修復到ACI交換矩陣的EVPN RMAC ExtCommunity傳輸問題

目錄

簡介

本檔案介紹從外部邊界閘道通訊協定(BGP)對等點接收時，路由器MAC擴充社群屬性設定錯誤對ACI網狀架構的影響。

背景資訊

使用BGP時，有一個選項可用來傳送具有通告給BGP對等體的首碼的社群和延伸社群屬性。這些社群屬性允許我們修改路由策略並動態改變路由流量的處理方式。

問題

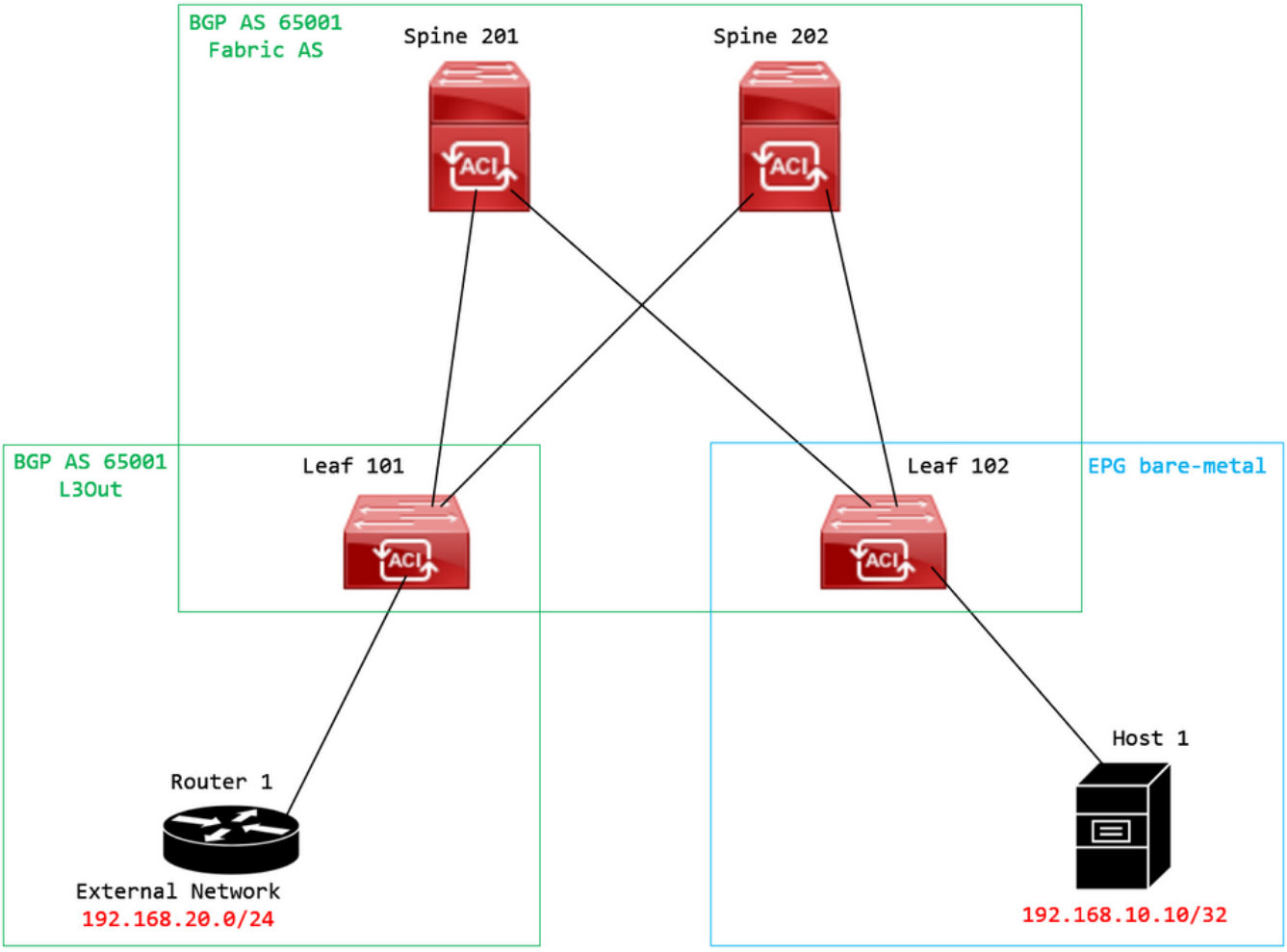
當路由器MAC擴展社群屬性從外部BGP對等體傳送到ACI交換矩陣時帶有IPv4 AFI字首時，交換矩陣中任何通過內部MP-BGP進程接收來自邊界枝葉的路由的枝葉上都會發生FIB和HAL程式設計錯誤。這是因為RMAC extcommunity屬性屬於BGP L2VPN EVPN地址系列，當將其注入BGP IPv4地址系列時，該屬性會被拒絕。這是由於違反了規則5.2（統一傳播模式），該規則在IETF標題為「EVPN與IPVPN互通」的文檔中進行了描述。在第15頁專案4c中，指出了具體問題：

```
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.
```

文檔連結：EVPN與IPVPN互通

以下是iBGP的問題範例，但是eBGP也出現了問題。

拓撲圖：



拓撲圖

在外部BGP對等裝置（路由器1）上設定路由映像，並設定EVPN RMAC extcommunity屬性：

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

在BGP鄰居IPv4地址系列配置下，配置BGP擴展社群，並在出站方向配置路由對映：

<#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
```

檢查BL 101上的BGP狀態：

<#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
```

檢查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

檢查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.

檢查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.

從EP（主機1）ping來自外部BGP對等體(192.168.20.20)的外部網路中的主機：

<#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.******

檢查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.******

解決方案

解決方式為停止將具有IPv4位址系列首碼的路由器MAC擴充社群屬性從外部BGP對等點傳送到ACI交換矩陣。

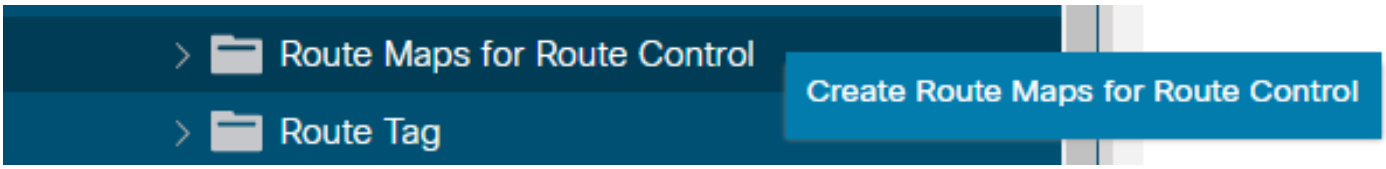
刪除先前配置的路由對映，並停止從外部BGP對等裝置（路由器1）傳送擴展社群。移除其中任一

組態（或同時移除兩者）會起作用：

```
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
```

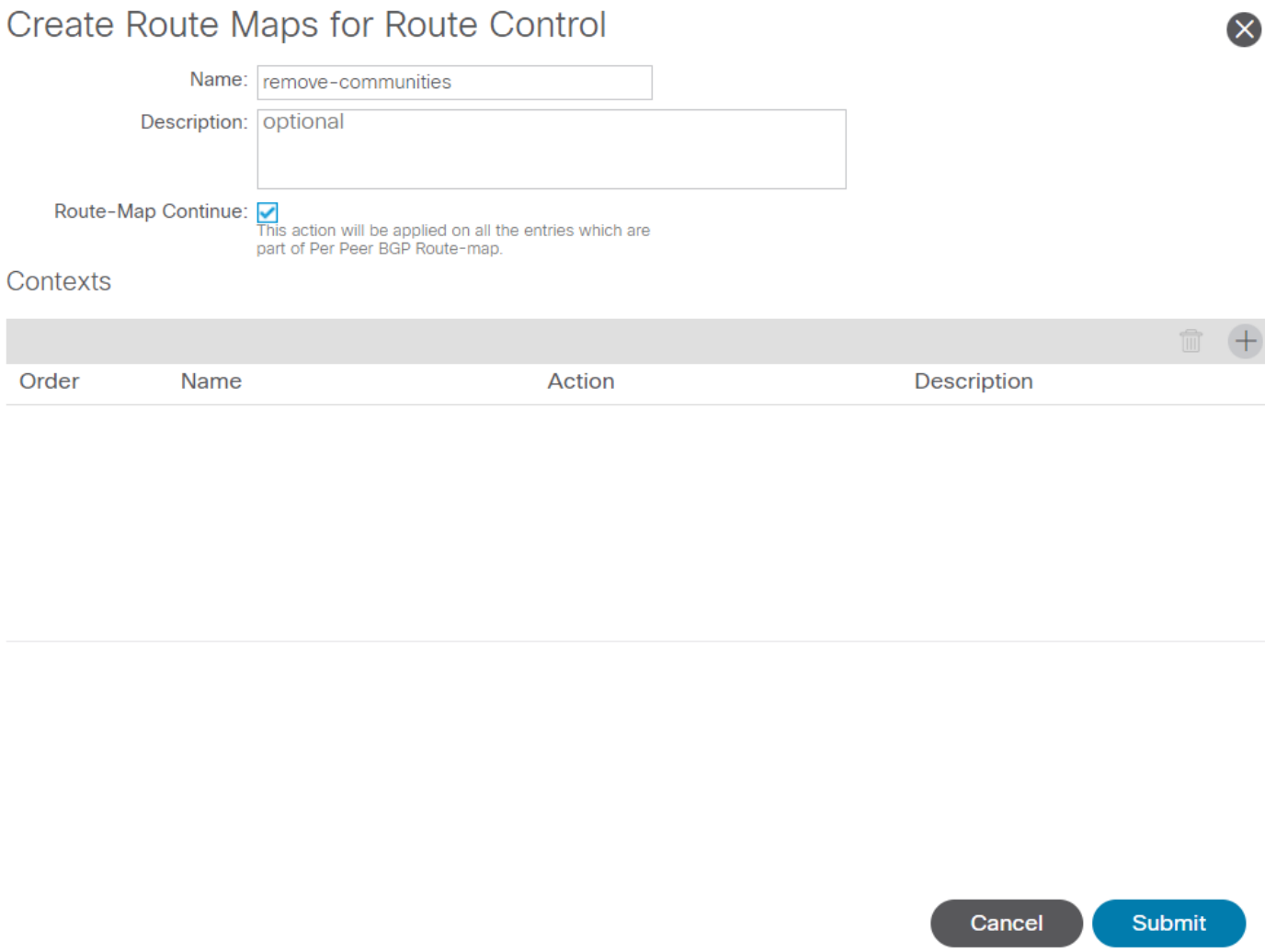
另一個（不太推薦）解決方案是，通過在ACI中配置的L3Out中建立路由對映，簡單過濾從外部BGP對等裝置接收的所有團體。

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



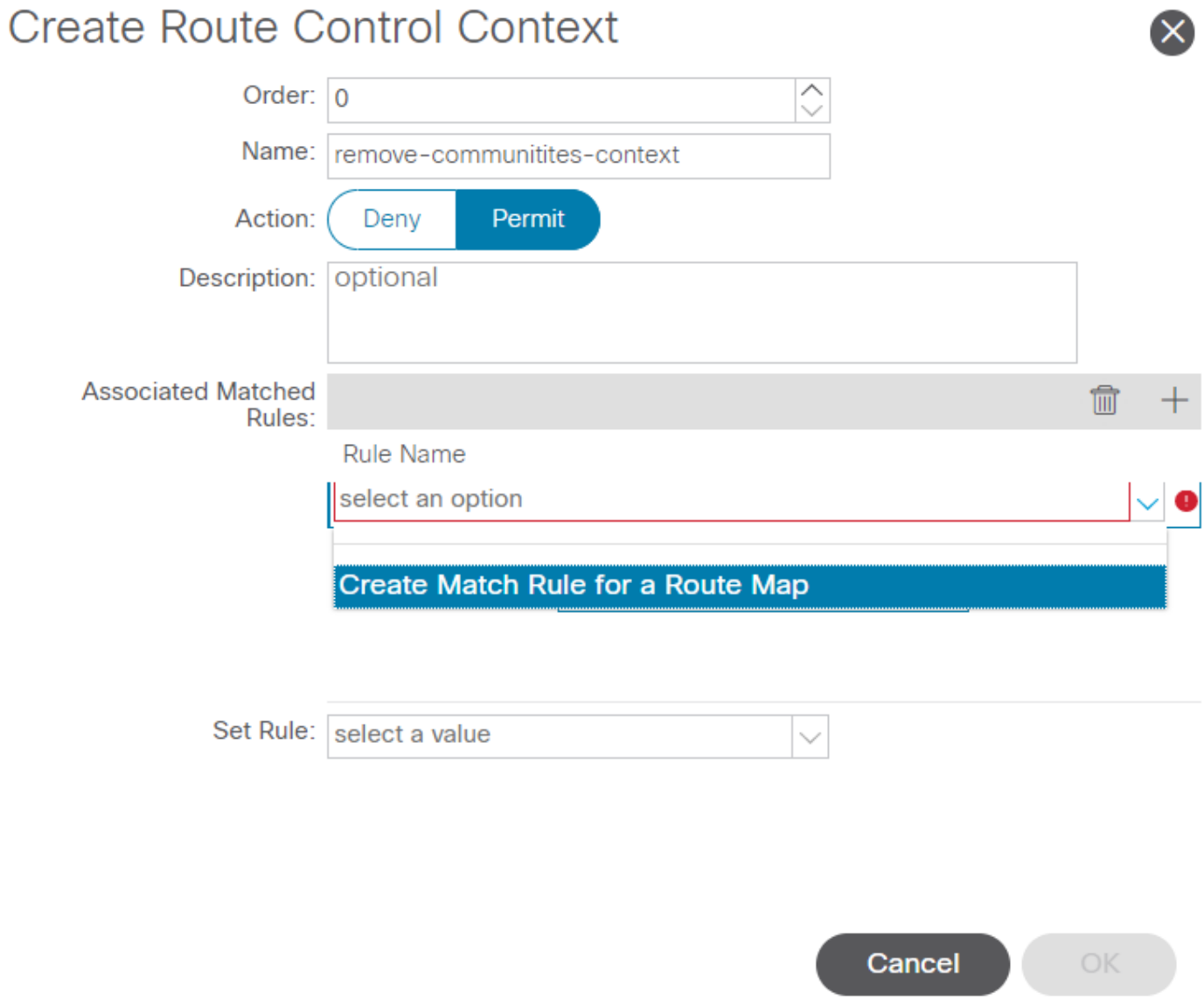
選擇為路由控制建立路由對映的選項

命名您的路由對映，啟用 Route-Map Continue 選項，然後新增上下文。選擇 + 圖示在Contexts表中：



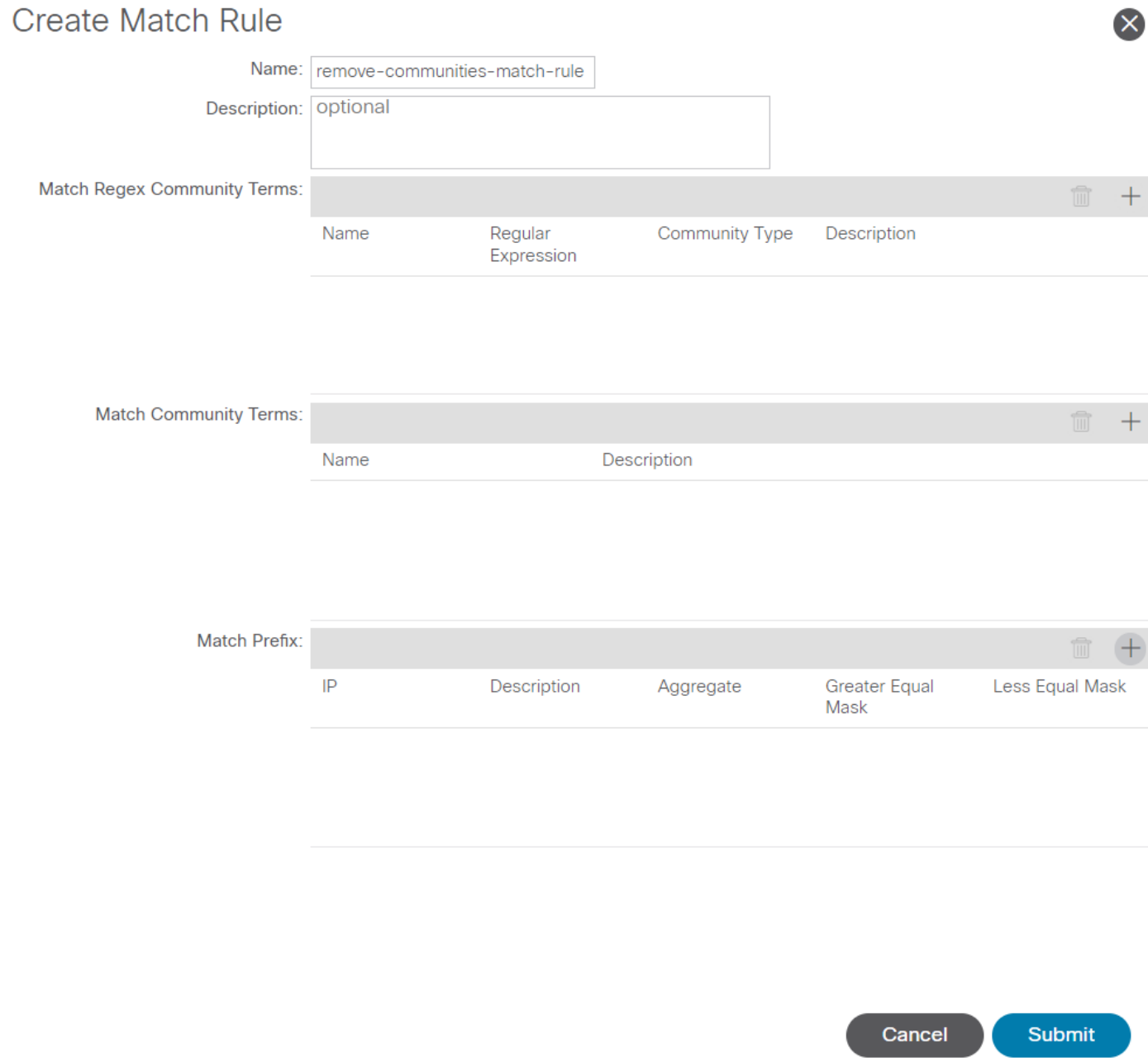
建立路由對映和建立上下文

命名上下文，並保留預設操作 Permit 選中，然後通過選擇 + 圖示 Associated Matched Rules 表，然後選擇 **Create Match Rule for a Route Map**：



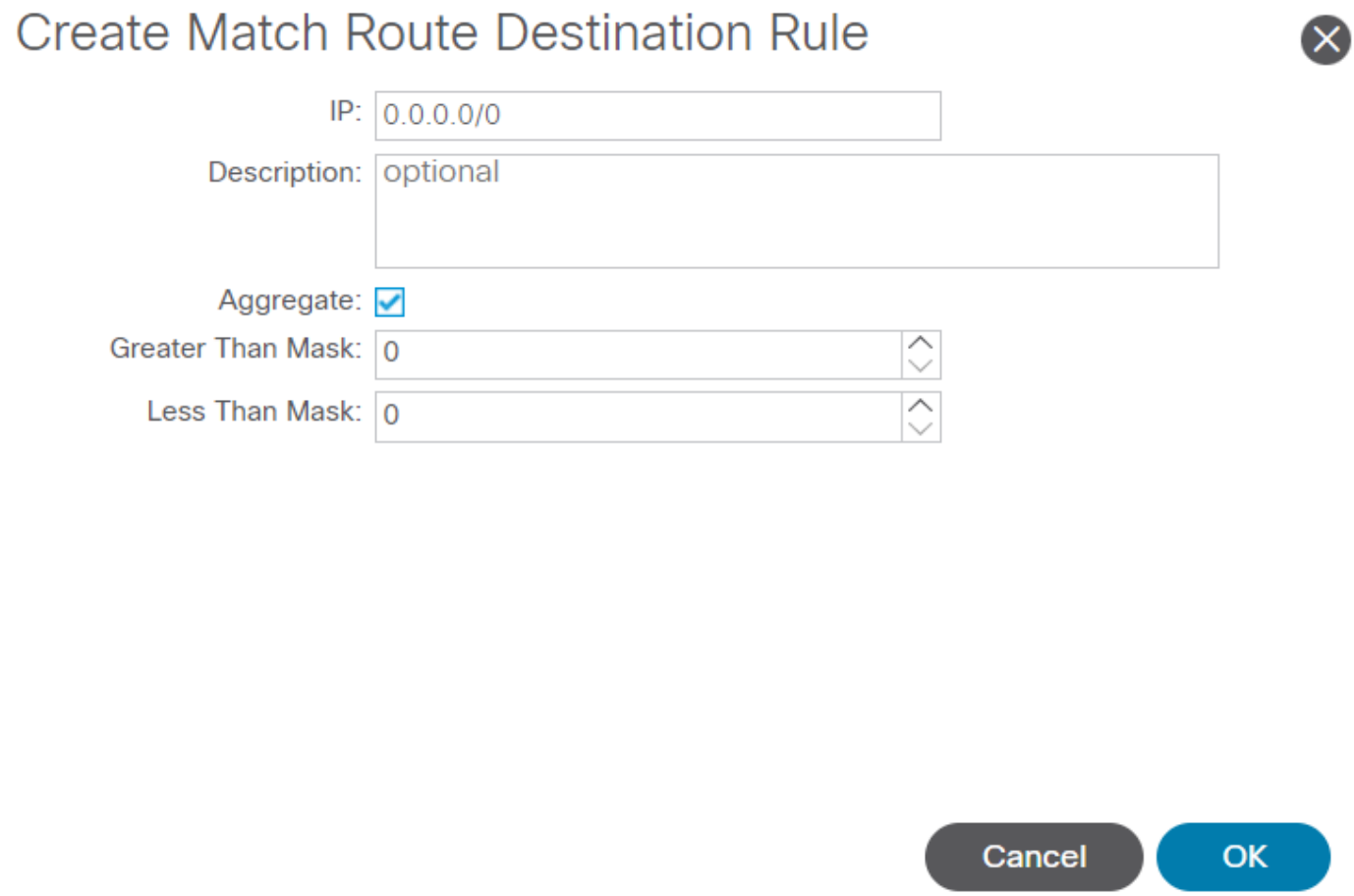
建立路由控制上下文並選擇為路由對映建立匹配規則的選項

將您的匹配規則命名為，然後通過選擇中的+圖示新增新的字首 Match Prefix **表：**



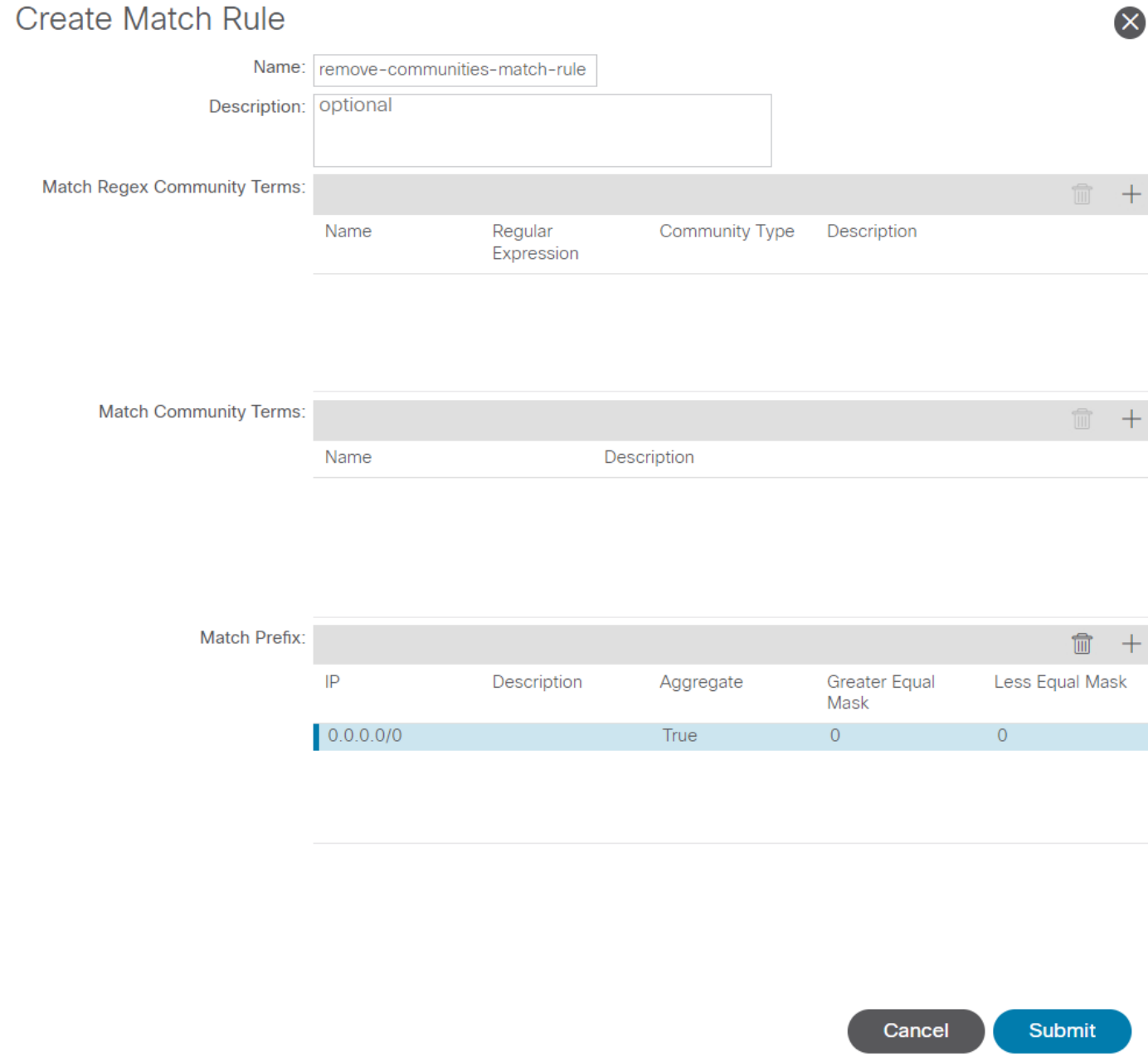
建立匹配規則並建立匹配字首

新增所需的字首。此示例說明如何新增所有字首的聚合：



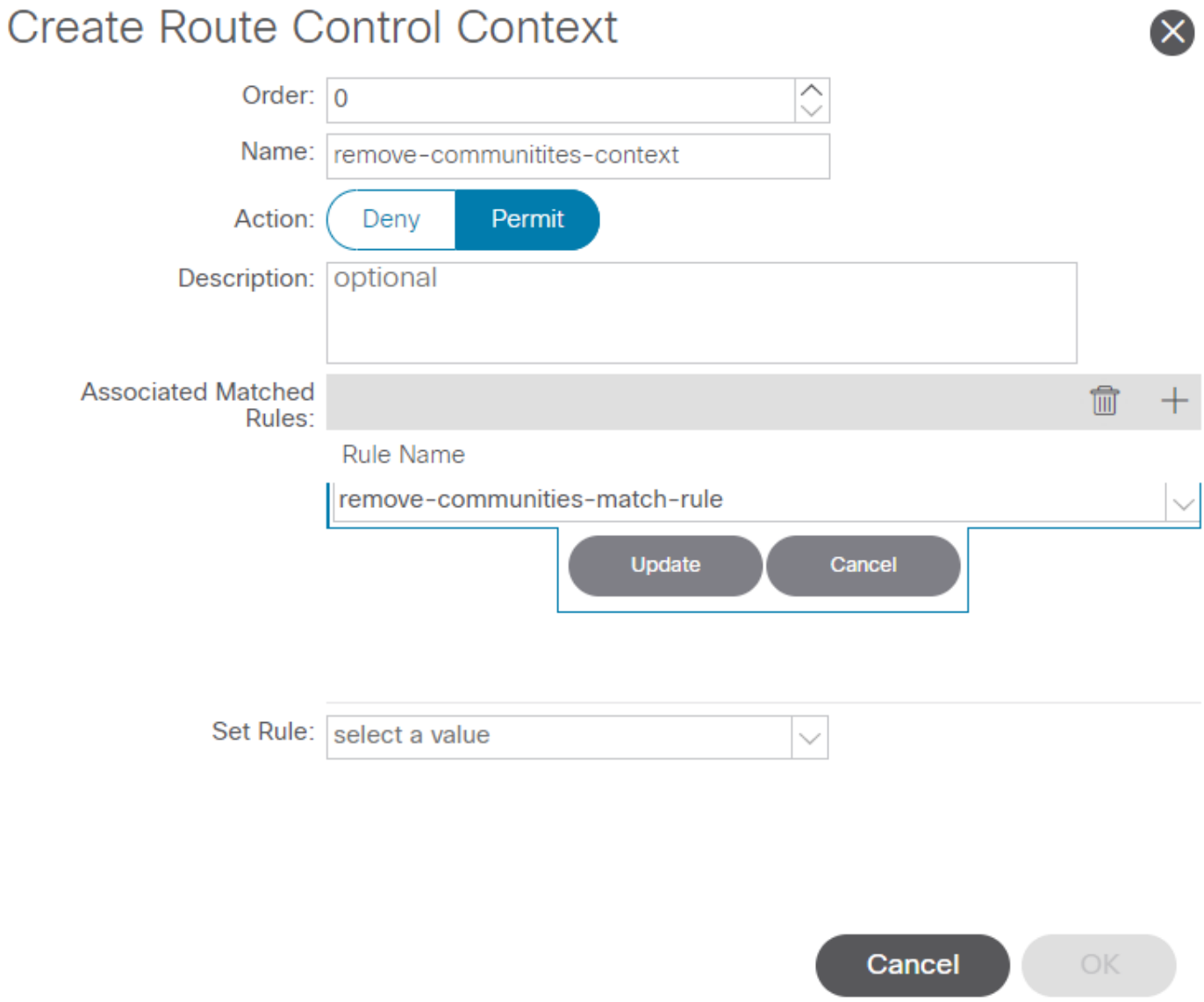
建立匹配路由目標規則

選擇之後 **OK** 在 Create Match Route Destination Rule 視窗中，您會看到您的字首已新增到 Match Prefix 中的表 Create Match Rule 視窗：



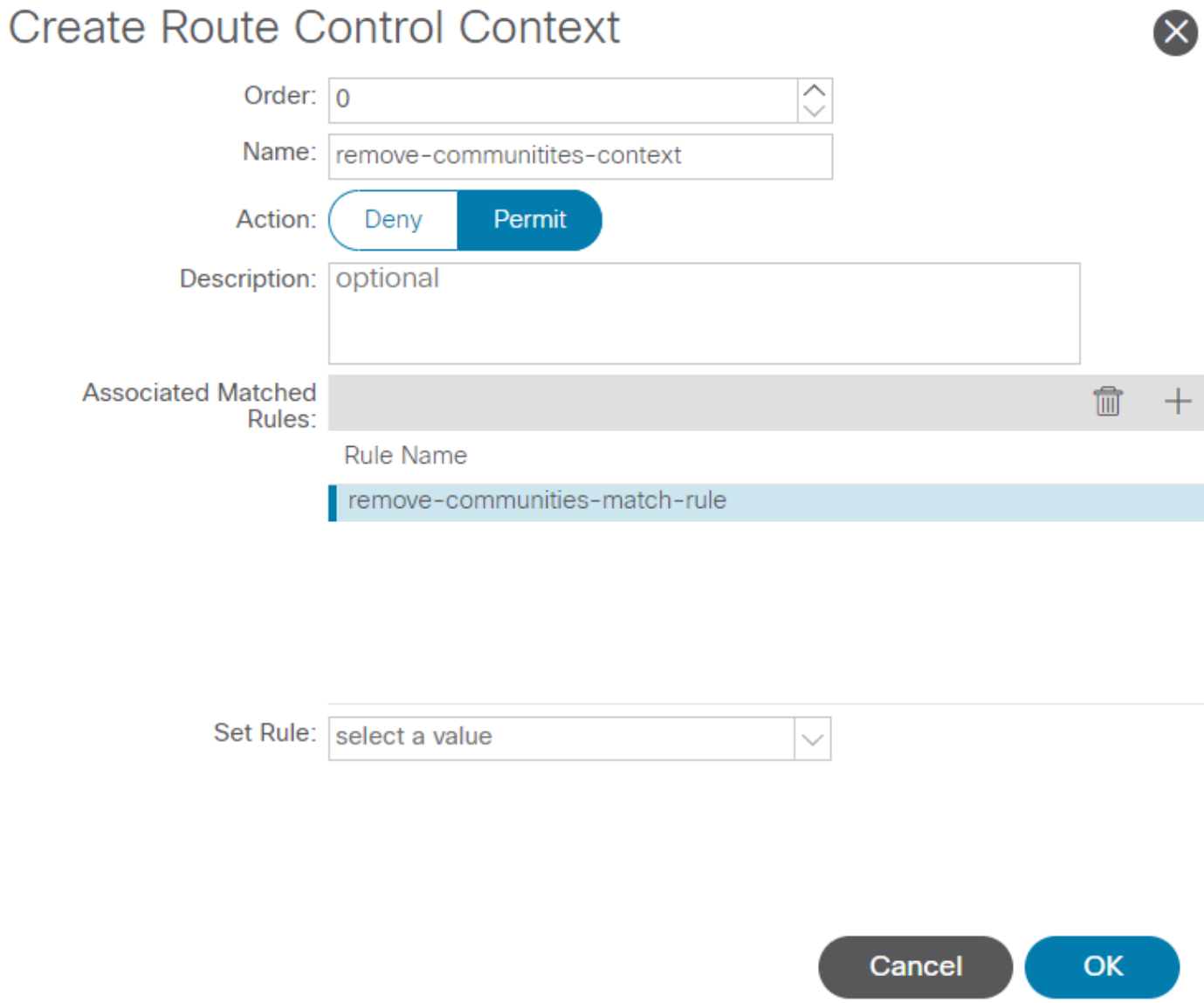
Match Prefix現在已新增到匹配規則

選擇之後 Submit 在 Create Match Rule 視窗，選擇 Update 在 Associated Matched Rules 中的表 Create Route Control Context 視窗：



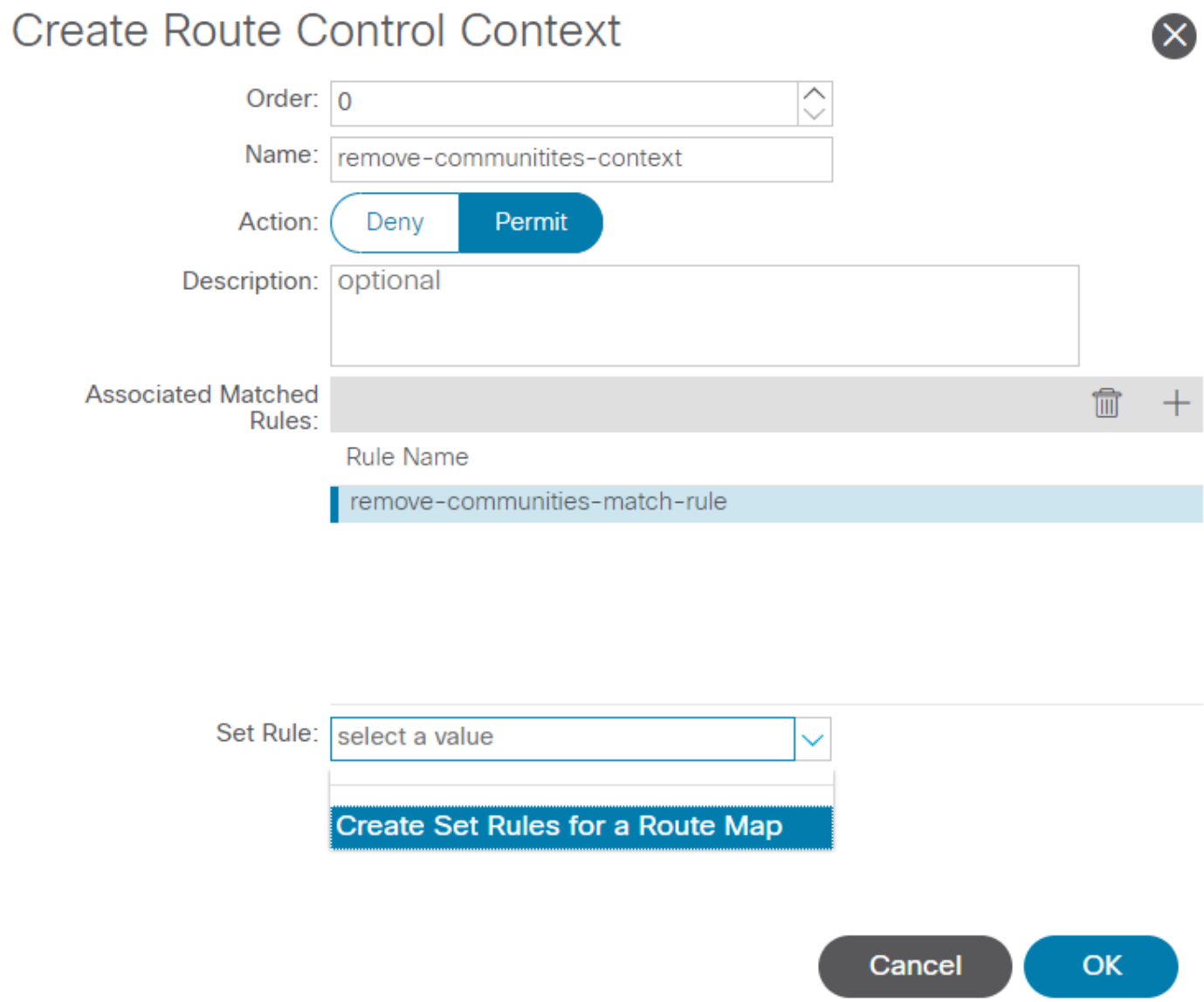
將關聯的匹配規則新增到路由控制上下文

現在，您的關聯匹配規則已新增到您的上下文：



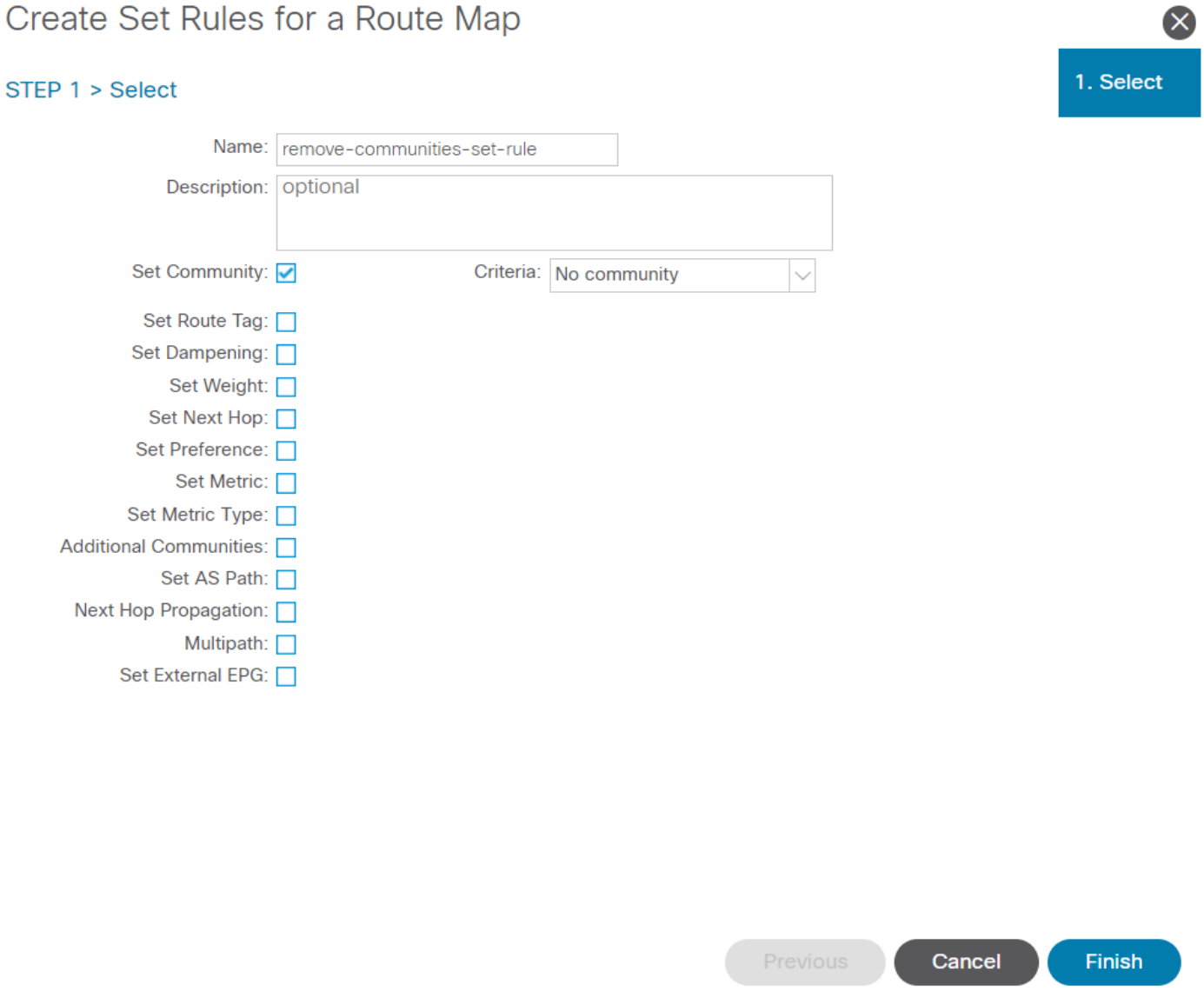
關聯匹配規則現在已新增到路由控制上下文

然後，選擇旁邊的下拉選單 Set Rule 並選取 Create Set Rules for a Route Map：



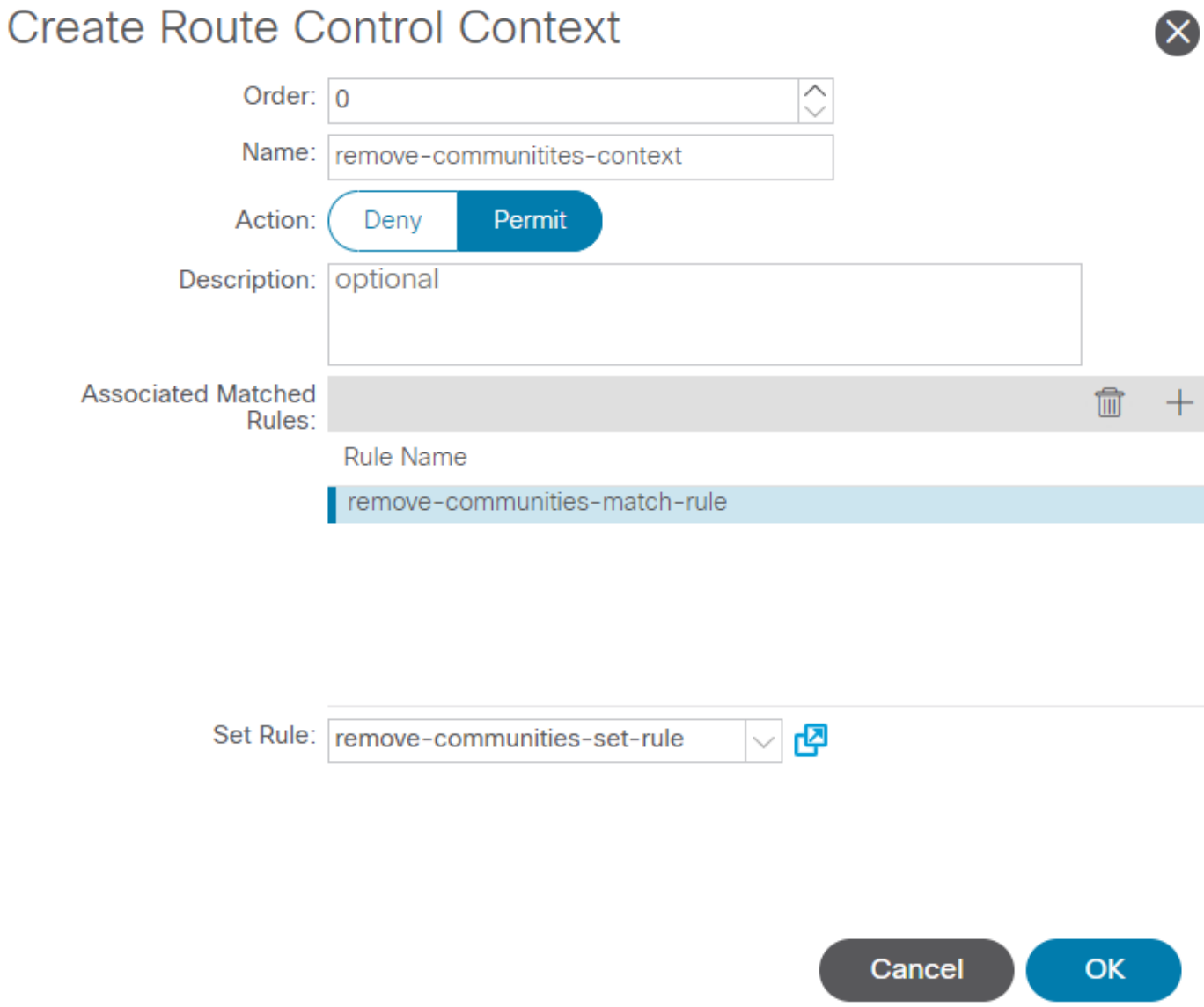
選擇選項為路由對映建立集規則

命名您的設定規則，然後選擇 Set Community 選項並保留預設條件 No community 已選取:



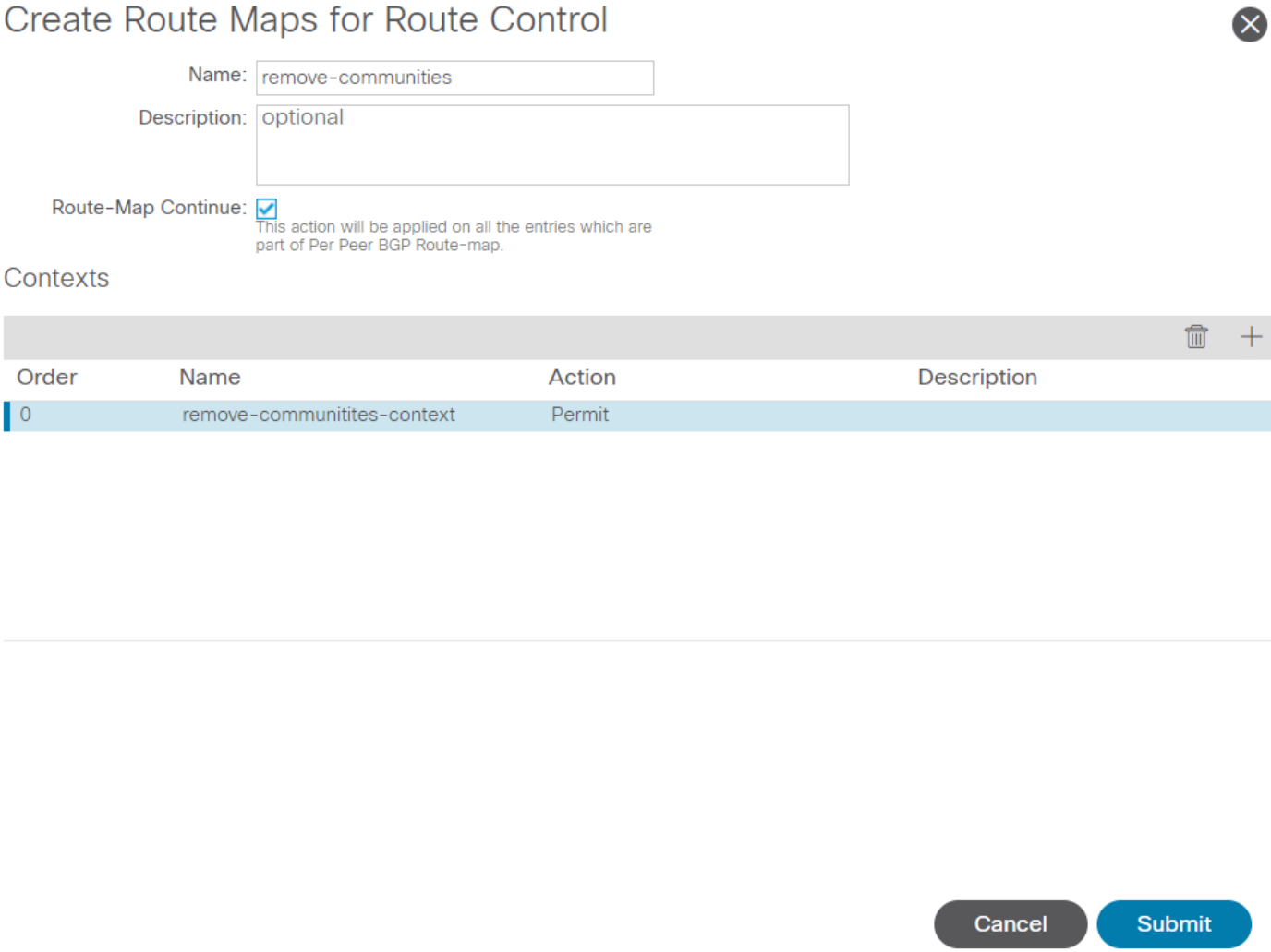
為路由對映建立集規則

在 Create Set Rules for a Route Map 視窗中，您會看到您在 Create Route Control Context 視窗：



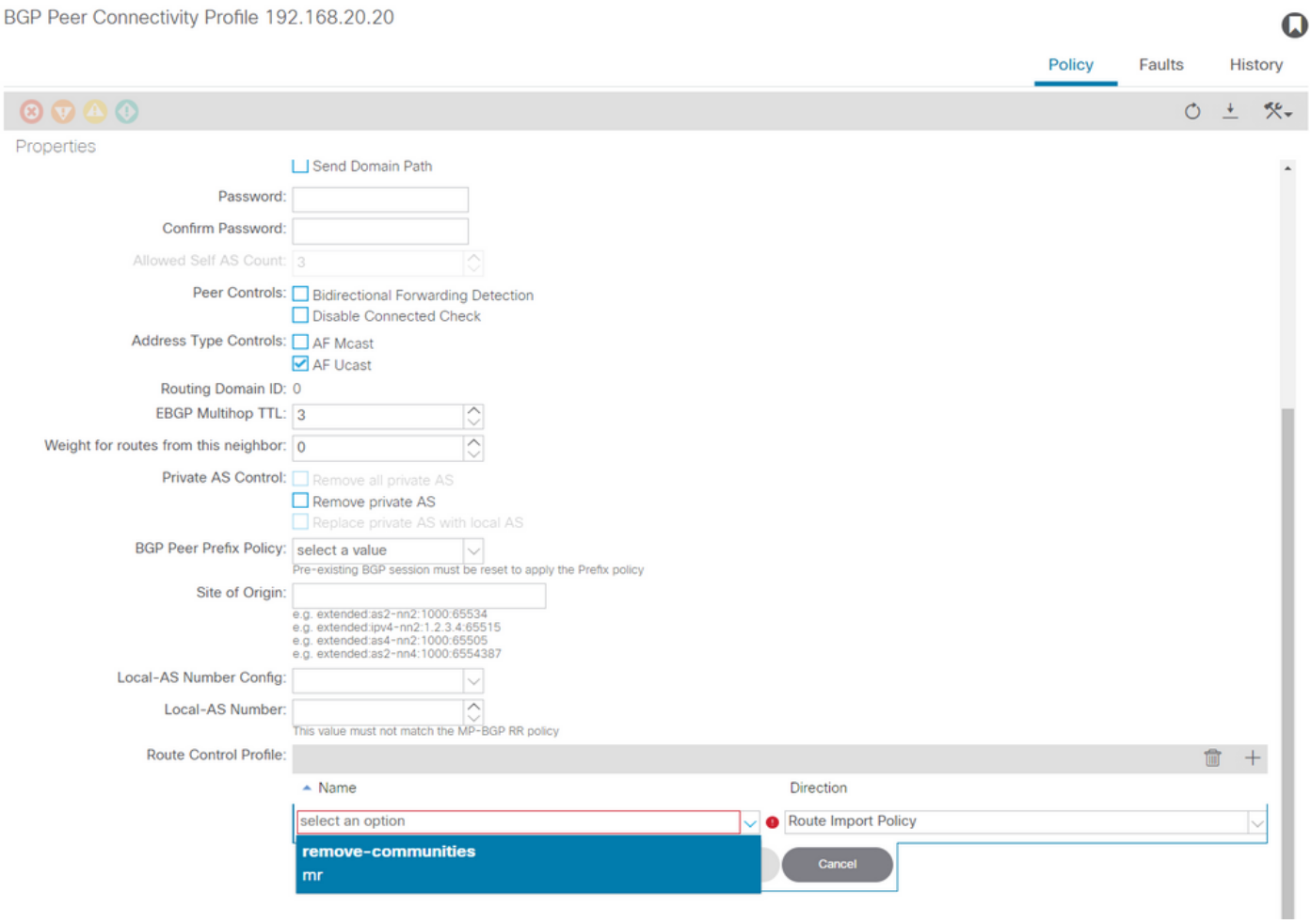
Set Rule現在已新增到路由控制上下文

選擇之後 OK 在 Create Route Control Context 視窗，您會看到您的上下文已新增到 Contexts 中的表 Create Route Maps for Route Control 視窗。最後，選擇 Submit 要完成配置，請執行以下操作：



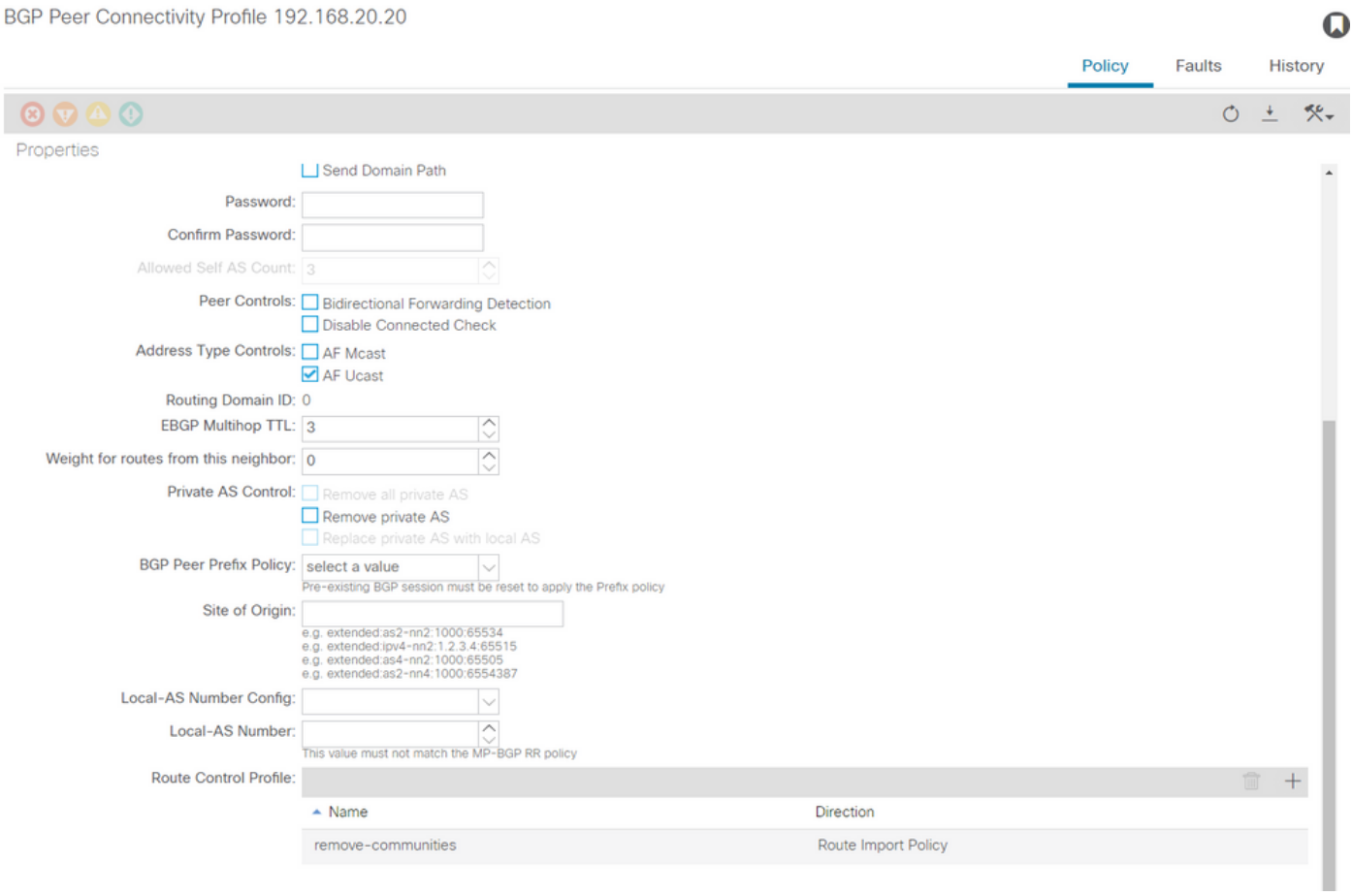
現在情景已新增到路由對映

導航到L3Out中的BGP對等連線配置檔案，然後選擇 + 圖示 Route Control Profile 表，然後使用預設方向新增您的路由對映 Route Import Policy 已選取:



將路由對映新增到BGP對等連線配置檔案

為路由對映選擇Update後，您會看到您的路由對映已新增到 Route Control Profile 表：



路由對映現在已新增到BGP對等連線配置檔案

*有關ACI中路由對映配置選項的詳細資訊，請參閱ACI交換矩陣L3Out白皮書

實施上述解決方案之一後，驗證問題是否已解決。

檢查BL 101上的BGP狀態：

<#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
```

檢查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.

註：單獨的rwVnid條目的缺失或存在並不確定問題是否發生。在許多情況下，一旦問題得到解決，便會將rwVnid條目從相關路由中刪除。然而，情況並非總是如此。請始終檢查FIB和HAL表以驗證問題是否已解決。

檢查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
```

從EP（主機1）ping來自外部BGP對等體(192.168.20.20)的外部網路中的主機：

<#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.******

相關資訊

- 此行為也記錄在此缺陷中：思科錯誤ID CSCvx28929
- 技術支援與文件 - Cisco Systems

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）。