# 修复ACI交换矩阵的EVPN RMAC ExtCommunity传输问题

## 目录



## 简介

本文档介绍从外部边界网关协议(BGP)对等体接收时，配置错误的路由器MAC扩展社区属性对ACI交换矩阵的影响。

## 背景信息

使用BGP时，可以选择发送带有通告给BGP对等体的前缀的社区和扩展社区属性。这些社区属性允许我们修改路由策略并动态改变路由流量的处理方式。

## 问题

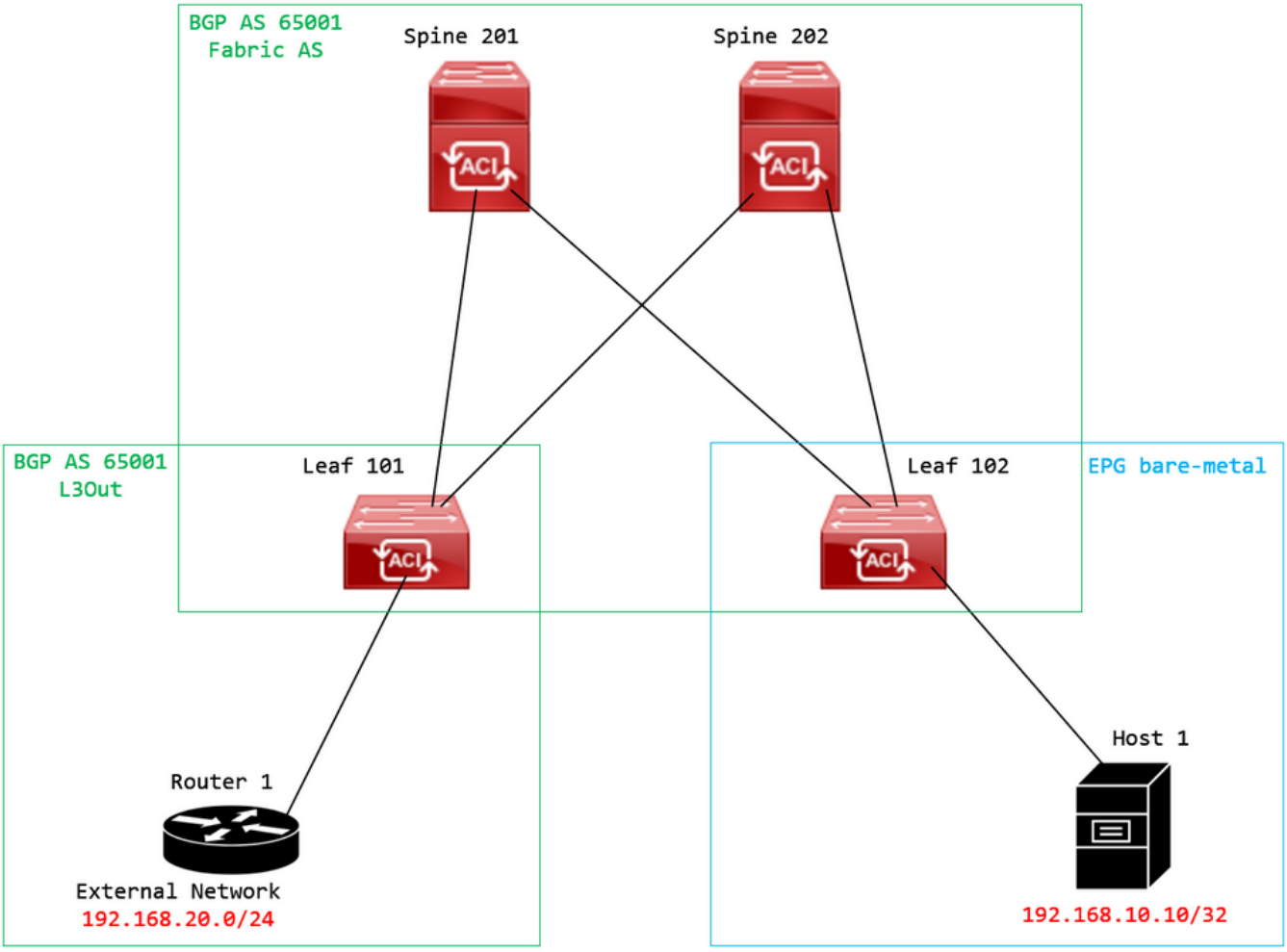
当从外部BGP对等体向ACI交换矩阵发送带有IPv4 AFI前缀的路由器MAC扩展社区属性时，交换矩阵中通过内部MP-BGP进程接收来自边界枝叶的路由的任何枝叶上都会发生FIB和HAL编程错误。这是因为RMAC extcommunity属性属于BGP L2VPN EVPN地址系列，当将其注入BGP IPv4地址系列时，该属性会被拒绝。这是由于违反了规则5.2(Uniform-Propagation-Mode),IETF文档标题为“EVPN与IPVPN互通”中对此进行了描述。第15页项目4c指出了具体问题：

```
4.  As discussed, Communities, Extended Communities and Large
      Communities SHOULD be kept by the gateway PE from the originating
      SAFI route.  Exceptions of Extended Communities that SHOULD NOT
      be kept are:

      C.  All the extended communities of type EVPN.

      The gateway PE SHOULD NOT copy the above extended communities
      from the originating ISF route to the re-advertised ISF route.
```

链接至文档：EVPN与IPVPN互通

以下是iBGP问题的示例，但是eBGP也存在此问题。

拓扑图：




拓扑图

在外部BGP对等设备（路由器1）上配置路由映射并设置EVPN RMAC extcommunity属性：

```
Router-1# show run | sec route-map
route-map RMAC permit 10
  set extcommunity evpn rmac aaaa.bbbb.cccc
```

在BGP邻居IPv4地址系列配置下，配置BGP扩展社区，并在出站方向配置路由映射：

<#root>

```
Router-1# show run bgp
<output omitted>
feature bgp

router bgp 65001
  vrf example
    router-id 192.168.20.20
    address-family ipv4 unicast
      network 192.168.20.0/24
```

```
    neighbor 192.168.30.30
      remote-as 65001
      update-source loopback1

address-family ipv4 unicast



send-community extended


        route-map RMAC out
```

检查BL 101上的BGP状态：

<#root>

```
leaf-101# show ip bgp 192.168.20.0 vrf example:example
BGP routing table information for VRF example:example, address family IPv4 Unicast
BGP routing table entry for 192.168.20.0/24, version 40 dest ptr 0xa0fec840
Paths: (1 available, best #1)
Flags: (0x80c001a 00000000) on xmit-list, is in urib, is best urib route, is in HW, exported
  vpn: version 2725, (0x100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type (0xa96485b8): internal 0x18 0x0 ref 0 adv path ref 2, path is valid, is best path
  AS-Path: NONE, path sourced internal to AS
    192.168.20.20 (metric 5) from 192.168.20.20 (192.168.20.20)
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0
      Extcommunity:
          RT:65001:2162688
          COST:pre-bestpath:163:1879048192


Router MAC:aaaa.bbbb.cccc



***Notice that the router mac is present here.***


          VNID:2162688

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

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    10.0.216.65        10.0.216.66
```

检查CL 102上的RIB:

<#root>

```
leaf-102# show ip route 192.168.20.0 vrf example:example
IP Route Table for VRF "example:example"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>
```

**192.168.20.0/24**

```
, ubest/mbest: 1/0
    *via
```

**10.0.210.70**

```
%overlay-1, [200/0], 00:00:43, bgp-65001, internal, tag 65001,
```

**rwVnid: vxlan-2162688**

```
         recursive next hop: 10.0.210.70/32%overlay-1
```

***Notice that we have the route here and our next-hop address is correct (showing the TEP IP of BL 101)***

```
leaf-102# acidiag fnvread | grep 101
     101        1             leaf-101      <output omitted>
```

**10.0.210.70/32**

```
    leaf         active   0
```

检查CL 102上的FIB:

<#root>

```
module-1(DBG-elam-insel6)# show forwarding route 192.168.20.0 vrf example:example
ERROR: no longest match in IPv4 table 0xf5df36b0
```

***No entry is present.***

检查CL 102上的HAL表：

<#root>

```
module-1(DBG-elam-insel6)# show platform internal hal l3 routes | grep 192.168.20.0
```

***No entry is present.***

从EP（主机1）对来自外部BGP对等体(192.168.20.20)的外部网络中的主机执行ping操作：

<#root>

```
Host-1# ping 192.168.20.20 vrf example
PING 192.168.20.20 (192.168.20.20): 56 data bytes
Request 0 timed out
Request 1 timed out
Request 2 timed out
Request 3 timed out
Request 4 timed out

--- 192.168.20.20 ping statistics ---
5 packets transmitted, 0 packets received, 100.00% packet loss
```

******No connectivity.******

检查CL 102上的ELAM:

<#root>

```
leaf-102# vsh_lc
module-1# debug platform internal roc elam asic 0
module-1(DBG-elam)# trigger reset
module-1(DBG-elam)# trigger init in-select 6 out-select 0
module-1(DBG-elam-insel6)# set outer ipv4 src_ip 192.168.10.10 dst_ip 192.168.20.20
module-1(DBG-elam-insel6)# start
module-1(DBG-elam-insel6)# stat
 ELAM STATUS
===========
Asic 0 Slice 0 Status Armed
Asic 0 Slice 1 Status Triggered

module-1(DBG-elam-insel6)# ereport
Python available. Continue ELAM decode with LC Pkg
 ELAM REPORT
<output omitted>
--------------------------------------------------------------------------------------------------------------
Lookup Drop
--------------------------------------------------------------------------------------------------------------
LU drop reason                          :
```

**UC_PC_CFG_TABLE_DROP**

******Notice the drop vector here.******

# 解决方案

解决方案是停止将具有IPv4地址系列前缀的路由器MAC扩展社区属性从外部BGP对等体发送到ACI交换矩阵。

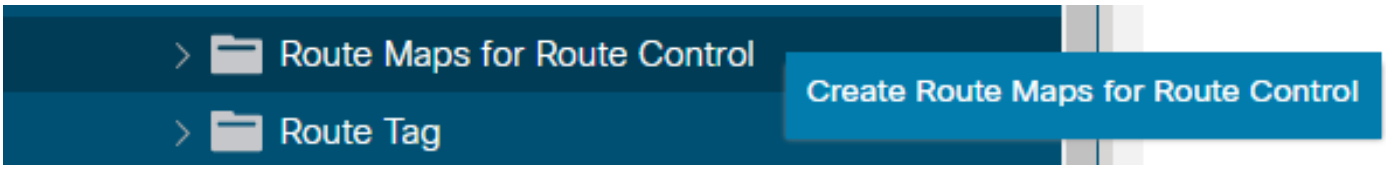
删除之前配置的路由映射并停止从外部BGP对等设备（路由器1）发送扩展社区。删除其中任一配

置或同时删除这两者均会起作用：

```
Router-1# show run bgp

feature bgp

router bgp 65001
  vrf example
    router-id 192.168.20.20
    address-family ipv4 unicast
      network 192.168.20.0/24
    neighbor 192.168.30.30
      remote-as 65001
      update-source loopback1
      address-family ipv4 unicast
```

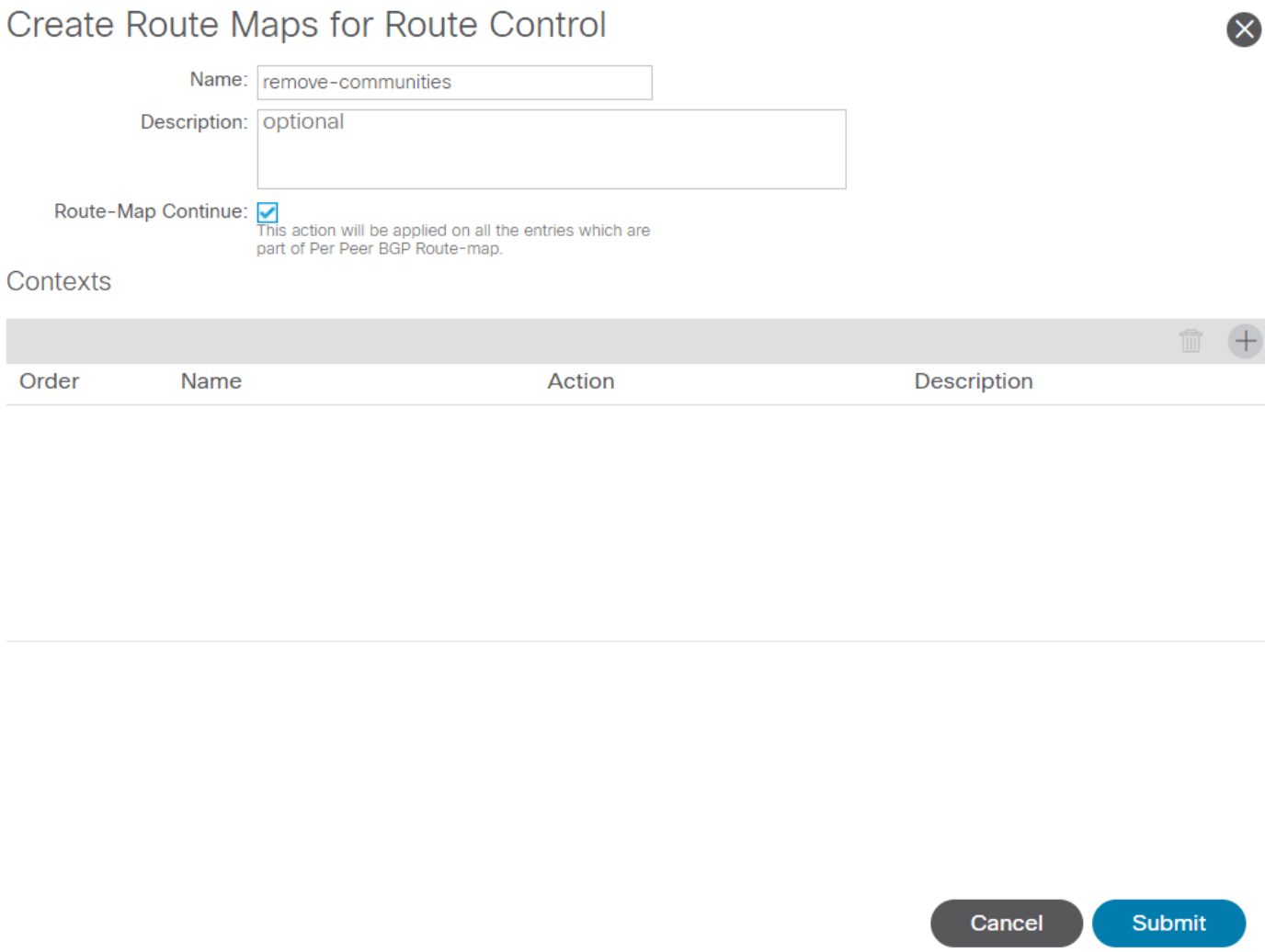
另一个（不太偏好）解决方案是，通过在ACI中配置的L3Out中创建路由映射，简单过滤从外部BGP对等设备接收的所有社区。

导航至 Tenant > Policies > Protocol > Route Maps for Route Control > Create Route Maps for Route Control：



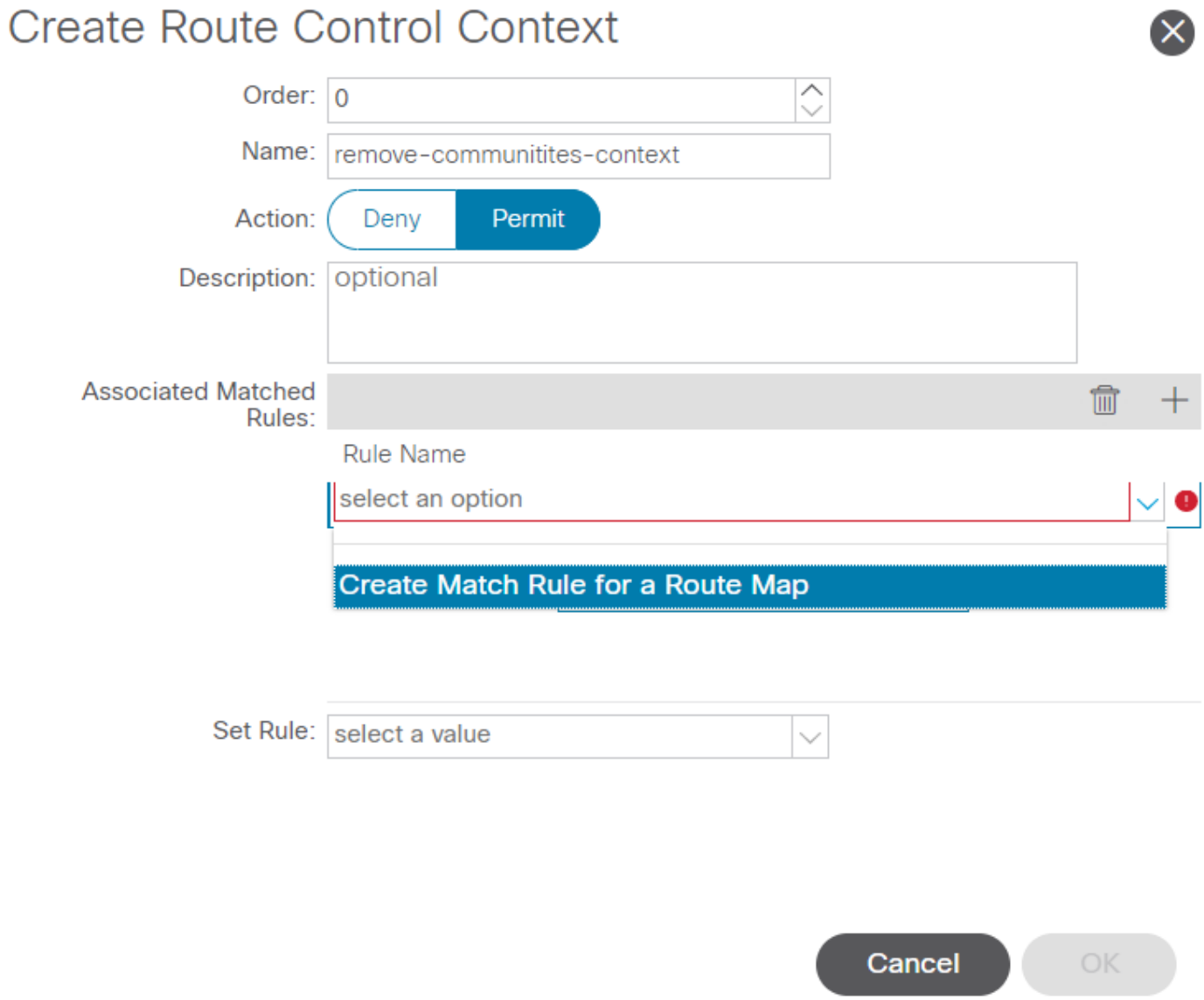

选择为路由控制创建路由映射的选项

命名路由映射，启用 Route-Map Continue 选项，然后添加情景。选择 + Contexts表中的图标：



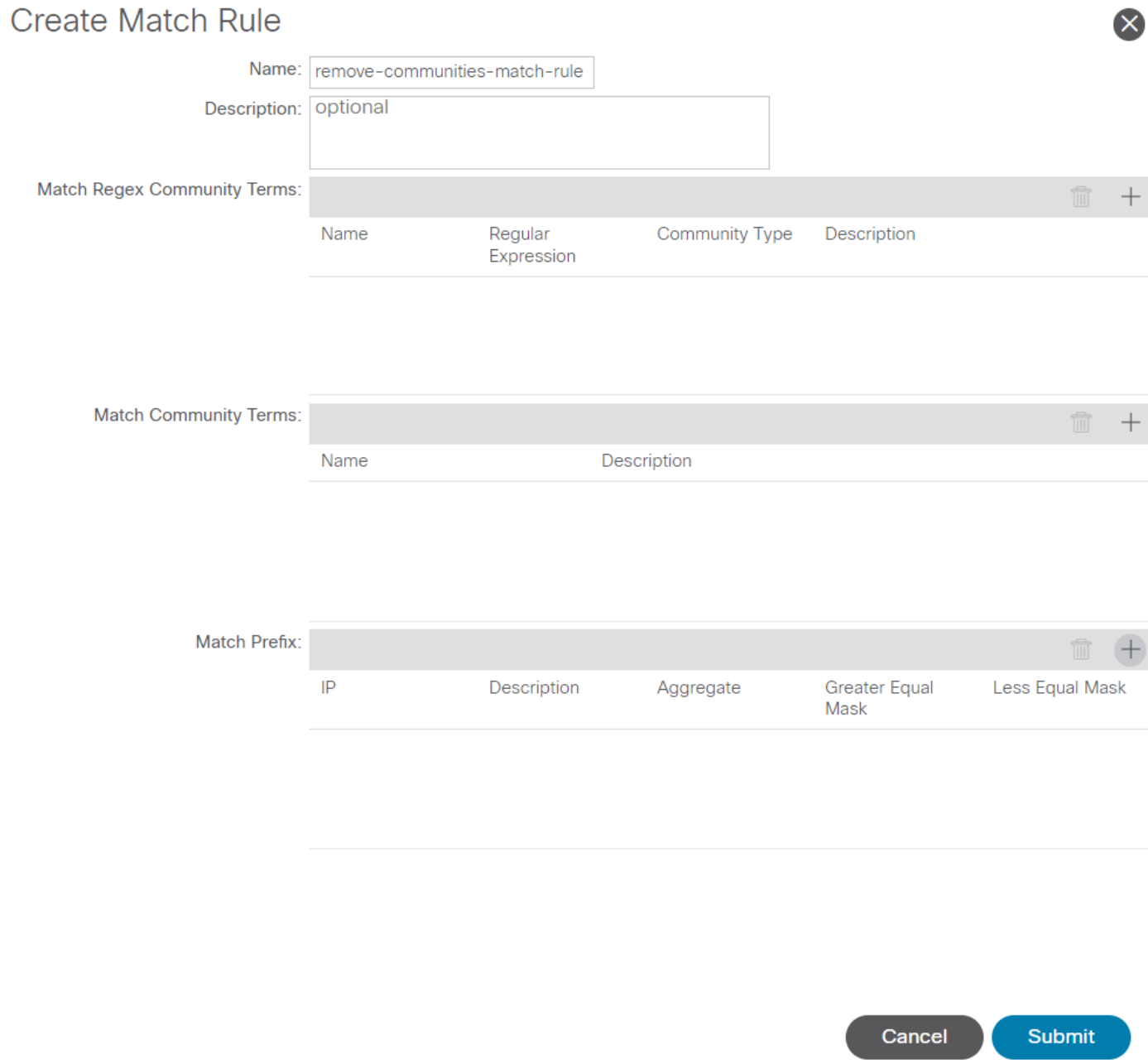

创建路由映射和创建情景

为您的情景命名，并保留默认操作 Permit 选中，然后通过选择 + 图标 Associated Matched Rules 表，然后选择 **Create Match Rule for a Route Map**:



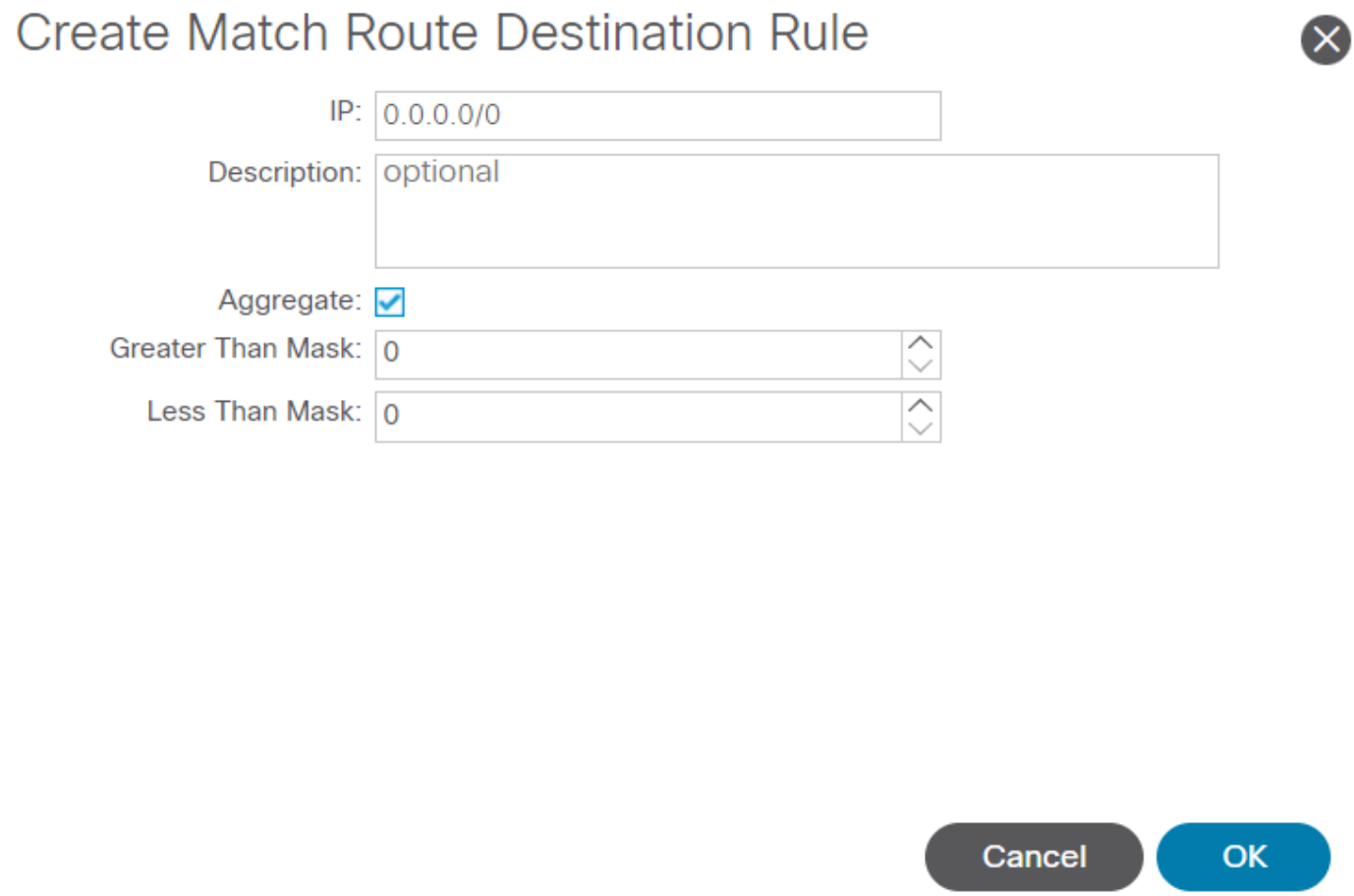

创建路由控制情景并为创建路由映射匹配规则选择选项

为您的匹配规则命名，然后通过选择中的+图标添加新前缀 Match Prefix 表：



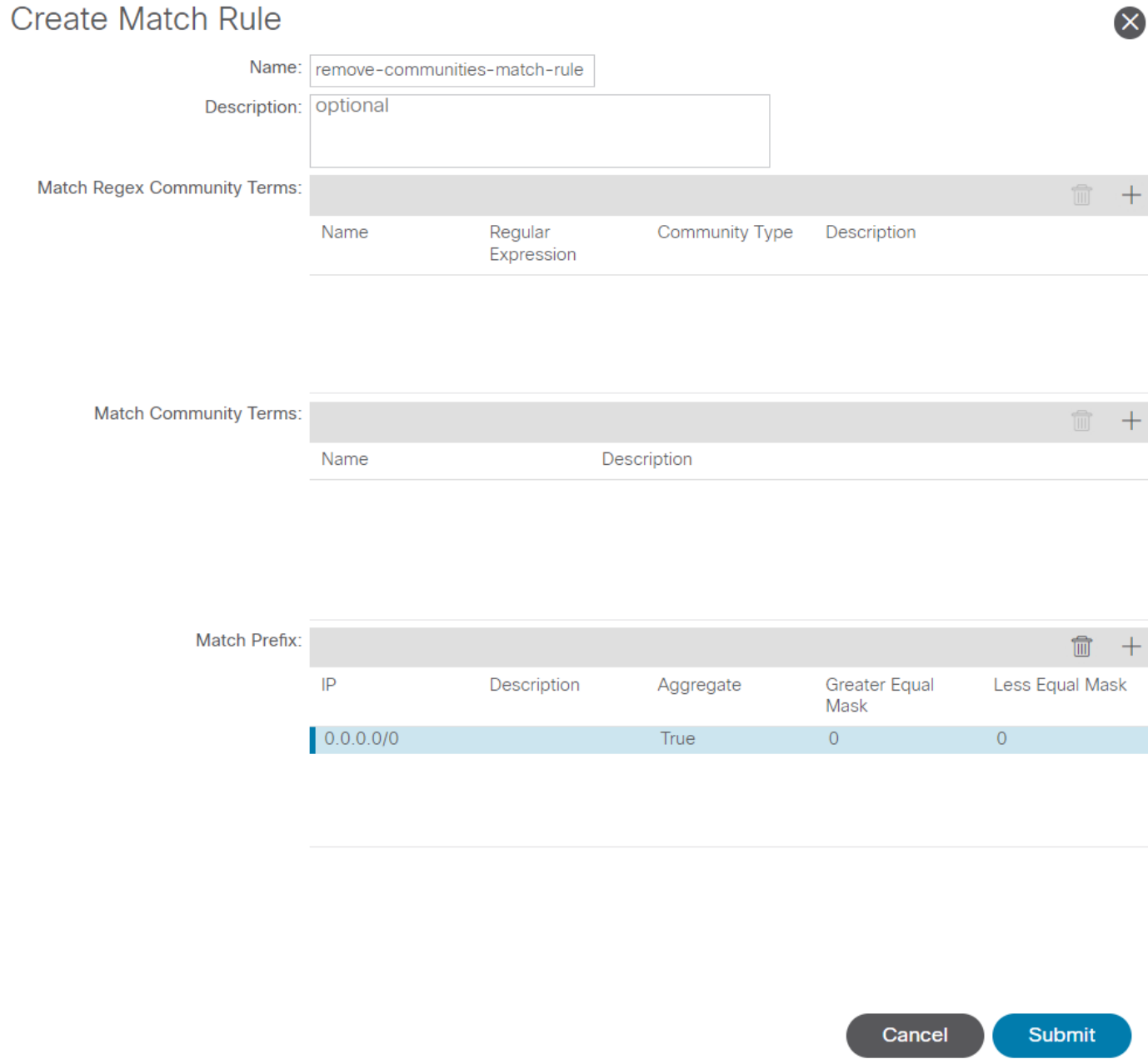

创建匹配规则并创建匹配前缀

添加所需前缀。此示例显示如何添加所有前缀的聚合：



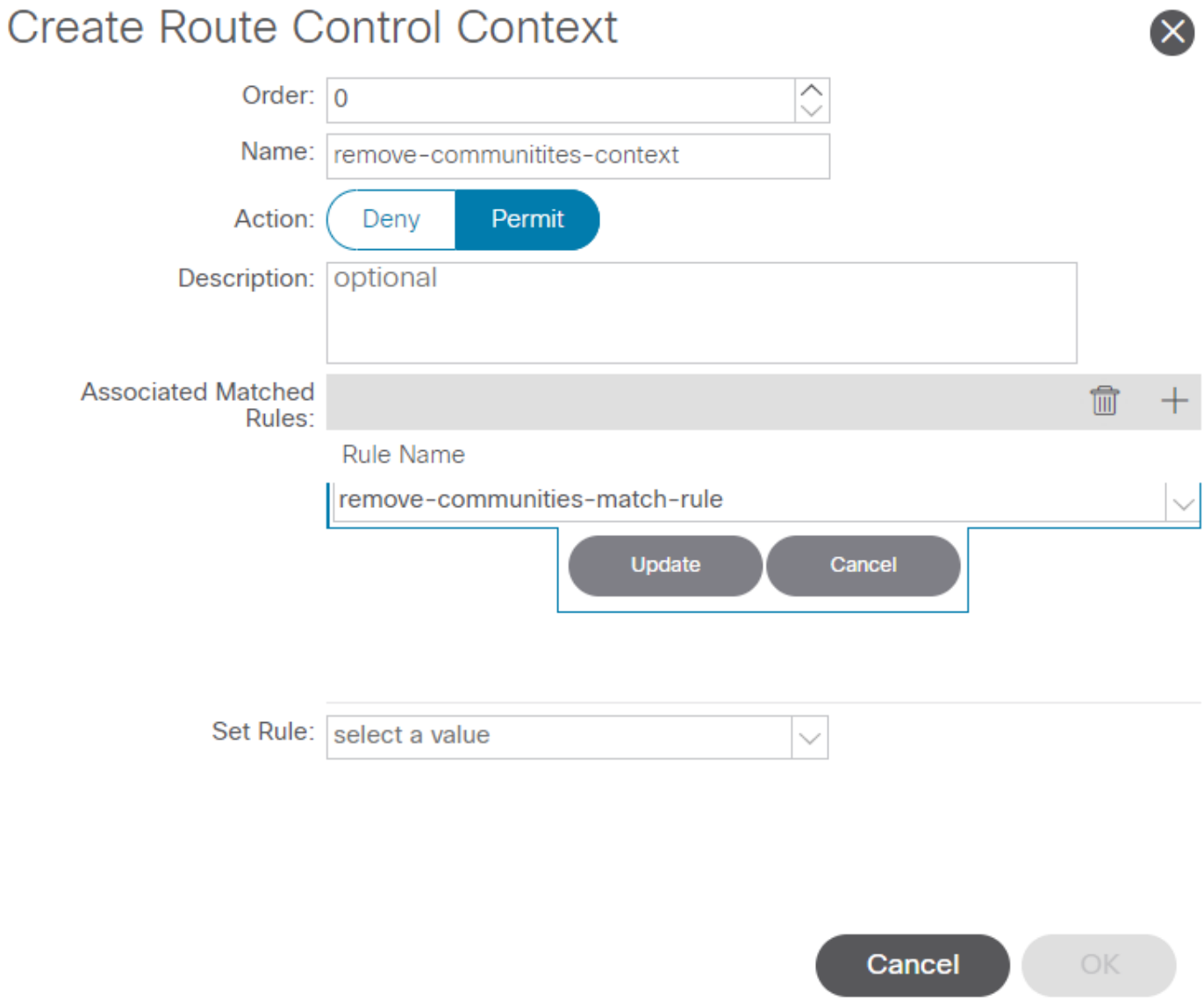

创建匹配路由目标规则

选择后 **OK** 如果 Create Match Route Destination Rule 窗口中，您会看到您的前缀已添加到 Match Prefix 中的表 Create Match Rule 窗口：



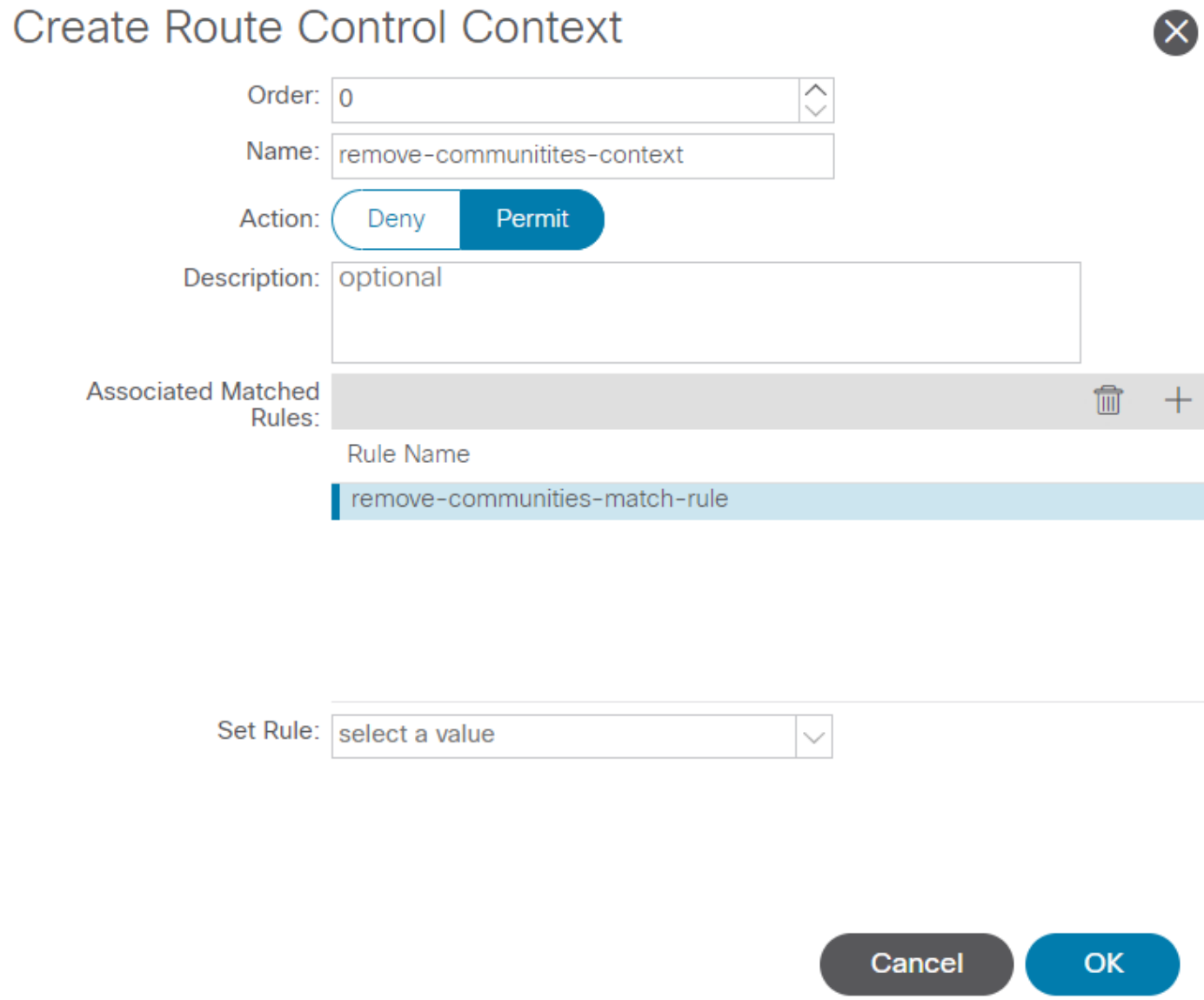

Match Prefix现已添加到Match Rule

选择后 Submit 如果 Create Match Rule 窗口，选择 Update 如果 Associated Matched Rules 中的表 Create Route Control Context 窗口：



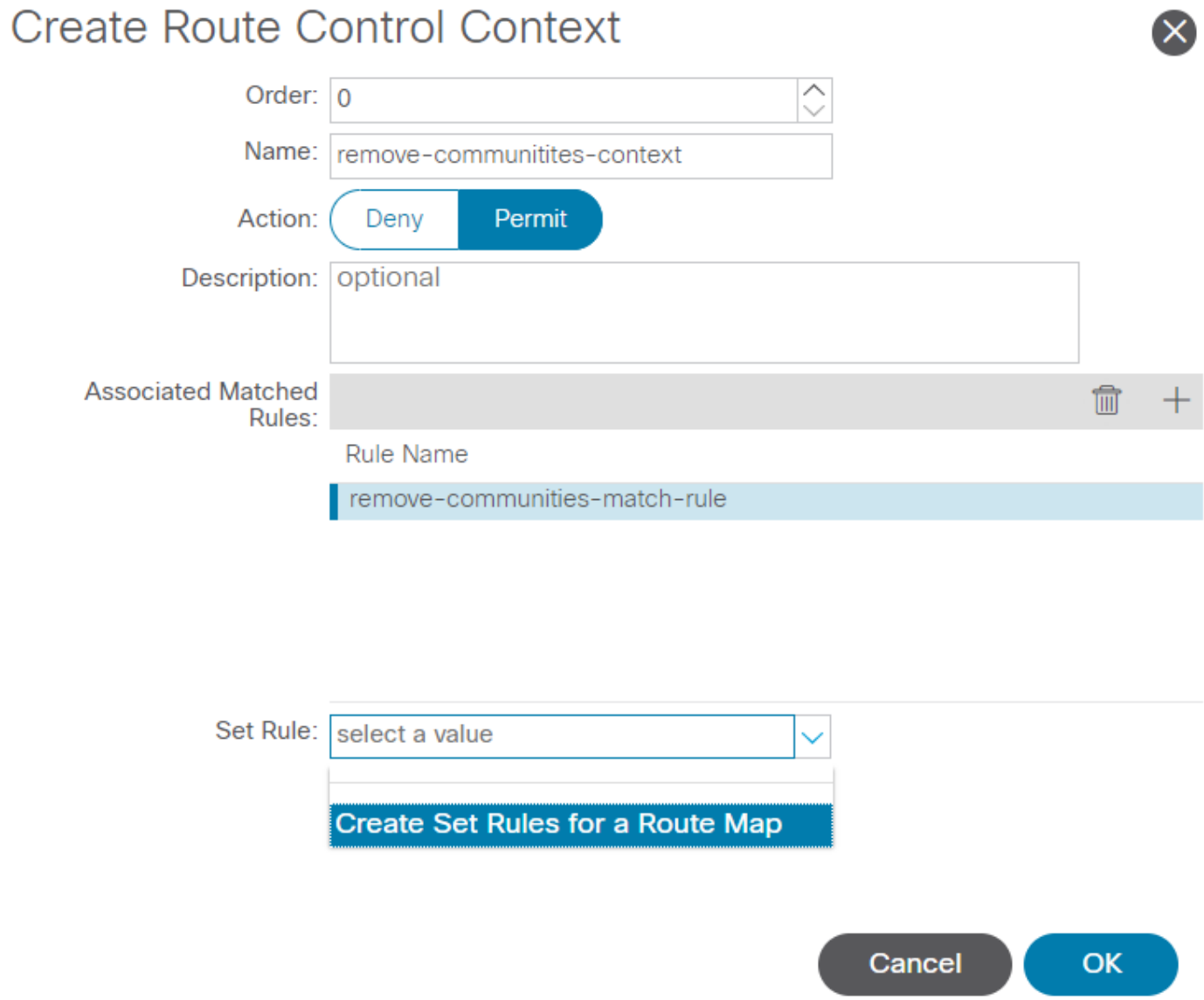

将关联的匹配规则添加到路由控制情景

关联匹配规则现在已添加到您的上下文：



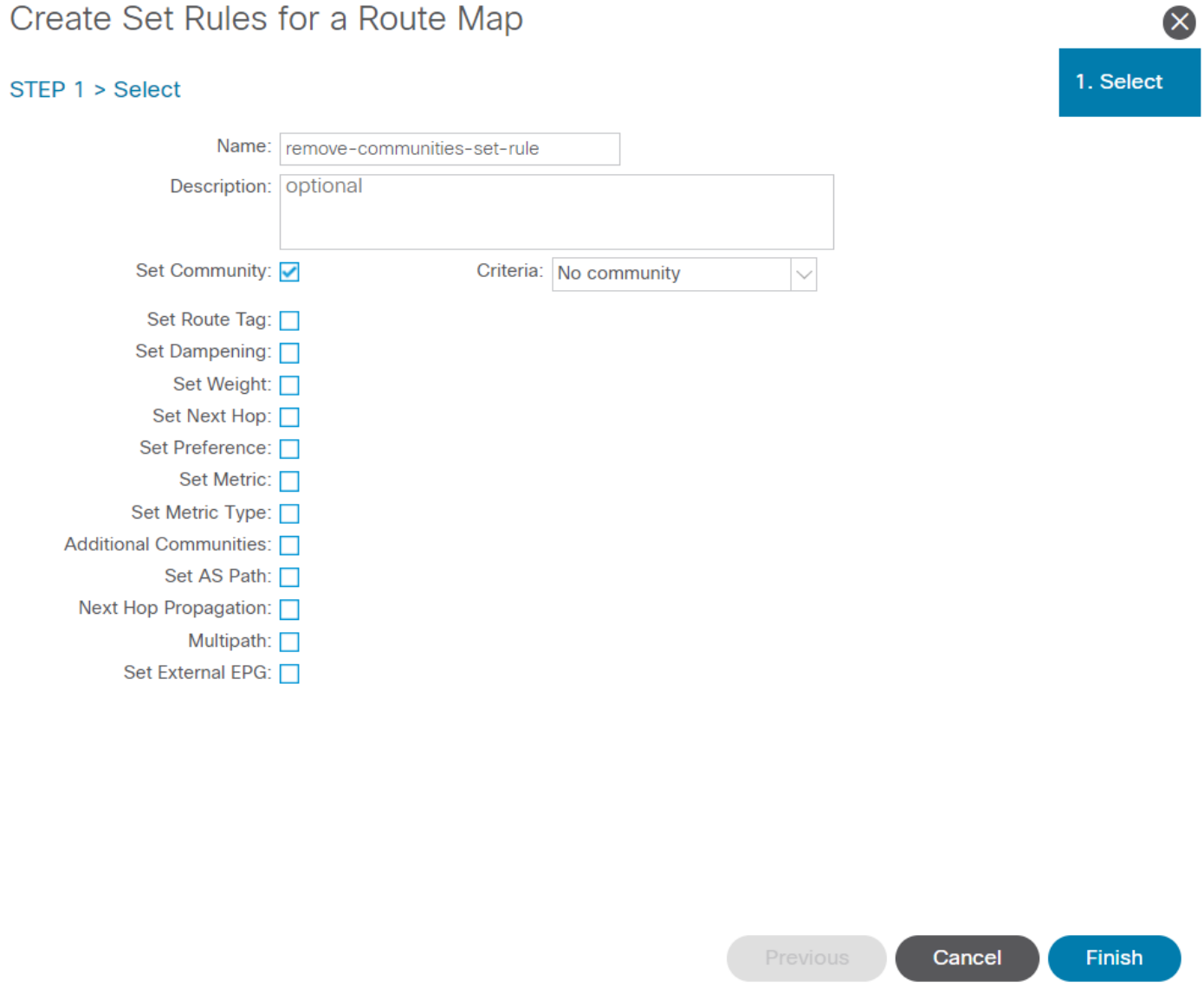

关联匹配规则现在已添加到路由控制情景

然后，选择旁边的下拉菜单 Set Rule 并选择 Create Set Rules for a Route Map:



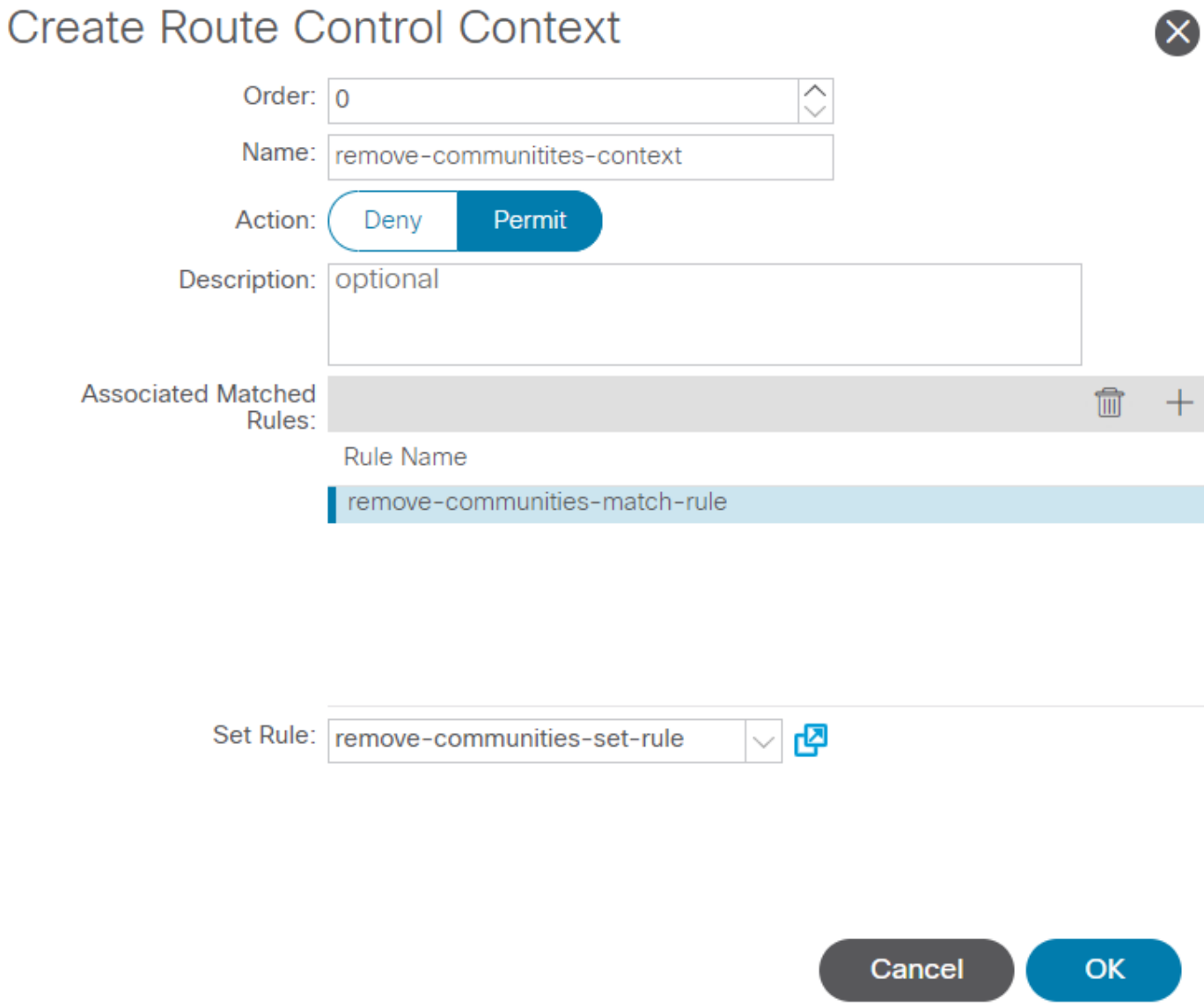

选择选项为路由映射创建集规则

为您的设置规则命名，然后选择 Set Community 选项并保留默认条件 No community 已选择：



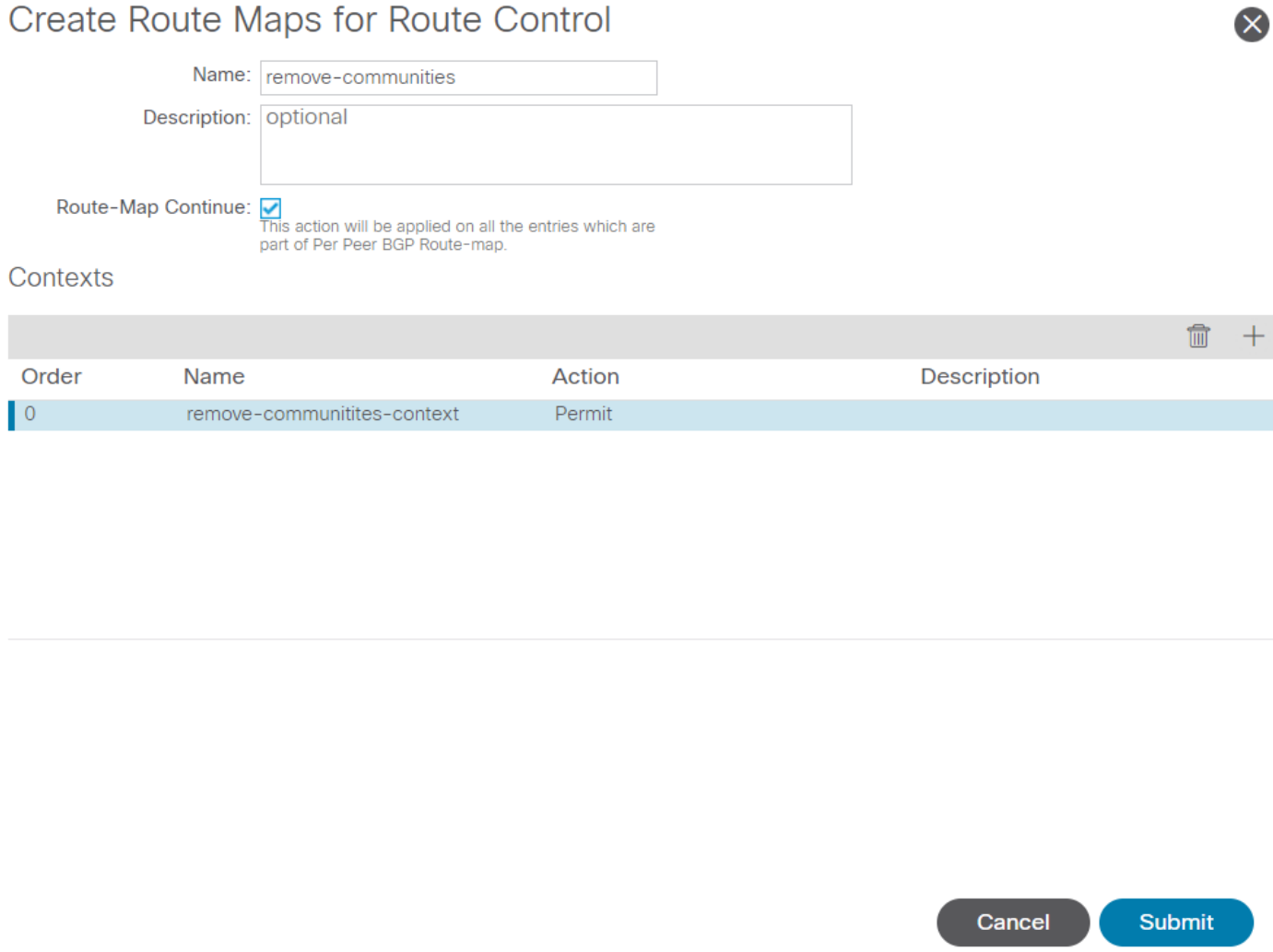

创建路由映射的设置规则

在 Create Set Rules for a Route Map 窗口中，您会看到您在 Create Route Control Context 窗口:



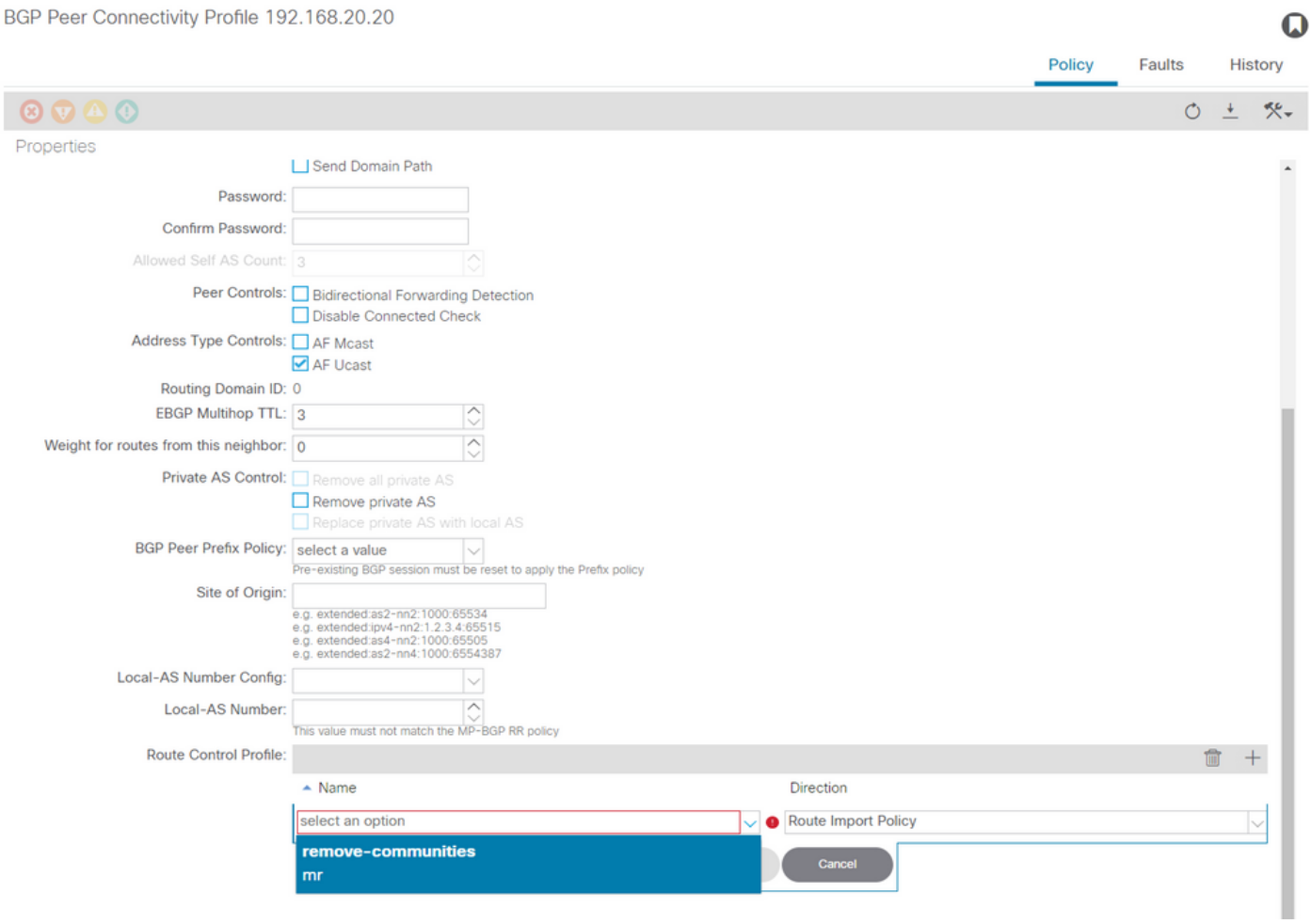

Set Rule现在已添加到路由控制上下文

选择后 OK 如果 Create Route Control Context 窗口，您会看到您的上下文已添加到 Contexts 中的表 Create Route Maps for Route Control 窗口.最后，选择 Submit 要完成配置，请执行以下操作：



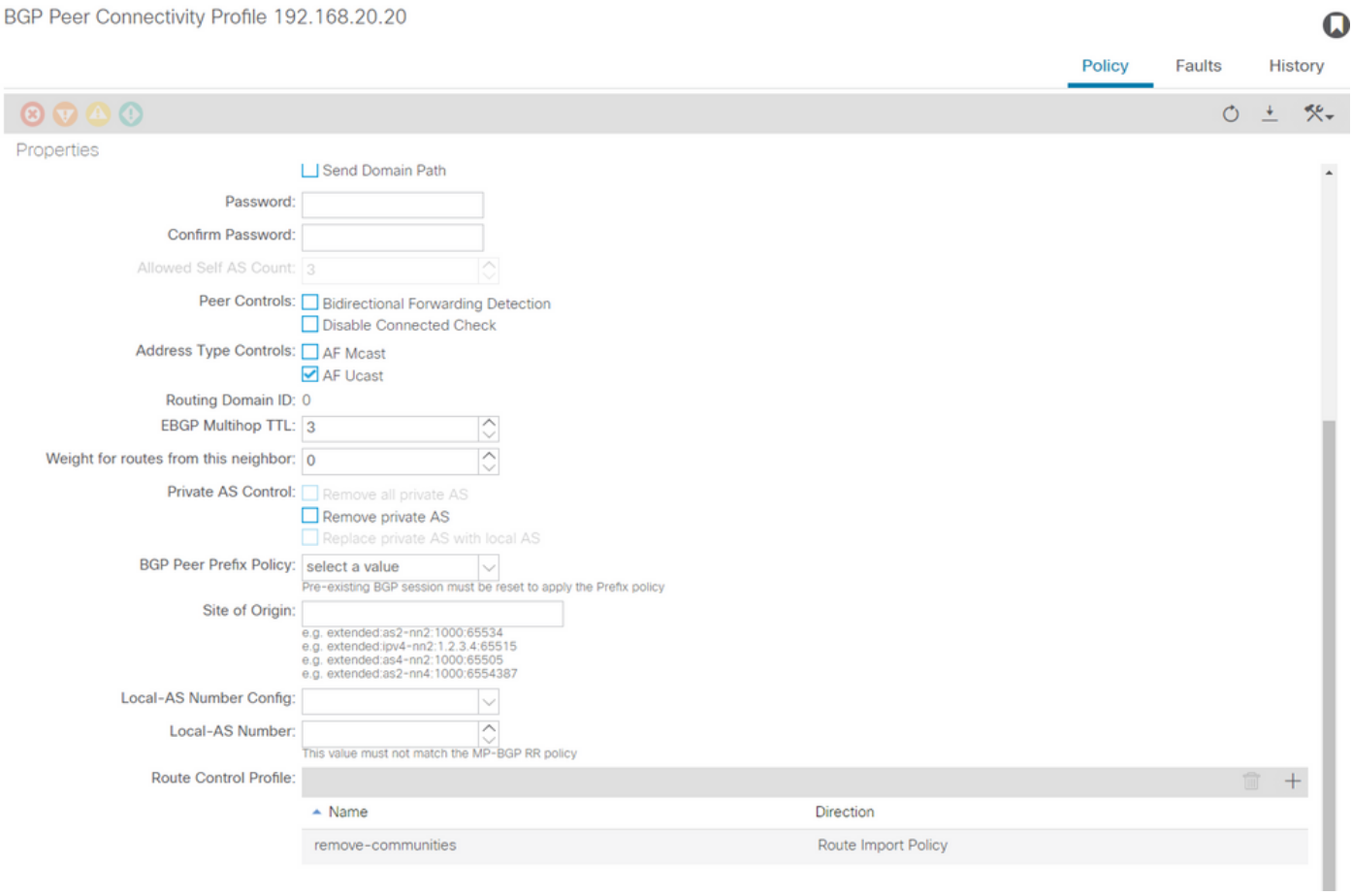

情景现在已添加到路由映射

导航至L3Out中的BGP对等连接配置文件，然后选择 + 图标 Route Control Profile 表，然后使用默认方向添加路由映射 Route Import Policy 已选择：




将路由映射添加到BGP对等连接配置文件

为路由映射选择Update后，您会看到路由映射已添加到 Route Control Profile 表：




路由映射现在已添加到BGP对等连接配置文件

*有关ACI中路由映射配置选项的详细信息，请参阅ACI交换矩阵L3Out白皮书

实施上述解决方案之一后，验证问题是否已解决。

检查BL 101上的BGP状态：

<#root>

```
leaf-101# show ip bgp 192.168.20.0 vrf example:example
BGP routing table information for VRF example:example, address family IPv4 Unicast
BGP routing table entry for 192.168.20.0/24, version 46 dest ptr 0xa0fec840
Paths: (1 available, best #1)
Flags: (0x80c001a 00000000) on xmit-list, is in urib, is best urib route, is in HW, exported
  vpn: version 2731, (0x100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type (0xa96485b8): internal 0x18 0x0 ref 0 adv path ref 2, path is valid, is best path
  AS-Path: NONE, path sourced internal to AS
    192.168.20.20 (metric 5) from 192.168.20.20 (192.168.20.20)
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0
      Extcommunity:
          RT:65001:2162688
          COST:pre-bestpath:163:1879048192


***Notice that no router mac is present here.***
```

```
          VNID:2162688

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

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    10.0.216.65        10.0.216.66
```

检查CL 102上的RIB:

<#root>

```
leaf-102# show ip route 192.168.20.0 vrf example:example
IP Route Table for VRF "example:example"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

192.168.20.0/24, ubest/mbest: 1/0
    *via 10.0.210.70%overlay-1, [200/0], 00:00:06, bgp-65001, internal, tag 65001
         recursive next hop: 10.0.210.70/32%overlay-1
```

***Notice that no rwVnid entry is present here.***

---

注意：单独的rwVnid条目不存在或存在并不确定问题是否发生。在许多情况下，一旦问题解决，rwVnid条目就会从有问题的路由中删除。然而，情况并非总是如此。请务必检查FIB和HAL表以验证问题是否已解决。

---

检查CL 102上的FIB:

<#root>

```
module-1(DBG-elam-insel6)# show forwarding route 192.168.20.0 vrf example:example

IPv4 routes for table example:example/base

------------------+------------------+----------------------+------------------------

Prefix            | Next-hop         | Interface/VRF        | Additional Info

------------------+------------------+----------------------+------------------------

*192.168.20.0/24


10.0.210.70

        overlay-1
```

***Notice that we have the route here and our next-hop address is correct (showing the TEP IP of BL 101)

```
Route Class-id:0x0
Policy Prefix 0.0.0.0/0

leaf-102# acidiag fnvread | grep 101
     101        1             leaf-101

10.0.210.70/32

    leaf         active   0
```

CL 102上的HAL表：

<#root>

```
module-1(DBG-elam-insel6)# show platform internal hal l3 routes | grep 192.168.20.0
|

4662

| 192.168.20.0/ 24| UC|  686|     20601| TRIE|   a5| 5/  0| 60a5|A|     8443|    86b6|  ef5| 1/  2|

***Notice that we have an entry here and it's in the correct VRF.***


module-1(DBG-elam-insel6)# hex

4662


0x

1236


module-1(DBG-elam-insel6)# show platform internal hal l3 vrf pi
=============================================================================================================
                                       | -- TOR --  | - Spine -  |                ACL                  |
      Vrf           Hw    I I Vrf     | SB    NB   | Proxy ACI  |      Ing               Egr          |
VrfId Name          VrfId I S Vnid    | BDId  BDId | Ou Bd Enc  | Lbl      Msk      Lbl      Msk      |
=============================================================================================================
26    example:example

1236

0 0 210000    0     0      0     1      0        0        0        0        0
```

从EP（主机1）对来自外部BGP对等体(192.168.20.20)的外部网络中的主机执行ping操作：

<#root>

```
Host-1# ping 192.168.20.20 vrf example
PING 192.168.20.20 (192.168.20.20): 56 data bytes
64 bytes from 192.168.20.20: icmp_seq=0 ttl=252 time=1.043 ms
64 bytes from 192.168.20.20: icmp_seq=1 ttl=252 time=1.292 ms
64 bytes from 192.168.20.20: icmp_seq=2 ttl=252 time=1.004 ms
```

```
64 bytes from 192.168.20.20: icmp_seq=3 ttl=252 time=0.769 ms
64 bytes from 192.168.20.20: icmp_seq=4 ttl=252 time=1.265 ms

--- 192.168.20.20 ping statistics ---
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min/avg/max = 0.769/1.074/1.292 ms
```

******Connectivity is there.******

CL 102上的ELAM:

<#root>

```
leaf-102# vsh_lc
module-1# debug platform internal roc elam asic 0
module-1(DBG-elam)# trigger reset
module-1(DBG-elam)# trigger init in-select 6 out-select 0
module-1(DBG-elam-insel6)# set outer ipv4 src_ip 192.168.10.10 dst_ip 192.168.20.20
module-1(DBG-elam-insel6)# start
module-1(DBG-elam-insel6)# stat
 ELAM STATUS
===========
Asic 0 Slice 0 Status Armed
Asic 0 Slice 1 Status Triggered

module-1(DBG-elam-insel6)# ereport
Python available. Continue ELAM decode with LC Pkg
 ELAM REPORT
<output omitted>
--------------------------------------------------------------------------------------------------------------------
Lookup Drop
--------------------------------------------------------------------------------------------------------------------
LU drop reason                          :
```

**no drop**

******Traffic forwards correctly.******

# 相关信息

- 此缺陷中还记录了此行为:Cisco Bug ID CSCvx28929
- 技术支持和文档 - Cisco Systems

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）。

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）。