exc-ext-in-default-import2tz0tz-dst
ipv6 address prefix-lists: IPv6-deny-all
Set clauses:

Border Leaf#
```

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）。