排除ACI中的L3Out子网分类故障

目录

简介

本文档介绍思科ACI的L3Out EPG中的外部子网分类。

缩写

- BD:网桥域
- EPG:终端组
- ExEPG:外部终端组
- RIB:路由信息库
- VRF:虚拟路由和转发
- 类 ID:标识EPG的标记

外部EPG分类

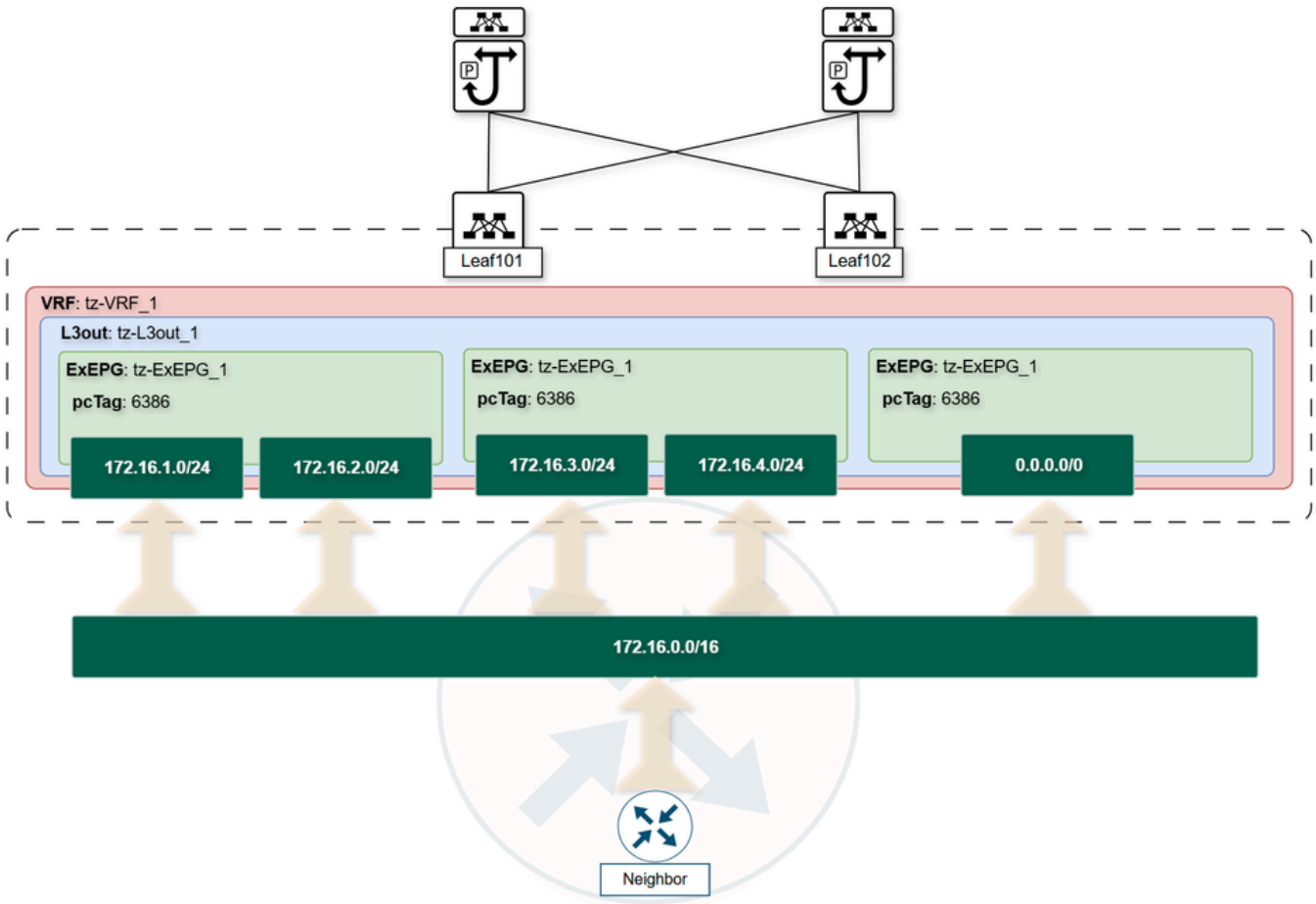
思科ACI中的外部EPG表示通过L3Outs连接的外部路由网络。类似于常规EPG对终端进行分类的方式，外部EPG根据每个VRF对外部子网进行分类，这意味着每个子网在其VRF环境中必须是唯一的。

一个常见的误解是外部EPG子网仅包含通过动态路由协议接受的前缀。但是，当创建L3Out时，有一个默认路由映射过滤传入通告；因此，默认情况下会接受动态路由协议通告的所有前缀。在ExEPG下定义子网的主要目的是分类，仅将唯一的pcTag分配给ExEPG中包含的子网，用于合同实施和策略应用。

此分类可实现精细的策略控制。例如，一个外部邻居可以向ACI通告一个超网，然后将其分段为多

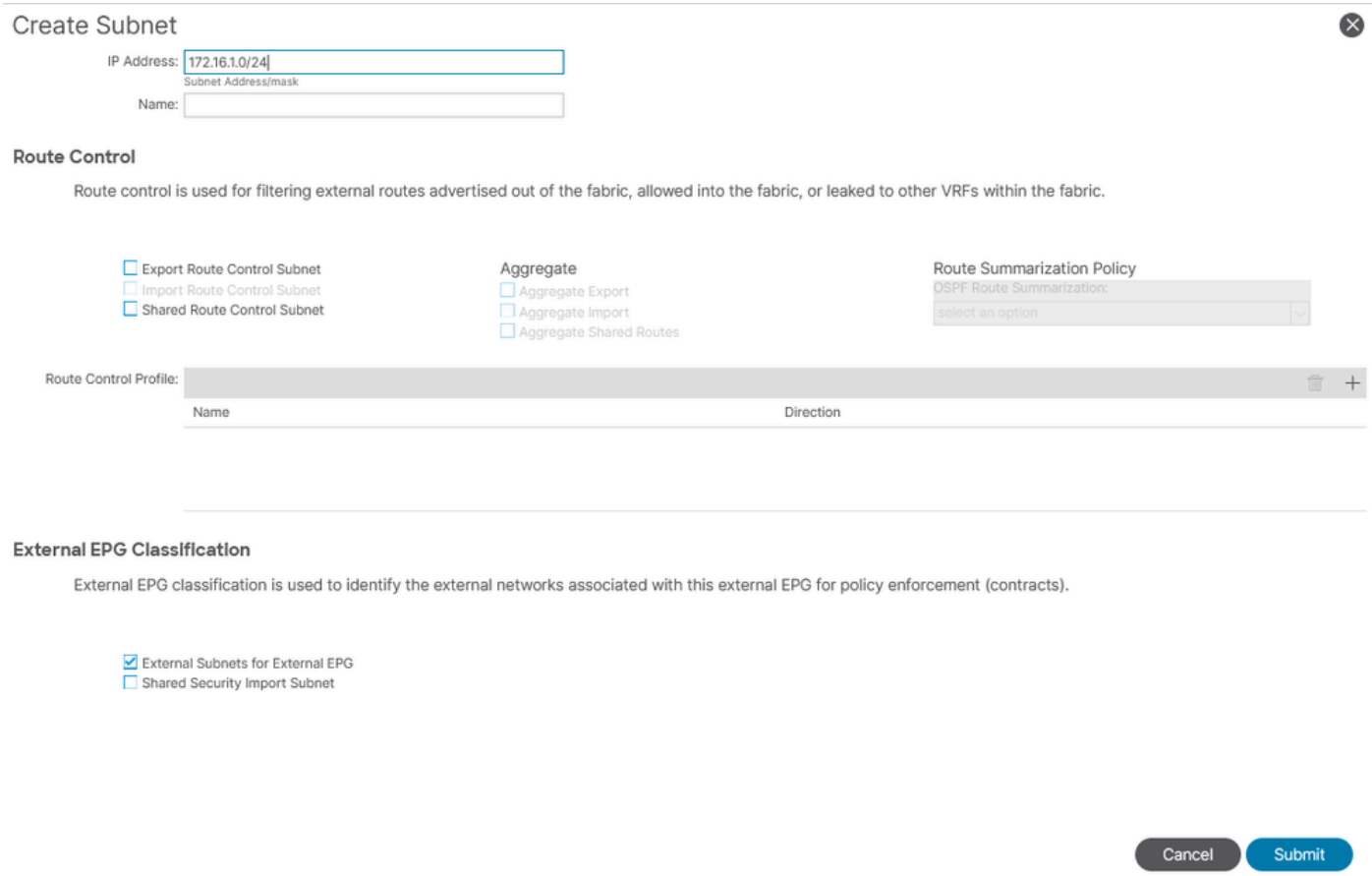
个ExEPG。这允许将不同的合同操作应用于不同的子网，例如允许特定内部EPG仅与指定的外部子网通信，或在到达最终目的地之前将指向特定前缀的流量重定向到PBR节点。

下图说明思科ACI如何根据外部EPG对外部子网进行分类，从而实现精确的流量分段和合同实施。



外部EPG子网标志

要在ACI中分类和管理ExEPG中的外部前缀，在ExEPG下创建子网前缀时配置特定子网标志。本节详细介绍每个标志及其预期用途：



- 外部EPG的外部子网：
  此标志表示子网位于ACI交换矩阵外部，且未在任何桥接域或EPG中配置。仅当前缀由路由邻居通告或静态注入RIB时，才必须使用该前缀。默认情况下启用此标志。

- 导出路由控制子网：
  此标志指定通过动态路由协议将子网从ACI通告给路由邻居。它不能与外部EPG的外部子网标志同时启用，因为这样做可能会导致第3层路由环路。由于ACI将子网分类为外部子网并通告给后端，因此尽管路由协议中存在环路避免机制，但这样做可能导致路由不一致。

- 共享路由控制子网：
  当子网前缀要在多个VRF之间共享时，会设置此标志，从而启用情景之间的路由渗透。

- 共享安全导入子网：
  与共享路由控制子网标志配合使用，可跨不同VRF共享外部子网的安全pcTags，从而促进一致的策略实施。

- 导入路由控制子网：
  此标记允许对从路由邻居接收的前缀进行精细控制。默认情况下，ACI接受所有传入路由通告；启用此标志需要激活路由控制实施以过滤传入的前缀。

- 聚合部分：
  此部分仅适用于quad-0(0.0.0.0/0)子网，它汇总了RIB中用于聚合导出或导入的所有前缀。当子网泄漏到其他VRF时，它们将汇总为聚合共享路由以优化路由表。

验证和故障排除命令

路由

首先，该路由必须存在于边界枝叶交换机上VRF的路由表中。例如，此命令显示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#
```

这确认该路由已安装在VRF路由表中并且可用于转发决策。

分类

路由出现在路由表后，分类会根据策略确定如何处理流量。在ACI中，分类与ExEPG及其关联子网关联。

要验证ExEPG下的子网分类，可以查询APIC以获取l3extInstP类（表示外部EPG实例）。其子类l3extSubnet列出在该ExEPG下配置的子网。例如：

```
<#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类的输出，则表明在外部EPG下未配置任何子网。如果没有配置子网，ACI无法将pcTag关联到传入流量子网，导致流量被丢弃，尽管路由表中存在路由。

需要注意的另一个重要方面是子网范围，这表示为有问题的子网设置的标志：

- 导入安全

该子网已标记为外部EPG的外部子网。

- export-rtctrl

该子网已标记有导出路由控制。

- import-rtctrl

该子网已使用Import Route Control标记。

- 共享安全

该子网已标记为共享安全导入子网。

- shared-rtctrl

该子网已使用共享路由控制进行标记。

路由协议和控制平面进程在收到来自所述邻居的前缀时更新路由表，然后将其编程到HAL L3转发表中。HAL L3路由表示编程到枝叶交换机上的硬件转发表(ASIC)中的实际第3层路由。这些路由源自路由协议和路由表计算，用于转发决策。

```
<#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#
```

随后，当在ExEPG中使用外部EPG的外部子网标志配置子网时，称为策略管理器(policy-mgr)的内部进程会使用此子网条目和关联的pcTag更新其前缀到pcTag的映射表。策略管理器用作交换矩阵集中式策略协调引擎，将高级策略定义转换为ACI交换矩阵中的可行配置。这通过基于配置的外部子网实施正确的流量分类和转发决策的pcTags，确保应用连接和网络行为的一致性和安全性。

<#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#
```

这确认前缀172.16.1.0/24正被邻居通告到ACI边界枝叶交换机，并且ACI已将前缀分类到pcTag32771

合同

分区规则是在交换矩阵内的EPG（包括ExEPG）之间实施合同策略的基本流程。外部EPG的VRF VNID（范围）和pcTag可用于定义和验证在源和目标EPG之间应用的通信规则。从本质上讲，分区规则将高级合同关系转换为在枝叶交换机上编程的特定的、可执行的规则。

需要考虑的一个重要方面是合同在交换矩阵中的安装位置。默认情况下，VRF配置策略控制实施方向设置为入口。此设置确定给定合同的分区规则安装在源终端所在的枝叶交换机上。

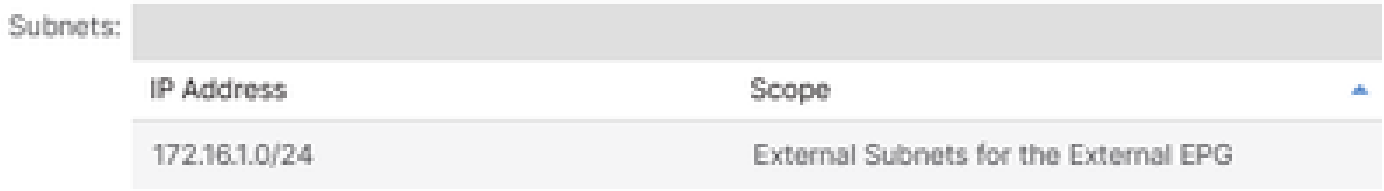
在本练习中，流量从L3Out进入，分区规则安装在连接到该L3Out的边界枝叶上，因为该枝叶充当进入交换矩阵的流量的源枝叶。

<#root>

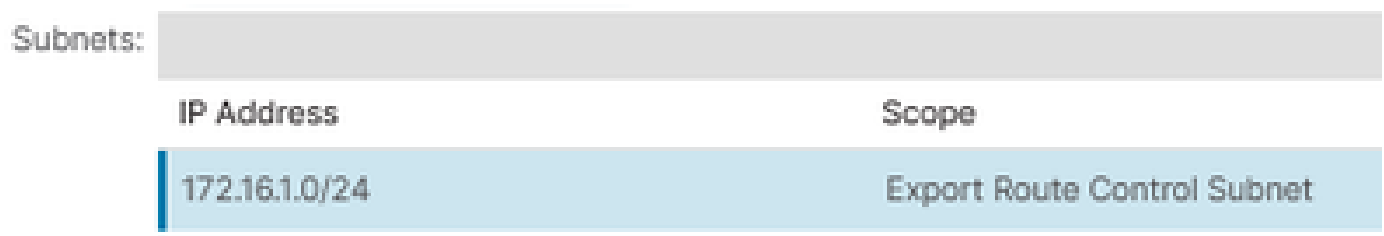
```
Leaf101#

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#
```

传输路由

中转路由使交换矩阵能够充当中转网络，方法是从一个L3Out获知的外部路由通告到另一个L3Out。要正确配置传输路由，传入的子网必须标有“外部EPG的外部子网”标志。



同时，向其他外部对等体通告此子网的L3Out必须在相应子网中启用导出路由控制子网标志。此标志允许通过该L3Out上配置的路由协议在交换矩阵外重新分发和通告子网。



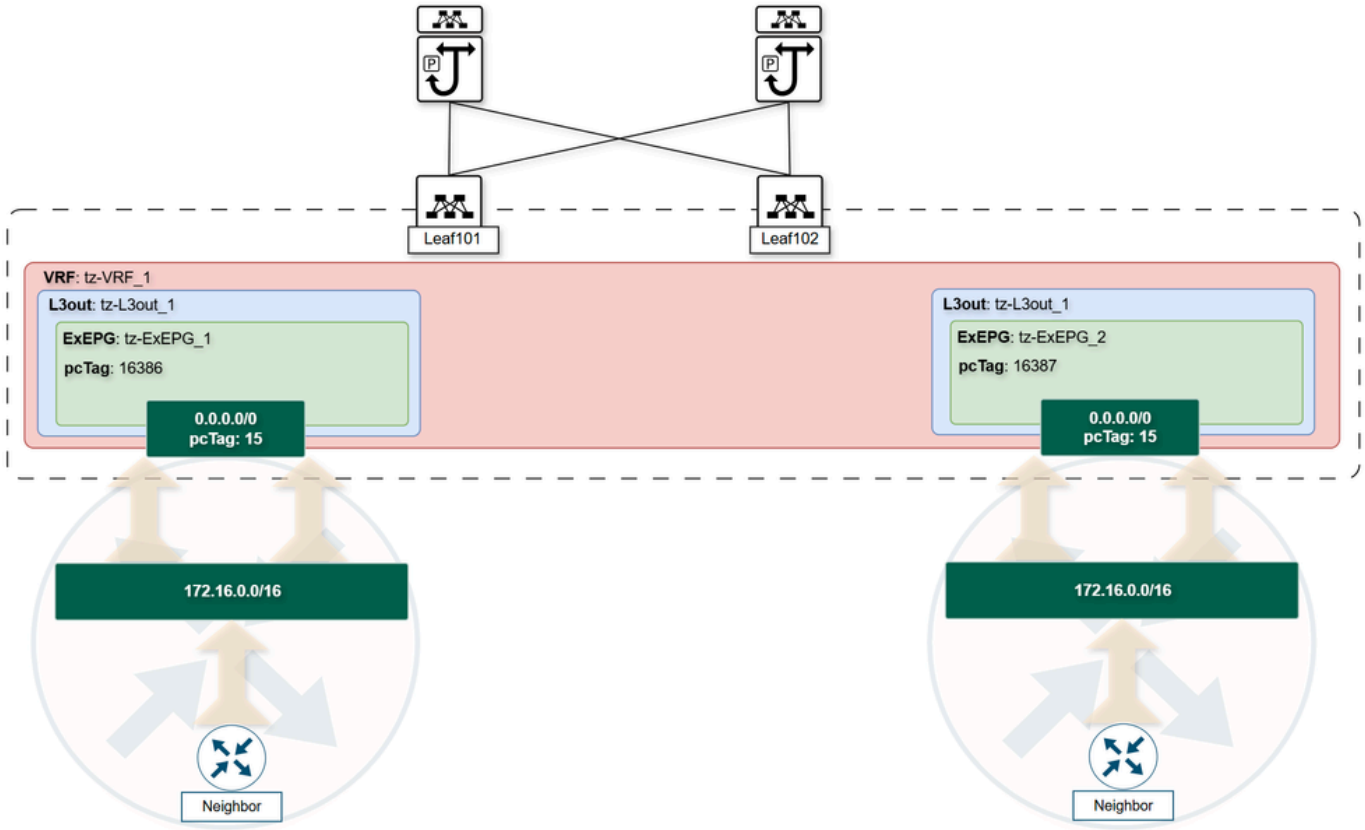
最后，需要配置接收和导出L3out之间的合同，以完成路由分发过程。

子网外部EPG分类中的常见问题

pcTag 15

以前，在本文档中，曾说过ExEPG子网出于策略实施原因帮助您将子网分类到正确的pcTag中。此分类的一个重要例外是使用“外部EPG的外部子网”标志配置的quad-0子网(0.0.0.0/0)。此子网始终分配有保留的pcTag 15，有效地用作VRF中所有外部流量的通配符。

下图表示在同一VRF内的多个ExEPG上配置带外部EPG的外部子网的quad-0所面临的问题：



- quad-0子网通常被误认为是默认路由。虽然有时确实如此（例如动态路由邻居仅向ACI L3Out通告默认路由时），但是quad-0子网在ACI中的角色更广泛，即是一种全捕获分类。
- 通常情况下，使用四0子网配置多个ExEPG以接受邻居通告的所有前缀。虽然这实现了广泛接受的目标，但是当在同一VRF中配置多个具有quad-0的ExEPG时，可能导致意外的非对称路由。当同一VRF中的多个ExEPG配置有quad-0作为外部子网时，ACI无法明确选择要用于特定目标子网的L3Out。而是任意选择一个L3Out。
- 如果随机选择的L3Out没有允许通信所需的合同，则此行为可能会导致非对称路由、流量间歇性甚至流量丢弃。

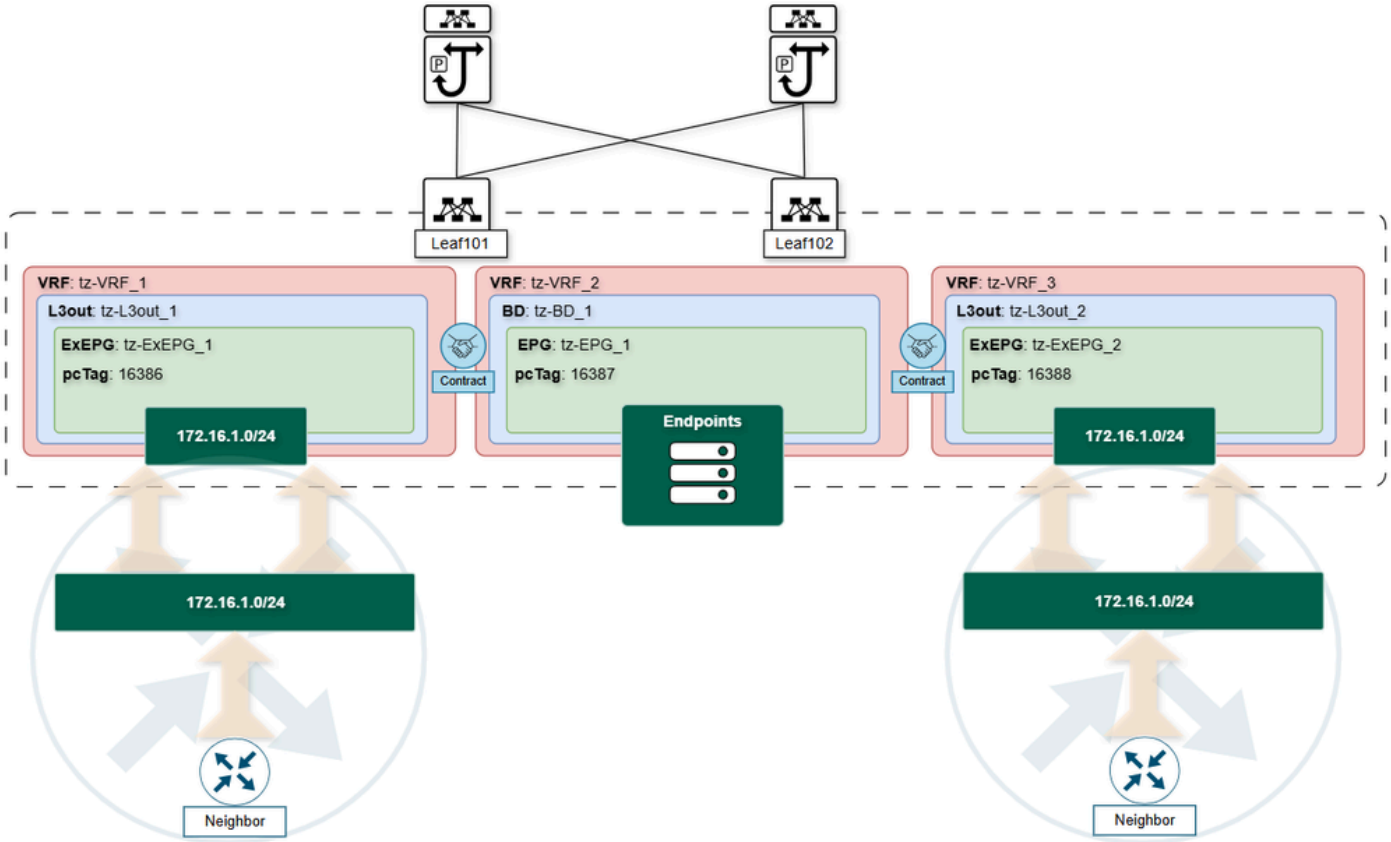
子网重叠

不允许在不同ExEPG之间配置相同的子网。尝试这样做会触发故障“F0467:Prefix Entry Already Used in Another EPG”，防止VRF中的子网重复。

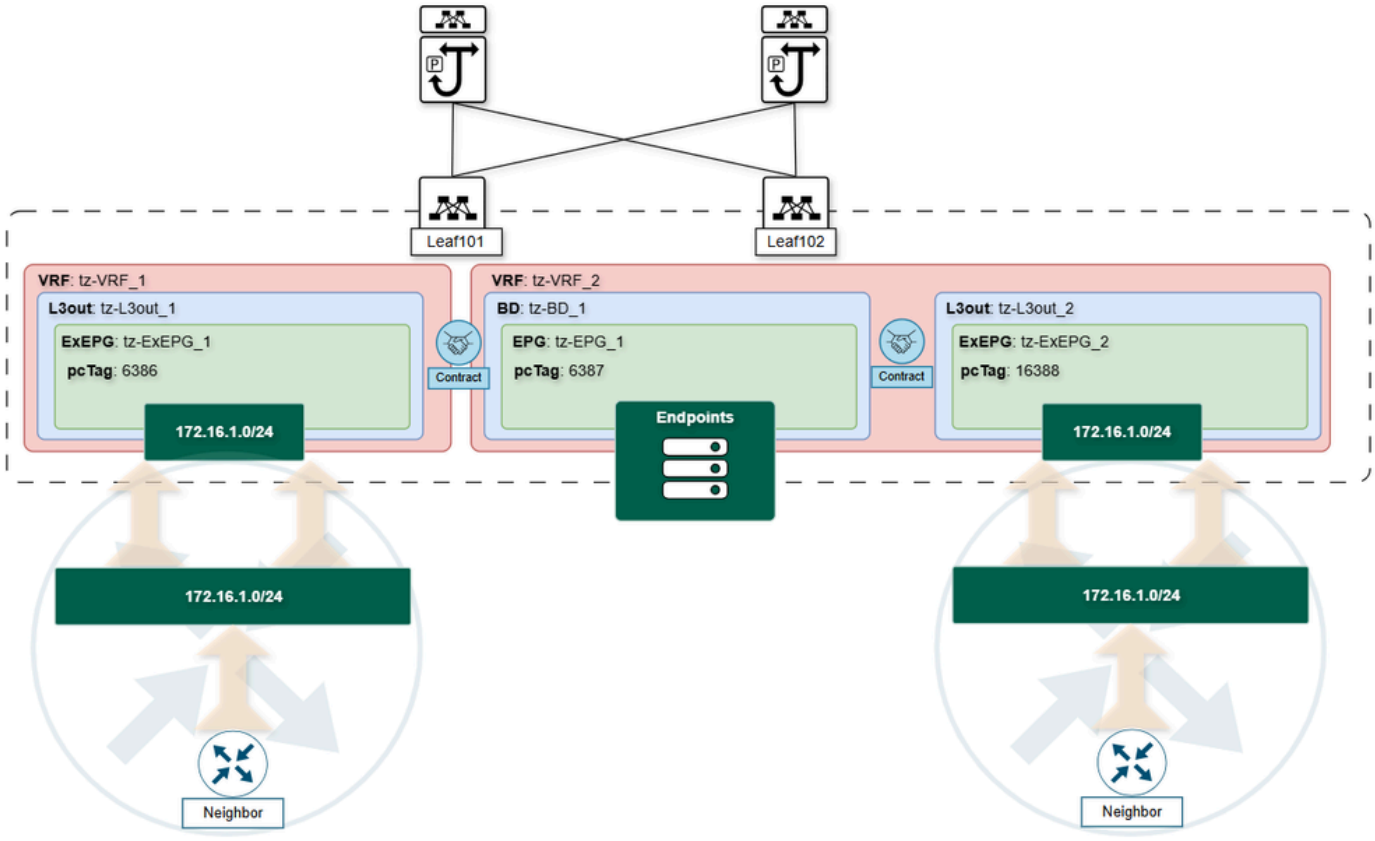
但是，由于每个VRF维护独立的路由表情景，因此不同VRF之间可能存在重叠的子网。这种分离允许在属于不同VRF的ExEPG中配置相同的子网。尽管如此，在涉及这些重叠子网的VRF路由泄漏时，谨慎至关重要，因为子网分类(pcTag)与路由信息(RIB)的冲突可能导致非对称转发决策。

主要场景包括：

- 从两个VRF到第三个VRF的路由泄漏：
  当两个VRF将同一子网泄漏到第三个VRF时，接收方VRF会根据APIC的共享策略安装接收到的第一个子网。如果处理此VRF的枝叶交换机重新启动或路由选择更改，路由表可能会更新到不同的L3Out，从而导致转发行为不一致。



- 与泄漏子网重叠的本地到VRF L3Out ExEPG:
  在使用路由泄漏的设计中，如果本地L3Out ExEPG配置了与泄漏子网相同的子网，则本地路由条目始终优先于泄漏的路由。



这些情况突出表明，非对称转发问题产生于分类和转发决策层，而不是路由表本身。虽然子网分类将子网与用于策略实施的特定L3Out和ExEPG相关联，但路由表可以指向不同的L3Out目标。这种不匹配会导致流量转发不一致，从而导致潜在的连接问题或策略实施差距。

导入路由控制默认行为更改

默认情况下，ACI接受来自邻居的所有传入路由通告。要控制接受的前缀，必须启用Route Control Enforcement:L3Out根对象上的入站：

导航到租户> [tenant name] > Networking > L3outs > [L3out name]。

此操作在所选路由协议下创建路由映射。

<#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#

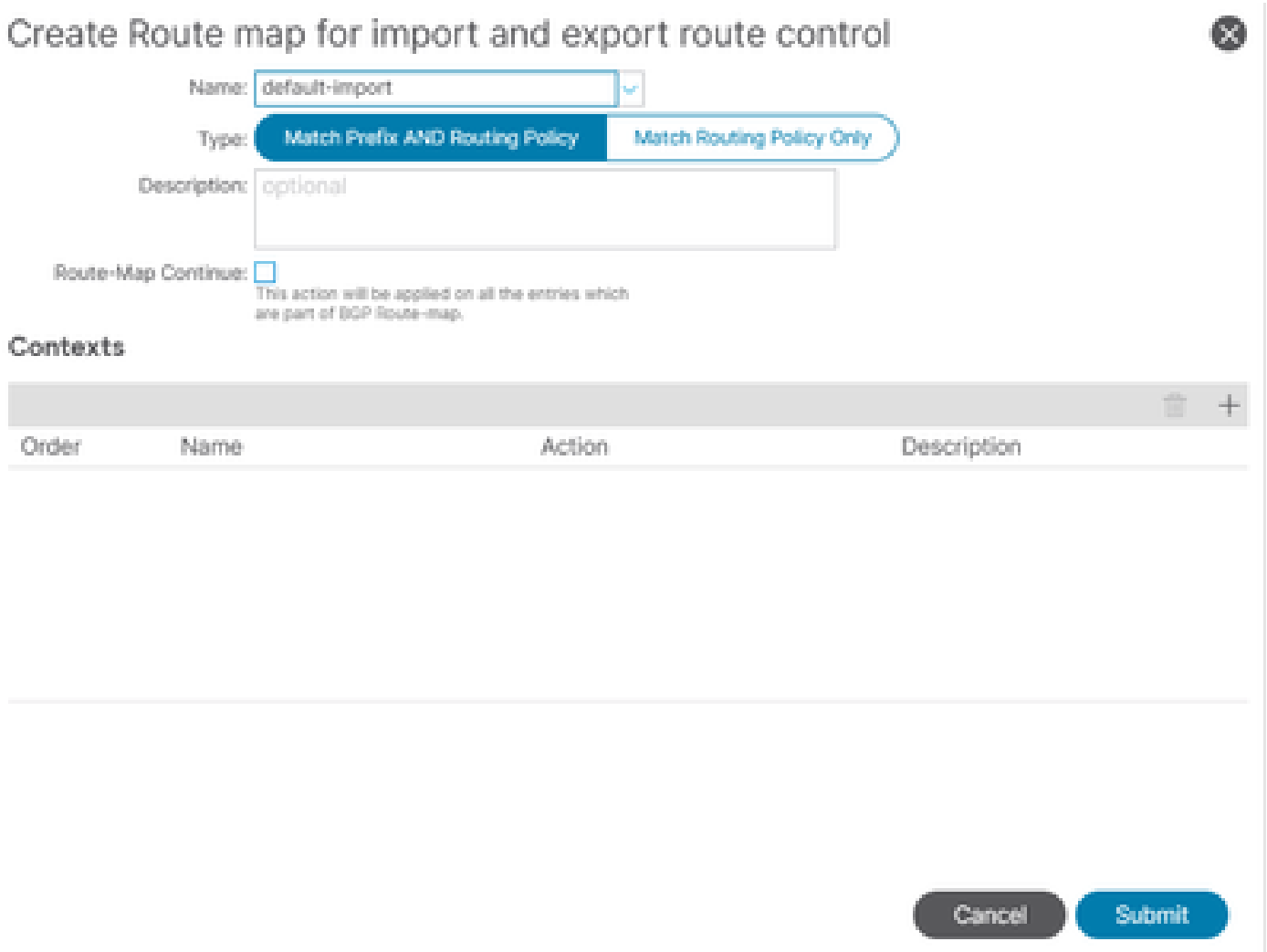
show ip prefix-list IPv4-peer49155-2981888-exc-ext-inferred-import-dst

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

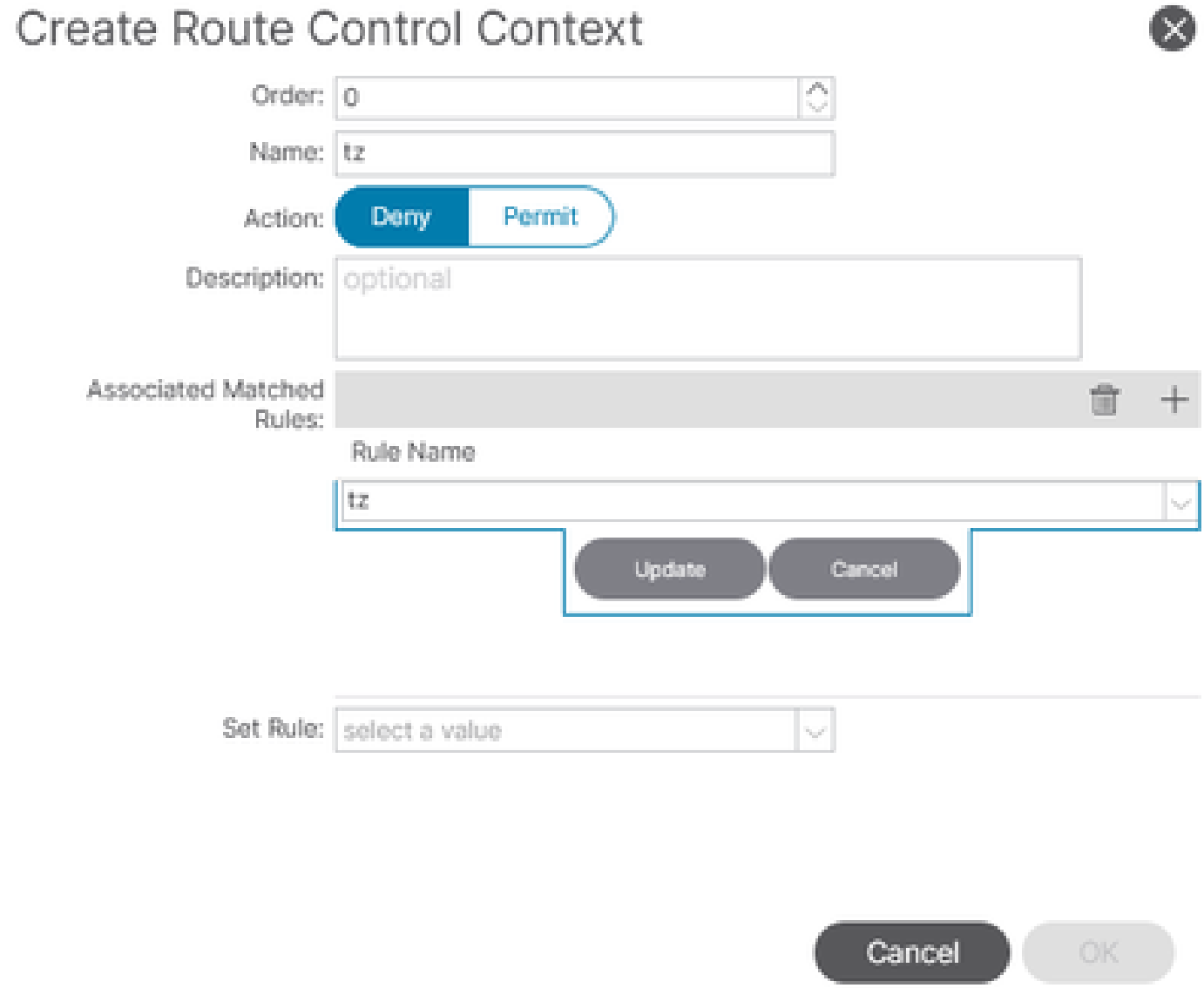
默认情况下，此导入路由映射允许所有传入前缀。要修改此行为，请执行以下操作：

导航到租户> [tenant name] >网络> L3outs > [L3out name] >用于导入和导出路由控制的路由映射

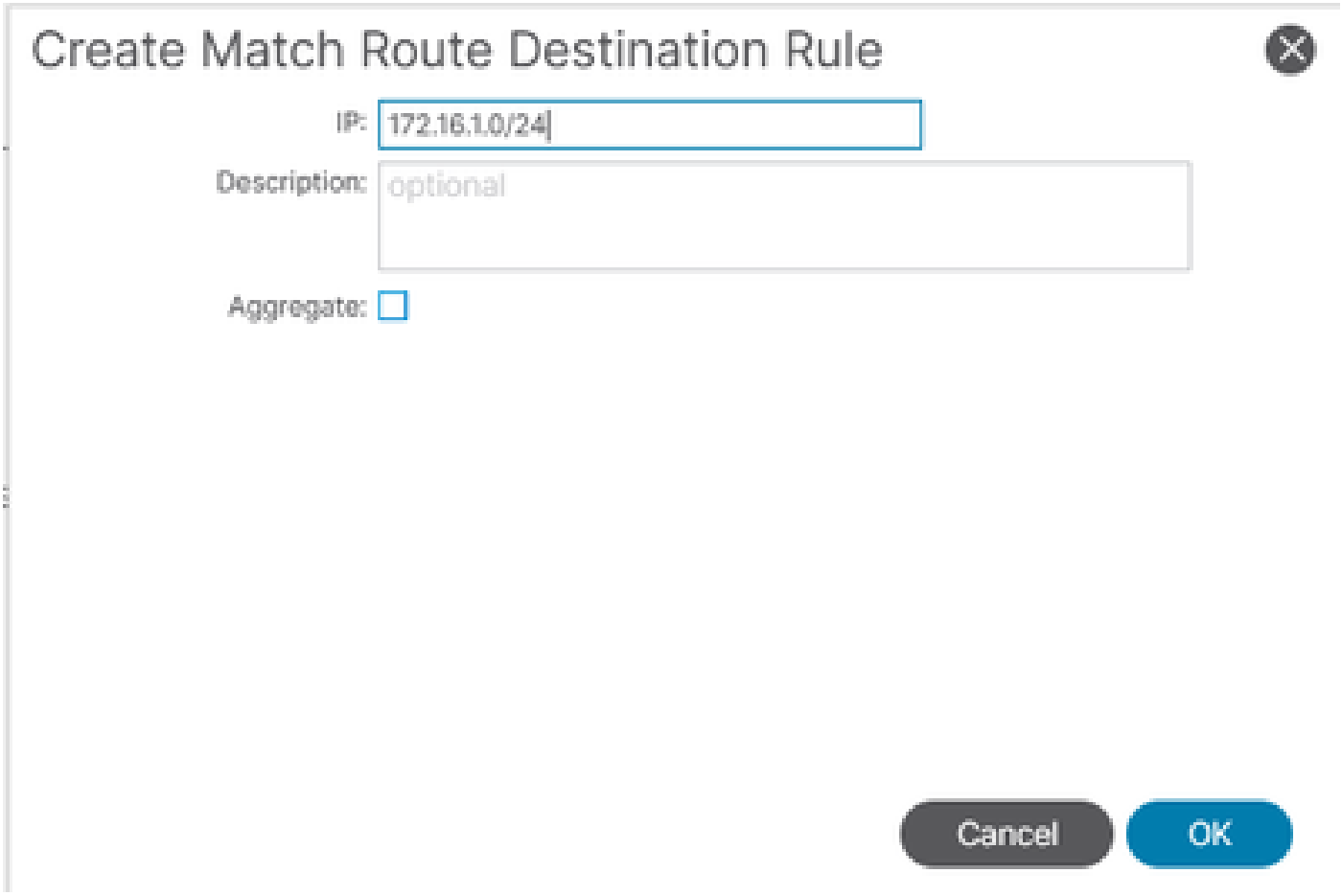
选择默认导入路由映射，或使用右上角的齿轮图标创建一个新的导入路由映射。



在Context部分中，创建一个新的Associated Matched Rule。



在Match Rules部分，滚动到Match Prefix，并添加要控制的特定子网。



提交策略后，导入路由映射操作会相应地更改，从而实施所需的前缀过滤。

<#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#
```

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）。