在Catalyst 9000系列交换机上实施BGP EVPN保护的重叠分段

目录

简介

本文档介绍如何在Catalyst 9000系列交换机上实施BGP EVPN VXLAN保护的重叠分段。

先决条件

要求

Cisco 建议您了解以下主题：

- BGP EVPN VxLAN概念
- BGP EVPN单播故障排除
- BGP EVPN VxLAN路由策略

使用的组件

本文档中的信息基于以下软件和硬件版本：

- Catalyst 9300
- Catalyst 9400
- Catalyst 9500
- Catalyst 9600
- Cisco IOS® XE 17.12.1及更高版本

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您的网络处于活动状态，请确保您了解所有命令的潜在影响。

背景信息

高级功能说明

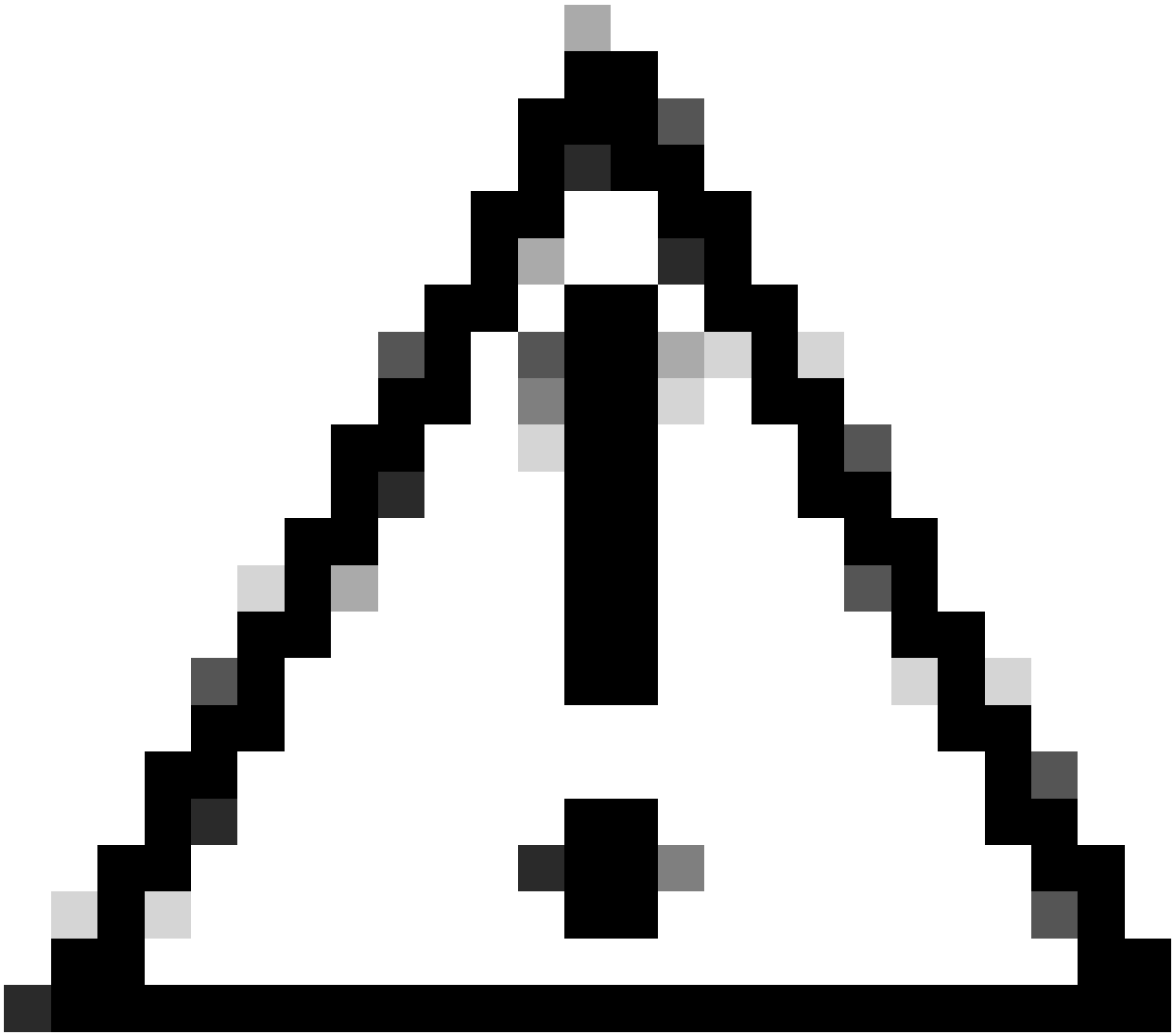
受保护分段功能是一种安全措施，可以防止端口相互转发流量，即使它们位于同一个VLAN和同一台交换机也是如此

- 此功能类似于“受交换机端口保护”或专用VLAN，但适用于EVPN交换矩阵。
- 此设计会强制所有流量发送到CGW，在发送到最终目的地之前，防火墙会先检查这些流量。
- 使用集中式安全设备，流量可控制、确定且易于检查。

文档详细信息

本文档是第2部分或第3部分相互关联的文档：

- 文档1：在Catalyst 9000系列交换机上实施BGP EVPN路由策略介绍如何在重叠中控制BGP BUM流量，必须首先进行配置
- 文档2：本文档。本文档基于文档1的重叠设计和策略，描述了“protected”关键字的实施
- 文档3：在Catalyst 9000系列交换机上实施BGP EVPN DHCP第2层中继介绍DHCP中继如何在仅第2层VTEP上工作



注意：在实施受保护分段配置之前，您必须实施文档1中的配置。

受保护分段类型

完全隔离

- 仅允许北向南通信，并且

- 使用“default-gateway advertise” CLI将网关通告到交换矩阵

大部分孤立

- 允许北向南通信（在此使用案例中，根据防火墙流量策略允许东/西流量）
- 允许从东向西通信（基于防火墙流量策略）
- 网关位于交换矩阵外部，且SVI不使用“默认网关通告” CLI进行通告

交换机行为

- 即使主机连接到同一台交换机，它们也无法直接相互通信(当主机位于同一VRF/Vlan/网段时，ARP请求不会发送到同一交换机上的其他端口)。
- L2 VTEP之间没有BUM流量(使用路由策略配置过滤IMET前缀)
- 来自主机的所有数据包都会中继到边界枝叶以转发。(这意味着主机1要与同一枝叶上的主机2通信，流量通过CGW固定)

路由类型2处理

- 接入枝叶通过E-Tree Extended Community和Leaf标志设置通告本地RT2。
- 接入枝叶不会安装任何接收的E-Tree Extended Community和Leaf标记设置在数据平面上的远程RT2。
- 接入枝叶不会相互在数据平面中安装RT2。
- 接入枝叶和边界枝叶(CGW)在数据平面中相互安装RT2。
- 无需在接入枝叶或边界枝叶上进行配置更改。

设计摘要

- 对于广播(BUM)，RT3拓扑是集中星型拓扑，用于强制广播流量（例如ARP）到达GCW。
- 为了考虑主机移动性，RT2在BGP控制平面上是全网状（当主机从一个VTEP移动到另一个VTEP时，RT2中的序列号会增加）
- 数据平面选择性地安装MAC地址。
    - 枝叶仅安装包含DEF GW属性的本地MAC和RT2
    - CGW没有受保护的KW，并将所有本地MAC和远程RT2安装在其数据平面中。

术语

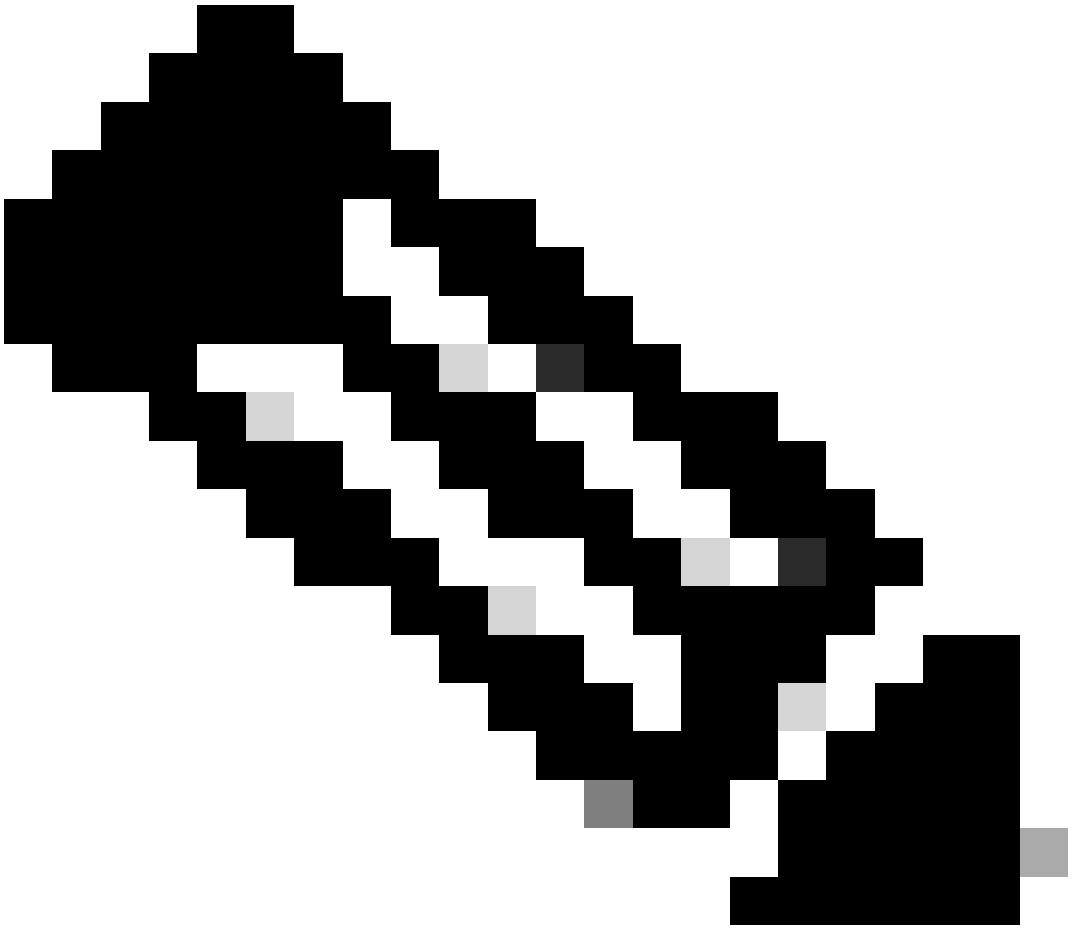
VRF	虚拟路由转发	定义与其他VRF和全局IPv4/IPv6路由域分离的第3层路由域
AF	地址系列	定义BGP处理的前缀类型和路由信息
AS	自治系统	一组属于一个网络或一组网络的可路由IP前缀，全部由单个实体或组织管理、控制和监督

EVPN	以太网虚拟专用网络	允许BGP传输第2层MAC和第3层IP信息的扩展是EVPN，并使用多协议边界网关协议(MP-BGP)作为协议来分发有关VXLAN重叠网络的可达性信息。
VXLAN	虚拟可扩展LAN（局域网）	VXLAN旨在克服VLAN和STP的固有局限性。推荐的IETF标准[RFC 7348]与VLAN提供相同的以太网第2层网络服务，但灵活性更高。从功能上讲，它是MAC-in-UDP封装协议，在第3层底层网络上作为虚拟重叠运行。
CGW	集中式网关	以及网关SVI不在每个枝叶上的EVPN的实施。相反，所有路由都由使用非对称IRB（集成路由和桥接）的特定枝叶完成
DEF网关	默认网关	通过“l2vpn evpn”配置部分下的“default-gateway advertise enable”命令添加到MAC/IP前缀的BGP扩展社区属性。
IMET (RT3)	包括组播以太网标记（路由）	也称为BGP类型3路由。此路由类型在EVPN中用于在VTEP之间传输BUM（广播/未知单播/组播）流量。
RT2	路由类型2	BGP MAC或MAC/IP前缀，表示主机MAC或网关MAC-IP
EVPN经理	EVPN管理器	用于各种其他组件的中央管理组件（例如：从SISF获知并向L2RIB发送信号）
SISF	交换机集成安全功能	EVPN使用的不可知主机跟踪表，用于了解枝叶上的本地主机
L2RIB	第2层路由信息库	在用于管理BGP、EVPN管理器、L2FIB之间交互的中间组件中
FED	转发引擎驱动程序	对ASIC（硬件）层进行编程
MATM	Mac地址表管理器	IOS MATM：仅安装本地地址和 FED MATM：硬件表，安装从控制平面获知的本地和远程地址，是硬件转发平面的一部分

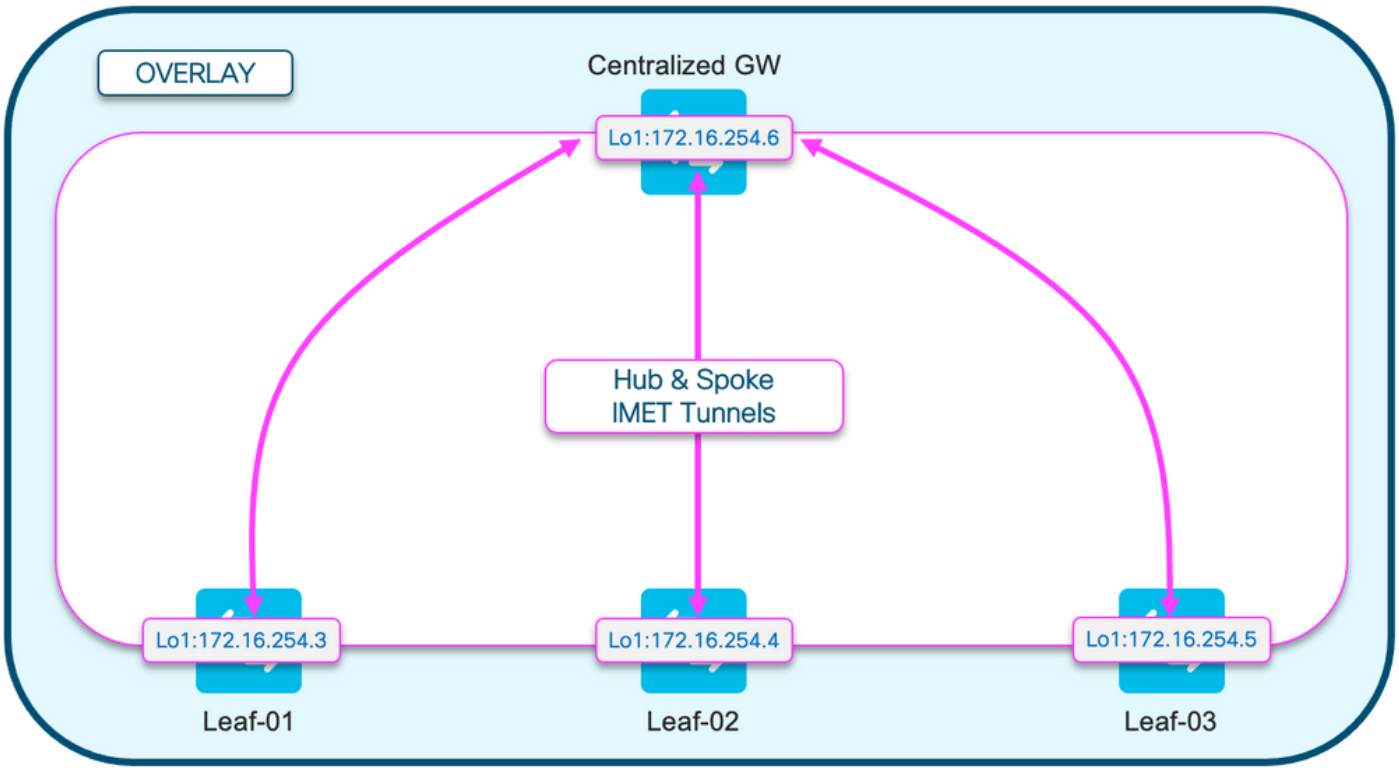
流程图

路由类型2 (RT2)图

下图显示了第2类MAC/MAC-IP主机前缀的全网状设计。

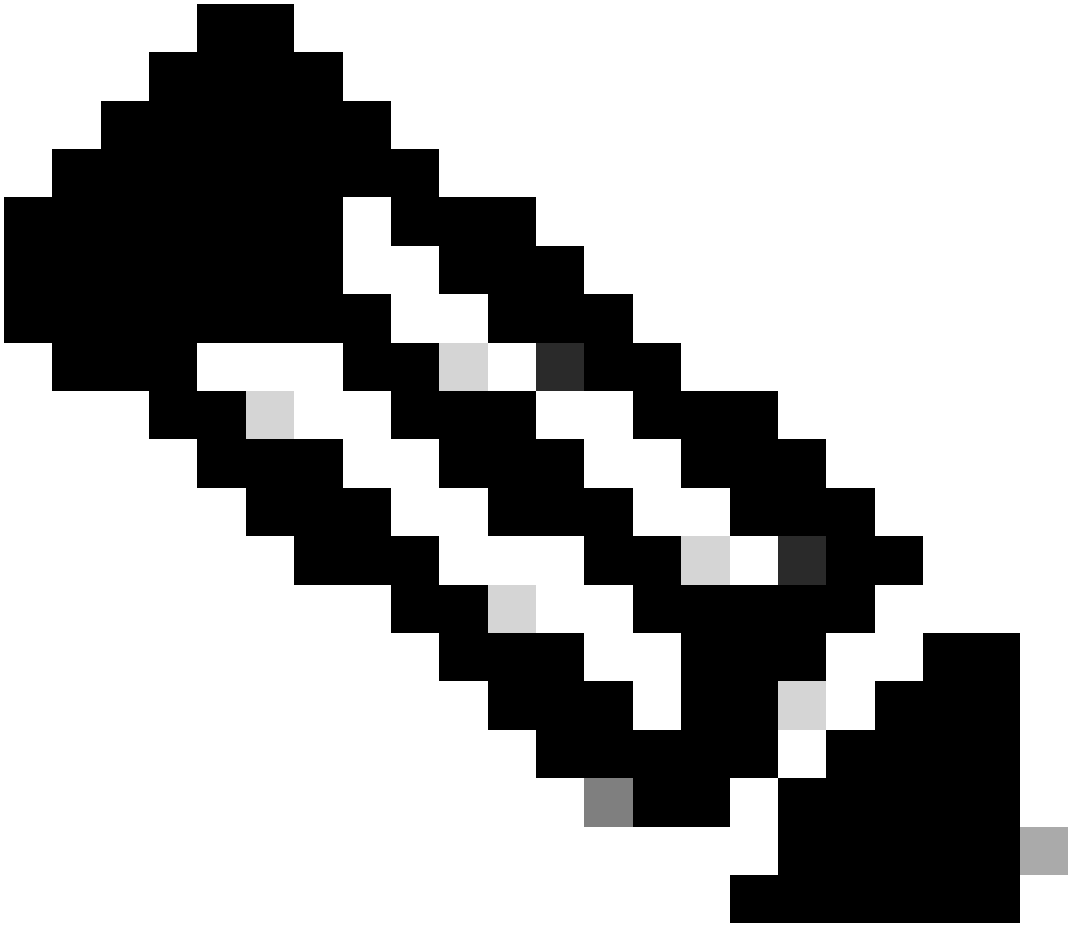


注意：支持移动性和漫游需要全网状

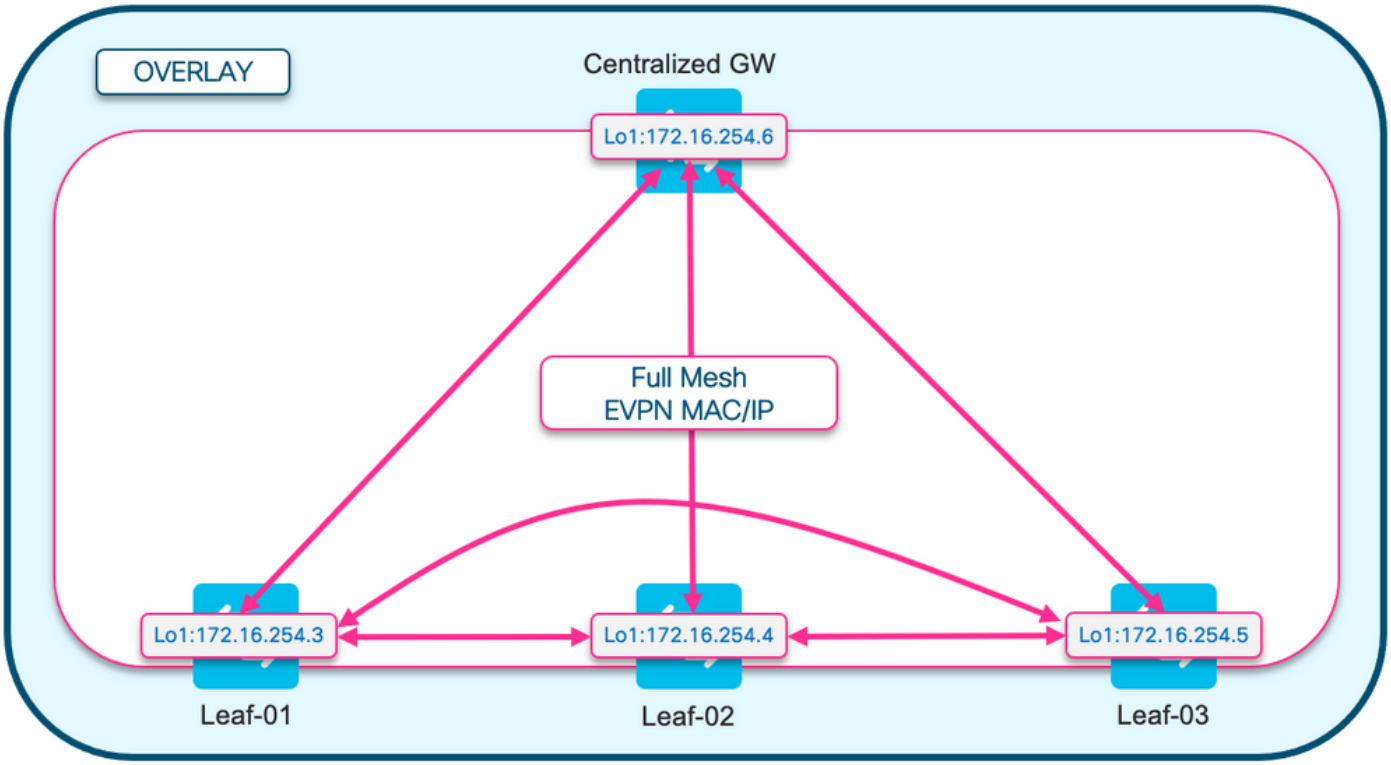


路由类型3 (RT3)图

此图显示广播IMET (RT3)隧道的星型设计

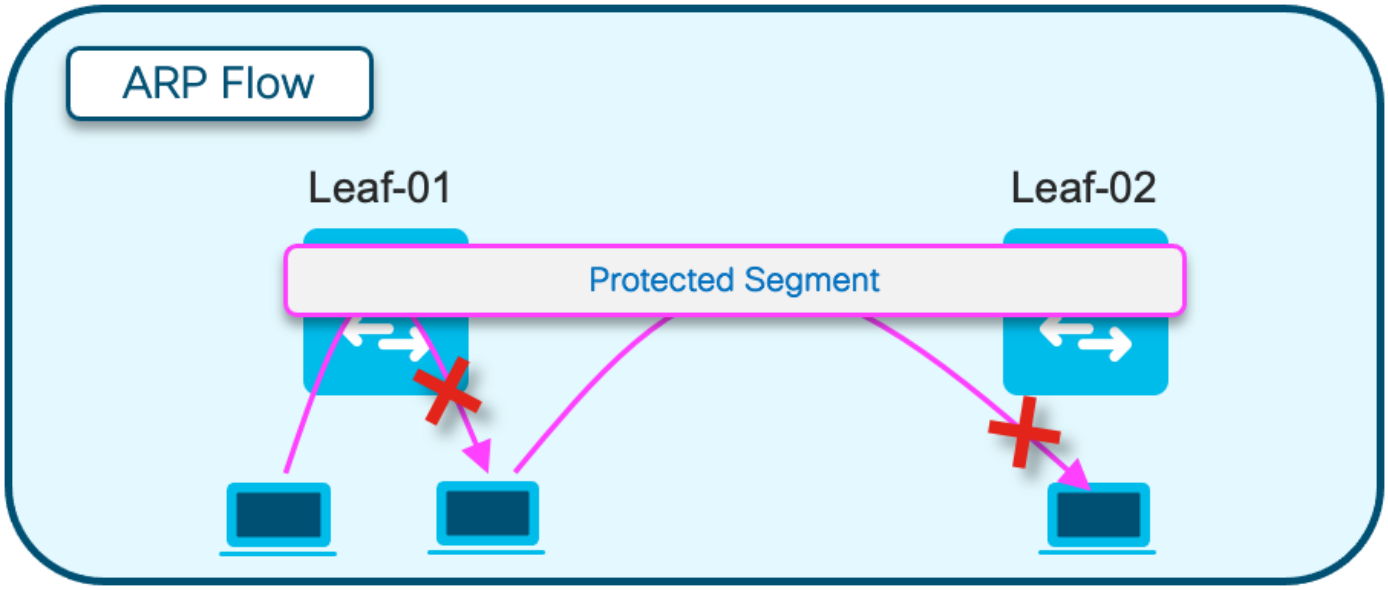


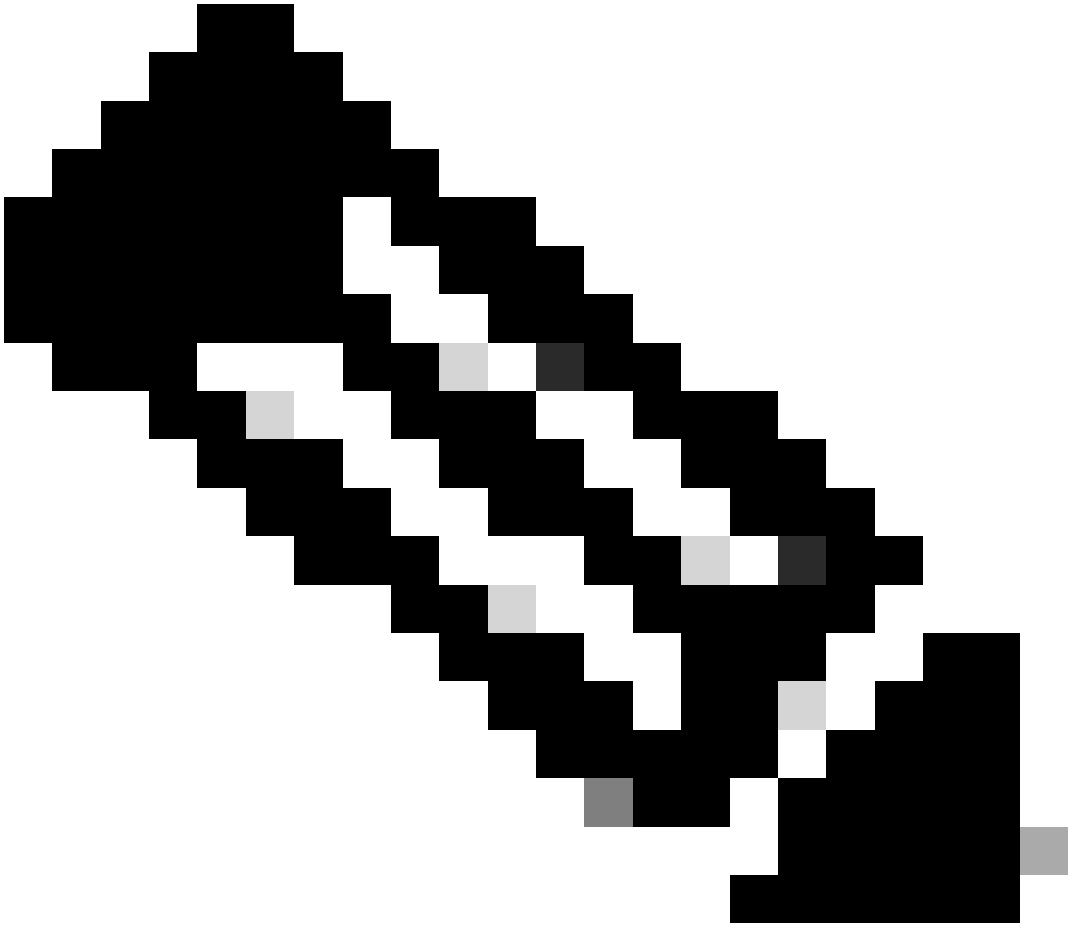
注意：需要使用集中星型广播来防止具有相同网段的枝叶之间直接相互发送广播。



地址解析(ARP)图

此图说明了ARP不允许到达同一EPVN网段中的任何主机。当另一台主机的主机ARP时，只有CGW获得此ARP并回复



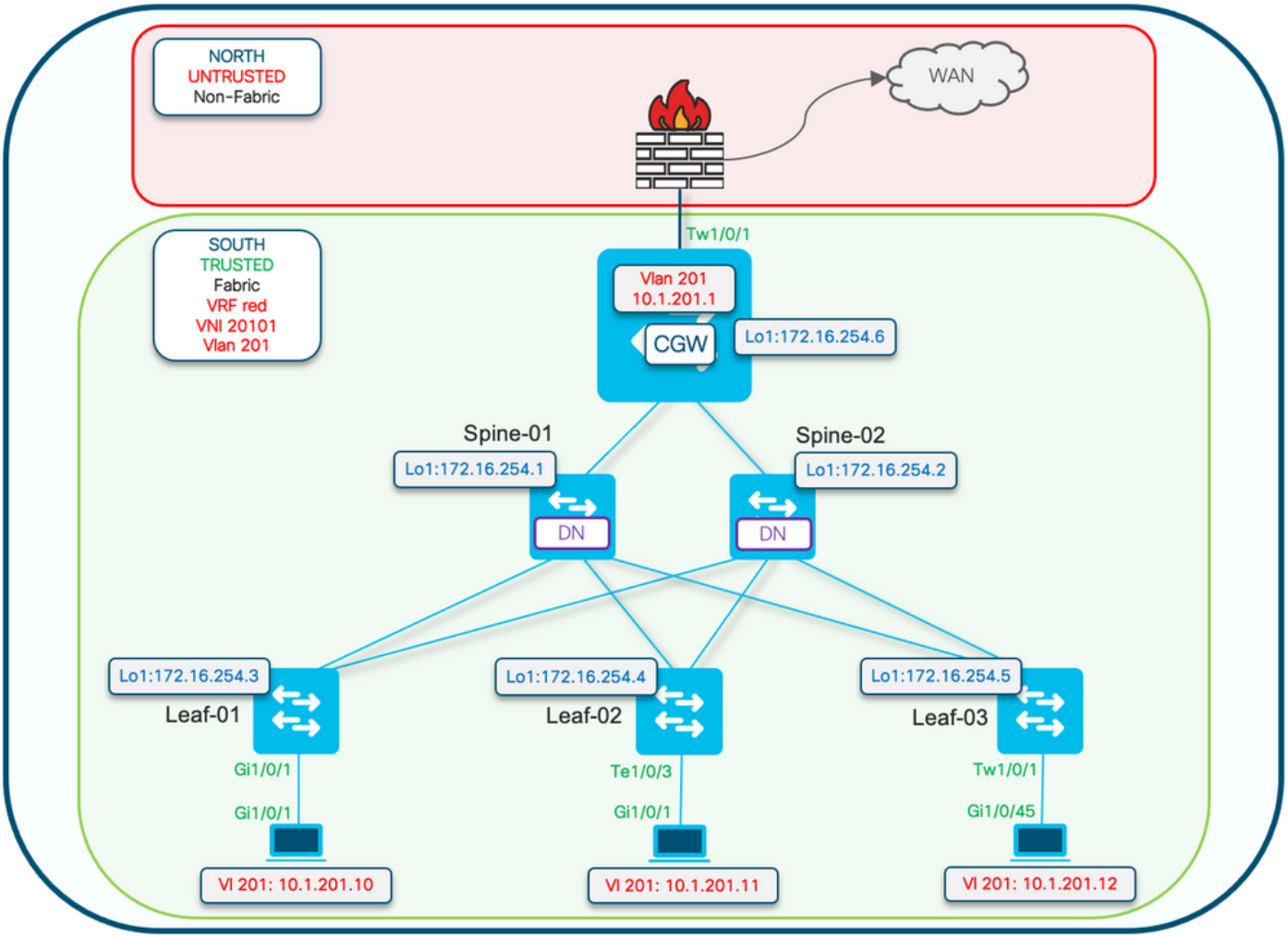


注意：此ARP行为更改通过使用“protected”关键字进行实例化。

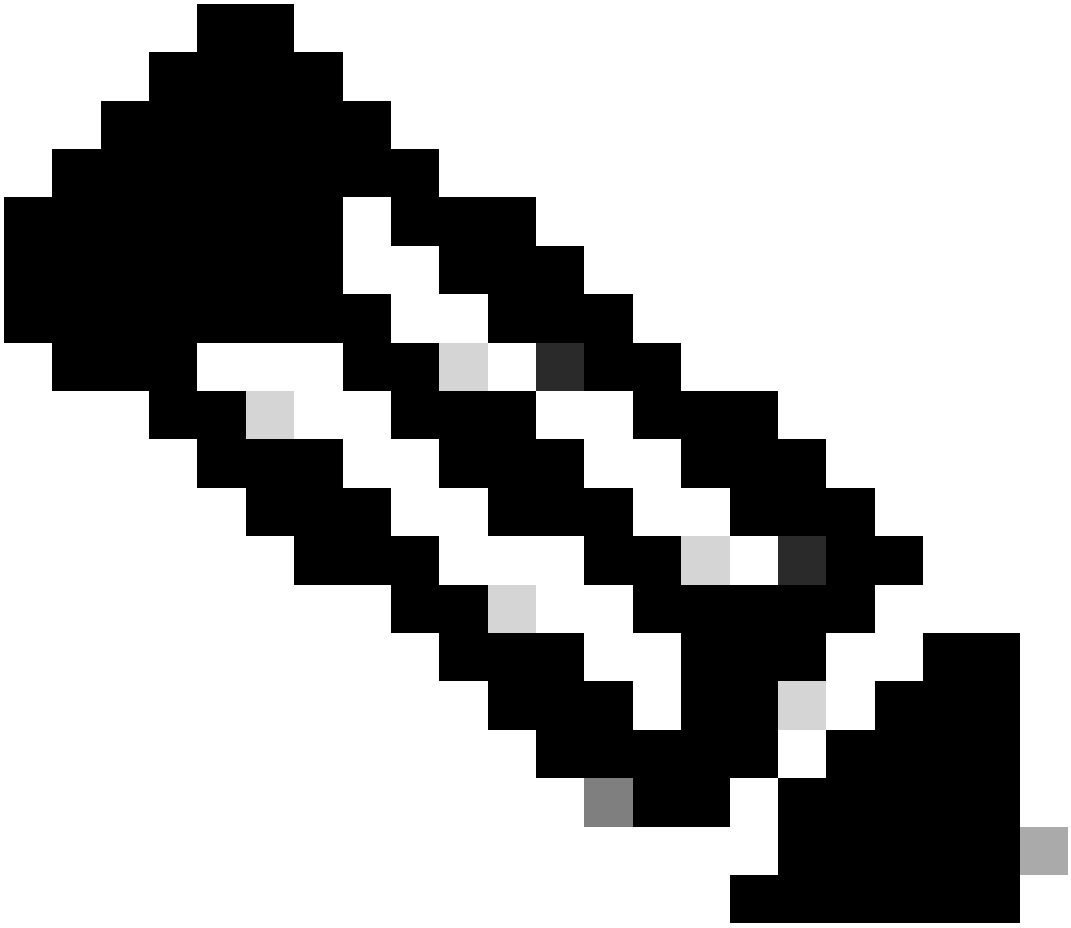
示例：成员evpn实例202 vni 20201 protected

配置（完全隔离）

网络图



受保护的配置关键字应用于枝叶交换机。CGW是混合设备，会安装所有mac地址。



注意：在Catalyst 9000系列交换机上实施BGP EVPN路由策略中显示了控制IMET前缀导入/导出的路由策略社区列表和路由映射配置。 本文档仅显示受保护的分段差异。

枝叶01（基本EVPN配置）

<#root>

```
Leaf-01#

show run | sec l2vpn
l2vpn evpn

 replication-type static

 flooding-suppression address-resolution disable  <-- Disables ARP caching so ARP is always sent up to t

 router-id Loopback1
l2vpn evpn

instance 201
```

```
 vlan-based
 encapsulation vxlan

replication-type ingress                          <-- Sets segment to use Unicast replication of BUM traff

 multicast advertise enable
```

<#root>

```
Leaf01#

show run | sec vlan config

vlan configuration 201
 member evpn-instance 201 vni 20101

protected  <-- protected keyword added
```

CGW（基本配置）

<#root>

```
CGW#

show running-config | beg l2vpn evpn instance 201

l2vpn evpn instance 201 vlan-based
 encapsulation vxlan
 replication-type ingress

 default-gateway advertise enable    <-- adds the BGP attribute EVPN DEF GW:0:0 to the MAC/IP prefix

 multicast advertise enable
```

<#root>

```
CGW#

show running-config | sec vlan config

vlan configuration 201
 member evpn-instance 201 vni 20101
```

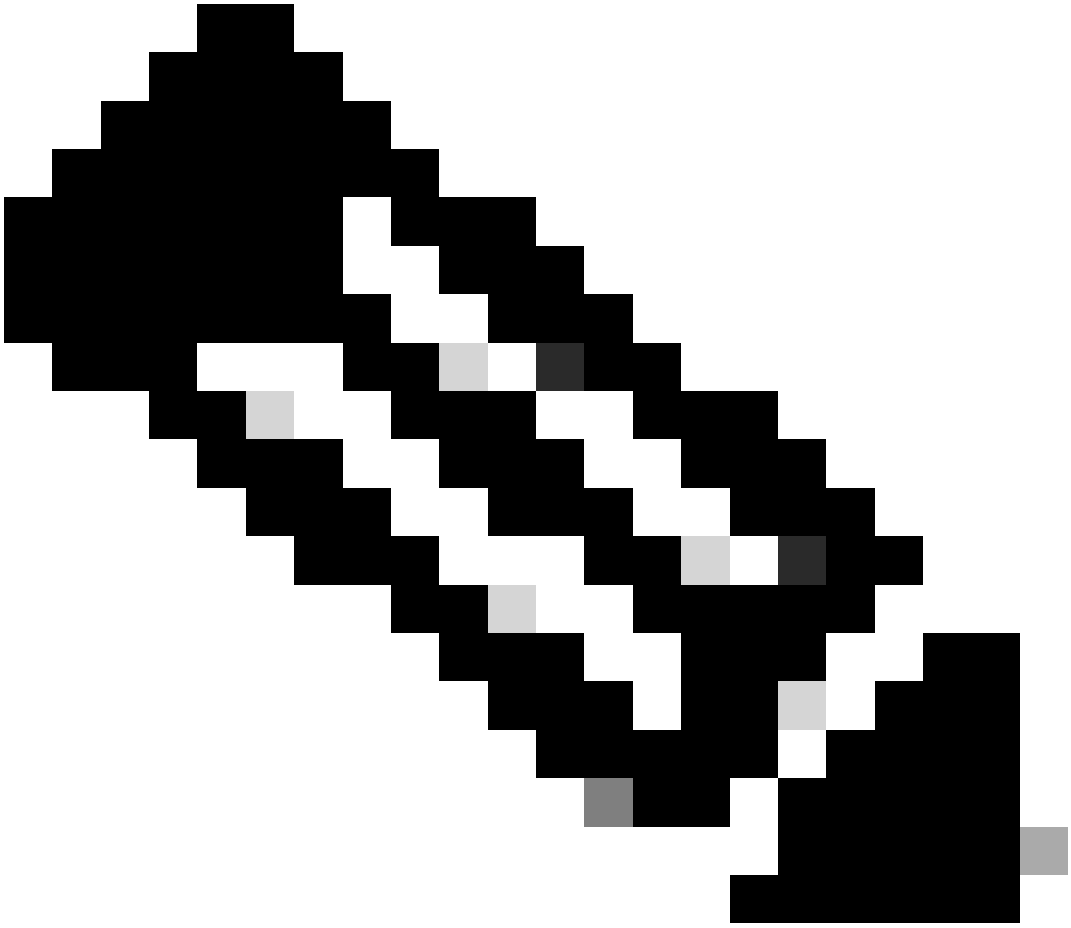
<#root>

```
CGW#

show run int nve 1

Building configuration...

Current configuration : 313 bytes
!
```

```
interface nve1
 no ip address
 source-interface Loopback1
 host-reachability protocol bgp
 member vni 20101 ingress-replication local-routing  <-- 'ingress-replication' (Unicast all BUM traffic)
```

<#root>

```
CGW#
show run interface vlan 201

Building configuration...

Current configuration : 231 bytes
!
interface Vlan201

mac-address 0000.beef.cafe          <-- MAC is static in this example for viewing simplicity. This is no
 vrf forwarding red                  <-- SVI is in VRF red
 ip address 10.1.201.1 255.255.255.0
 no ip redirects
 ip local-proxy-arp                 <-- Sets CGW to Proxy reply even for local subnet ARP requests
 ip pim sparse-mode
 ip route-cache same-interface      <-- This is auto added when local-proxy-arp is configured. However,
 ip igmp version 3
 no autostate
```



注意：在CGW上没有应用BGP策略。允许CGW接收和发送所有前缀类型(RT2、RT5/RT3)。

验证（完全隔离）

EVI详细信息

<#root>

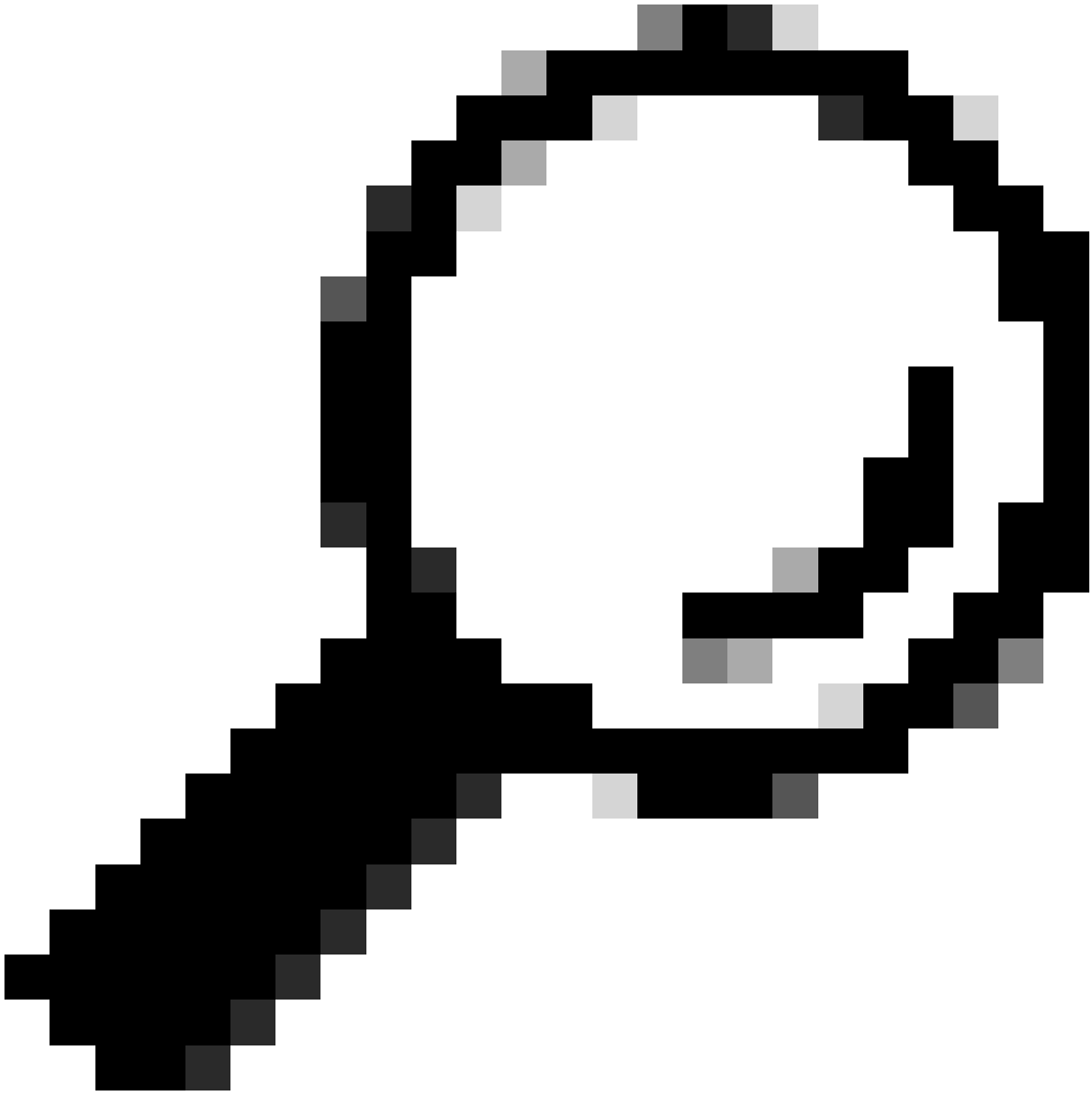
```
Leaf01#

sh l2vpn evpn evi 201 detail

EVPN instance:       201 (VLAN Based)
  RD:                172.16.254.3:201 (auto)
  Import-RTs:        65001:201
  Export-RTs:        65001:201
  Per-EVI Label:     none
  State:             Established
```

```
  Replication Type:  Ingress
  Encapsulation:     vxlan
  IP Local Learn:    Enabled (global)
  Adv. Def. Gateway: Disabled (global)
  Re-originate RT5:  Disabled
  Adv. Multicast:    Enabled
  AR Flood Suppress: Disabled (global)


Vlan:              201
    Protected:       True (local access p2p blocked)  <-- Vlan 201 is in protected mode

<...snip...>
```

本地RT2生成（本地主机到RT2）

验证从本地主机学习到RT2生成的组件依赖关系链：

- SISF（当枝叶没有SVI时，SISF仍会通过来自主机的ARP帧来收集主机信息）
- EVPN经理
- L2RIB
- 调试输出中显示“BGP

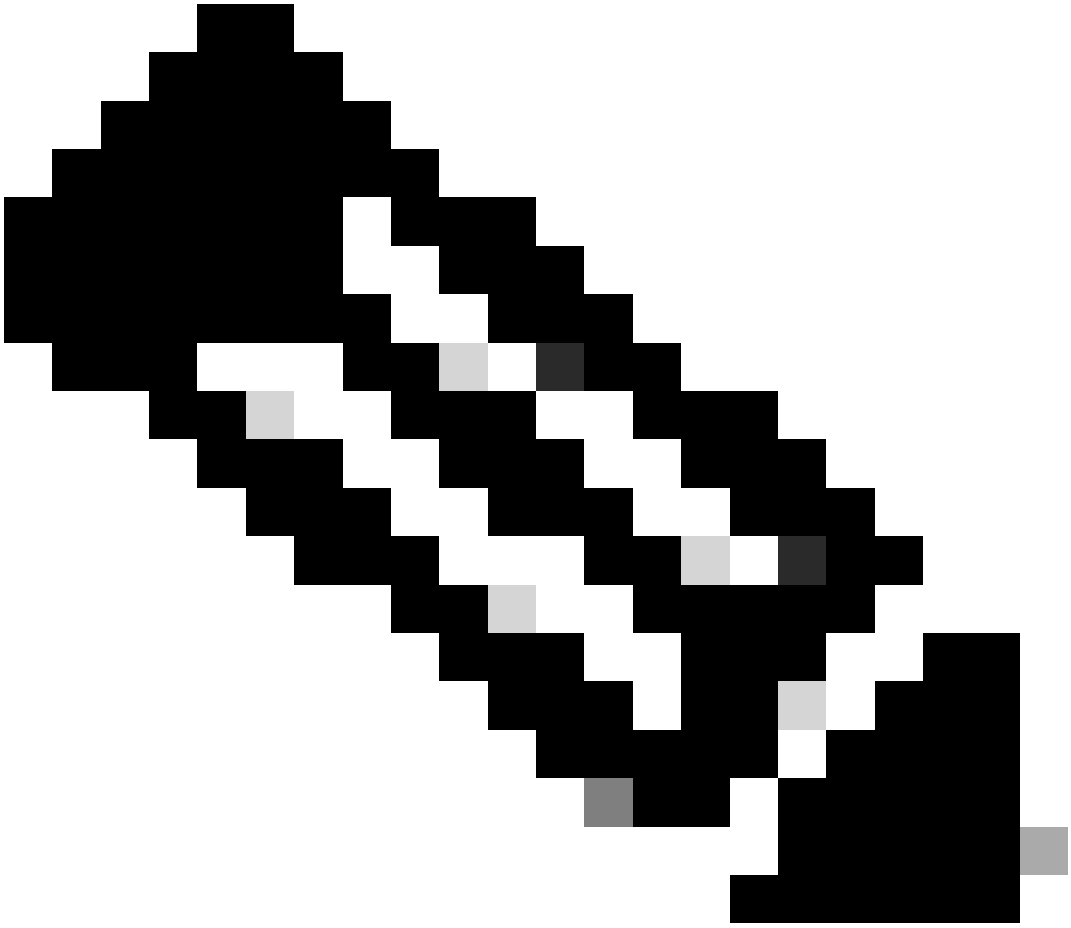


提示：如果之前的组件未正确编程，则整个依赖关系链会中断（例如：SISF没有en条目，则BGP无法创建RT2）。

SISF

验证SISF是否已在数据库中获知主机（从DHCP或ARP获知主机信息）

- SISF从IOS-MATM learning获取MAC条目，然后向上发送到EVPN管理器（必须使用策略“evpn-sisf-policy”进行MAC可访问）。
- SISF收集本地VTEP上的IP/MAC绑定并使用EVPN管理器，这些信息应编程为通过BGP到其他枝叶的/32路由。



注意：在此场景中，主机有一个静态IP，因此SISF使用ARP来收集主机详细信息。在大部分隔离部分中显示DHCP和DHCP监听。

<#root>

```
Leaf01#

show device-tracking database vlanid 201

vlanDB has 1 entries for vlan 201, 1 dynamic
Codes: L - Local, S - Static, ND - Neighbor Discovery, ARP - Address Resolution Protocol, DH4 - IPv4 DHC
Preflevel flags (prlvl):
0001:MAC and LLA match     0002:Orig trunk            0004:Orig access
0008:Orig trusted trunk    0010:Orig trusted access   0020:DHCP assigned
0040:Cga authenticated     0080:Cert authenticated    0100:Statically assigned


    Network Layer Address                  Link Layer Address     Interface  vlan       prlvl      age

ARP

 10.1.201.10
```

```
0006.f601.cd43


Gi1/0/1

    201        0005       3mn        REACHABLE  86 s

<-- Gleaned from local host ARP Request
```

EVPN管理器

EVPN Mgr学习本地MAC并将其安装到L2RIB中。EVPN Mgr也从L2RIB获取远程MAC，但条目仅用于处理MAC移动性

确认EVPN管理器已使用SISF条目更新

```
<#root>

Leaf01#

show l2vpn evpn mac evi 201

MAC Address    EVI   VLAN  ESI                       Ether Tag  Next Hop(s)
-------------- ----- ----- ------------------------- ---------- ---------------

0006.f601.cd43 201   201

0000.0000.0000.0000.0000 0

         Gi1/0/1:201    <-- MAC in VLan 201 local interface Gi1/0/1:service instance 201

<...snip...>
```

L2RIB

- L2RIB从EVPN管理器获取本地MAC并发送到BGP和L2FIB。
- L2RIB还负责从BGP学习远程MAC以更新EVPN管理器和L2FIB。
- L2RIB需要“本地”和“远程”，其他组件才能正确更新。
- L2RIB组件位于本地和远程MAC学习之间，具体取决于需要更新的方向/组件

验证L2RIB是否已使用EVPN管理器中的本地MAC更新

```
<#root>

Leaf01#

show l2route evpn mac topology 201      <-- View the overall topology for this segment

  EVI       ETag

Prod
```

```
    Mac Address                                          Next Hop(s) Seq Number
----- ---------- ----- -------------- ------------------------------------------------- ----------
  201          0
BGP
 0000.beef.cafe                                  V:20101 172.16.254.6          0
<-- produced by BGP who updated L2RIB (remote learn)
  201          0
L2VPN
 0006.f601.cd43                                          Gi1/0/1:201          0
<-- produced by EVPN Mgr who updated L2RIB (local learn)


Leaf01#
show l2route evpn mac mac-address 0006.f601.cd43 detail
EVPN Instance:            201
Ethernet Tag:             0
Producer Name:            L2VPN                <-- Produced by local
MAC Address:              0006.f601.cd43       <-- Host MAC Address
Num of MAC IP Route(s):   1
Sequence Number:          0
ESI:                      0000.0000.0000.0000.0000
Flags:                    B()
Next Hop(s):              Gi1/0/1:201 (E-LEAF) <-- Port:Instance and info about the Role (Leaf)
```

调试输出中显示“BGP

验证BGP是否由L2RIB更新

<#root>

```
Leaf01#
show bgp l2vpn evpn route-type 2 0 0006.f601.cd43 *
BGP routing table entry for [2][172.16.254.3:201][0][48][0006F601CD43][0][*]/20, version 268232
Paths: (1 available, best #1,
table evi_201
)
<-- In the totally isolated evi context
  Advertised to update-groups:
     2
  Refresh Epoch 1
  Local
```

```
 0.0.0.0 (via default) from 0.0.0.0

 (172.16.255.3)

<-- from 0.0.0.0 indicates local


      Origin incomplete, localpref 100, weight 32768, valid, sourced,

local

, best

<-- also indicates local


      EVPN ESI: 00000000000000000000, Label1 20101
      Extended Community: RT:65001:201 ENCAP:8

EVPN E-Tree:flag:1

,label:0

<-- EVPN e-Tree attribute with Leaf flag = 1 (added to indicate this is a host address)

      Local irb vxlan vtep:
        vrf:not found, l3-vni:0
        local router mac:0000.0000.0000
        core-irb interface:(not found)


 vtep-ip:172.16.254.3                                                    <-- Local VTEP Loopback

      rx pathid: 0, tx pathid: 0x0
      Updated on Sep 14 2023 20:16:17 UTC
```

远程RT2学习（默认网关RT2）

调试输出中显示“BGP

验证BGP已获取CGW RT2前缀

<#root>

```
Leaf01#

show bgp l2vpn evpn route-type 2 0 0000.beef.cafe 10.1.201.1

BGP routing table entry for [2][172.16.254.3:201][0][48][0000BEEFCAFE][32][10.1.201.1]/24, version 1141
Paths: (1 available, best #1,

table evi_201

)

<-- EVI context is 201

  Flag: 0x100
  Not advertised to any peer
  Refresh Epoch 2
  Local, imported path from [2][172.16.254.6:201][0][48][0000BEEFCAFE][32][10.1.201.1]/24 (global)
    172.16.254.6 (metric 3) (via default) from 172.16.255.1 (172.16.255.1)
```

```
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      EVPN ESI: 00000000000000000000,

Label1 20101                <-- Correct segment identifier

      Extended Community: RT:65001:201 ENCAP:8

EVPN DEF GW:0:0    <-- Default gateway attribute is added via the 'default gateway advertise CLI'

      Originator: 172.16.255.6, Cluster list: 172.16.255.1
      rx pathid: 0, tx pathid: 0x0
      Updated on Sep 1 2023 15:27:45 UTC
```

L2RIB

验证BGP更新的L2RIB

- L2RIB从EVPN管理器获取本地MAC并发送到BGP和L2FIB。L2RIB还负责从BGP学习远程MAC以更新EVPN管理器和L2FIB。
- L2RIB需要“本地”和“远程”，其他组件才能正确更新。
- L2RIB组件位于本地和远程MAC学习之间，具体取决于需要更新的方向和组件。

<#root>

```
Leaf01#

show l2route evpn default-gateway host-ip 10.1.201.1

  EVI       ETag  Prod    Mac Address                                Host IP
----- ---------- ----- -------------- ------------------------------------ ------------------------

201

          0

BGP


0000.beef.cafe


10.1.201.1

                              V:20101 172.16.254.6

<-- L2RIB has the MAC-IP of the Gateway programmed
```

L2FIB

在L2FIB中验证

- 负责用MAC更新FED以在硬件中进行编程的组件。

- L2FIB安装到FED-MATM中的远程MAC条目不会传送到IOS-MATM。（IOS-MATM仅显示本地MAC，而FED-MATM同时显示本地和远程MAC）。
- L2FIB输出仅显示远程MAC（它不负责对本地MAC进行编程）。

<#root>

```
Leaf01#

show l2fib bridge-domain 201 address unicast 0000.beef.cafe


MAC Address                   :

0000.beef.cafe                                          <-- CGW MAC

Reference Count               : 1
Epoch                         : 0

Producer                      : BGP                                                    <-- Learned fro

Flags                         : Static
Adjacency                     :

VXLAN_UC

  PL:2973(1) T:VXLAN_UC [MAC]20101:

172.16.254.6  <-- CGW Loopback IP

PD Adjacency                  : VXLAN_UC  PL:2973(1) T:VXLAN_UC [MAC]20101:172.16.254.6
Packets                       : 6979
Bytes                         : 0
```

FED

在FED MATM中验证

- 在配置了“protected关键字”的枝叶的硬件级别，您应该只能看到CGW默认网关MAC和本地主机MAC。
- 交换机查看DEF GW属性的RT2前缀，以确定哪些远程MAC适合安装。

<#root>

```
Leaf01#

show platform software fed switch active matm macTable vlan 201

VLAN   MAC

Type

  Seq#    EC_Bi  Flags  machandle           siHandle            riHandle            diHandle

Con

-------------------------------------------------------------------------------------------------

201    0000.beef.cafe
```

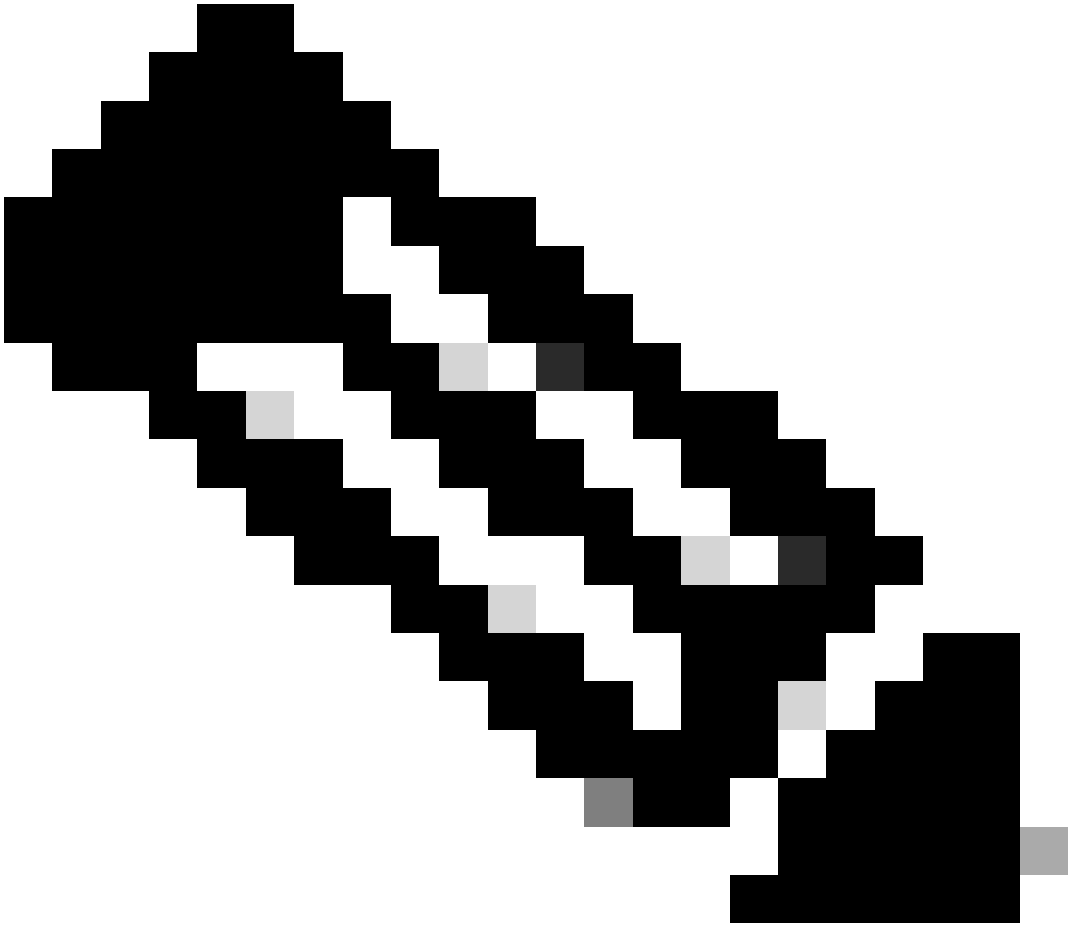
```
0x5000001

       0      0     64  0x7a199d182498      0x7a199d183578

0x71e059173e08

      0x0                          0       82

VTEP 172.16.254.6

 adj_id 9

No
<-- Only remote MAC installed in Fed is the Default Gateway (0x5000001 type) Conn = No (meaning not dire

201    0006.f601.cd01

0x1

    2458      0      0  0x7a199d1a2248      0x7a199d19eef8      0x0                 0x7a199c6f7cd8

201    0006.f601.cd43        0x1    8131      0      0  0x7a199d195a98      0x7a199d19eef8      0x0

<-- Two local MAC addresses (0x1 type) Conn = Yes (directly connected)



Total Mac number of addresses:: 5
Summary:
Total number of secure addresses:: 0
Total number of drop addresses:: 0
Total number of lisp local addresses:: 0
Total number of lisp remote addresses:: 3
*a_time=aging_time(secs)  *e_time=total_elapsed_time(secs)
Type:

MAT_DYNAMIC_ADDR           0x1

  MAT_STATIC_ADDR            0x2  MAT_CPU_ADDR               0x4  MAT_DISCARD_ADDR           0x8
MAT_ALL_VLANS             0x10  MAT_NO_FORWARD            0x20  MAT_IPMULT_ADDR           0x40  MAT_RESV
MAT_DO_NOT_AGE           0x100  MAT_SECURE_ADDR          0x200  MAT_NO_PORT              0x400  MAT_DROP
MAT_DUP_ADDR            0x1000  MAT_NULL_DESTINATION    0x2000  MAT_DOT1X_ADDR          0x4000  MAT_ROUT
MAT_WIRELESS_ADDR      0x10000  MAT_SECURE_CFG_ADDR    0x20000  MAT_OPQ_DATA_PRESENT   0x40000  MAT_WIRE
MAT_DLR_ADDR          0x100000  MAT_MRP_ADDR          0x200000  MAT_MSRP_ADDR         0x400000  MAT_LISP

MAT_LISP_REMOTE_ADDR 0x1000000

  MAT_VPLS_ADDR        0x2000000

MAT_LISP_GW_ADDR     0x4000000

<-- the addition of these values = 0x5000001

MAT_LISP_REMOTE_ADDR 0x1000000
MAT_LISP_GW_ADDR 0x4000000
MAT_DYNAMIC_ADDR 0x1
```

数据平面邻接

确认FED条目后的最后一步是解析重写索引(RI)

<#root>

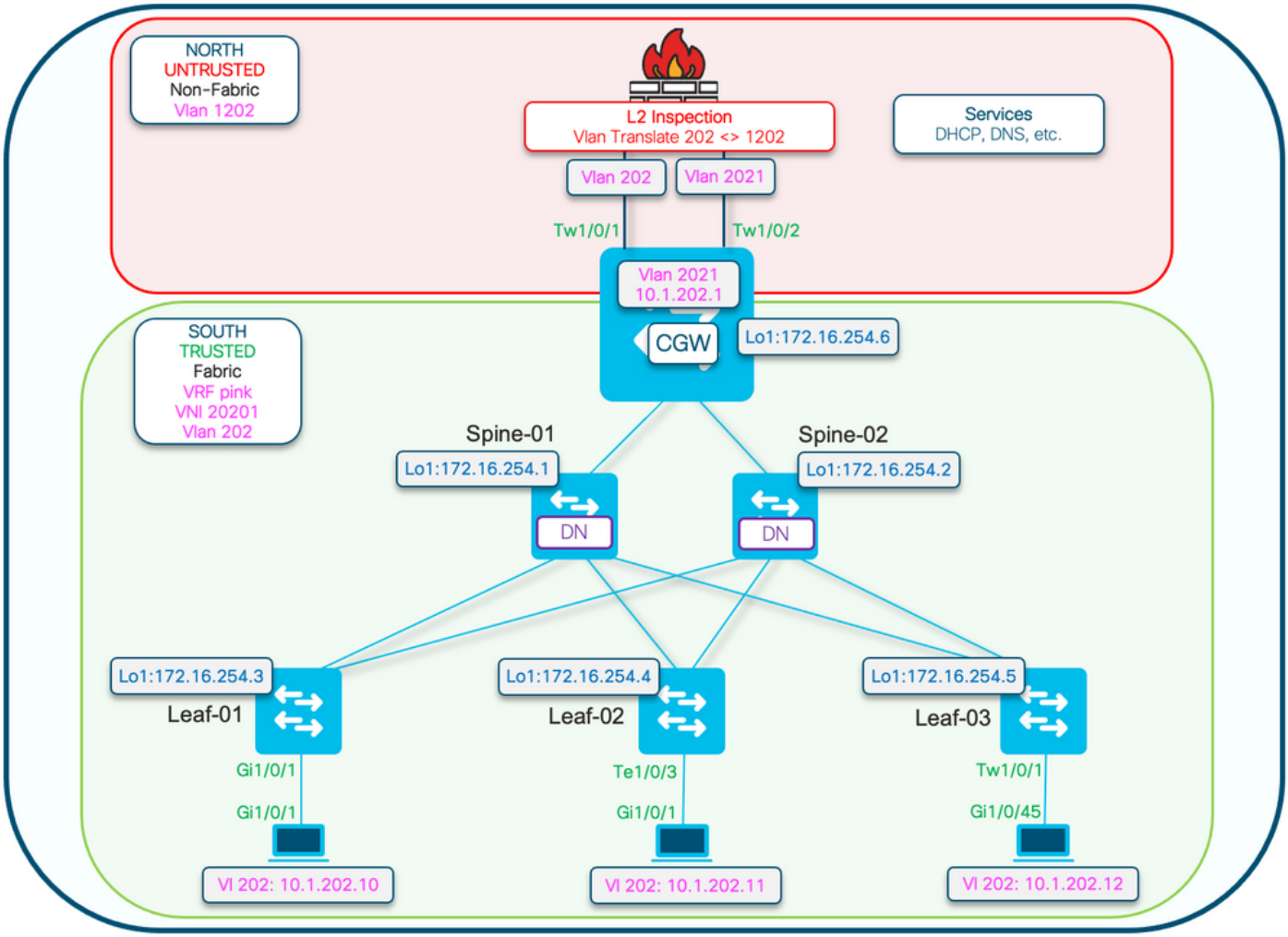
```
Leaf01#

sh platform hardware fed switch active fwd-asic abstraction print-resource-handle 0x71e059173e08 0
<-- 0x71e059173e08 is taken from previous FED command riHandle for the CGW MAC

Handle:0x71e059173e08 Res-Type:ASIC_RSC_RI Res-Switch-Num:255 Asic-Num:255 Feature-ID:AL_FID_L2_WIRELESS
priv_ri/priv_si Handle: 0x71e05917b8d8Hardware Indices/Handles: index0:0x38  mtu_index/l3u_ri_index0:0x0
Features sharing this resource:58 (1)]


Brief Resource Information (ASIC_INSTANCE# 0)
---------------------------------------
ASIC#:0 RI:56 Rewrite_type:AL_RRM_REWRITE_LVX_IPV4_L2_PAYLOAD_ENCAP_EPG(116) Mapped_rii:LVX_L3_ENCAP_L2_

 Src IP:     172.16.254.3      <-- source tunnel IP
 Dst IP:     172.16.254.6      <-- dest tunnel IP

 iVxlan dstMac:     0x9db:0x00:0x00
 iVxlan srcMac:     0x00:0x00:0x00
 IPv4 TTL:     0
 iid present:     0

 lisp iid:     20101            <-- Segment 20101

 lisp flags:     0

 dst Port:     4789             <-- VxLAN

 update only l3if:     0
 is Sgt:     0
 is TTL Prop:     0
 L3if LE:     53 (0)
 Port LE:     281 (0)
 Vlan LE:     8 (0)
```

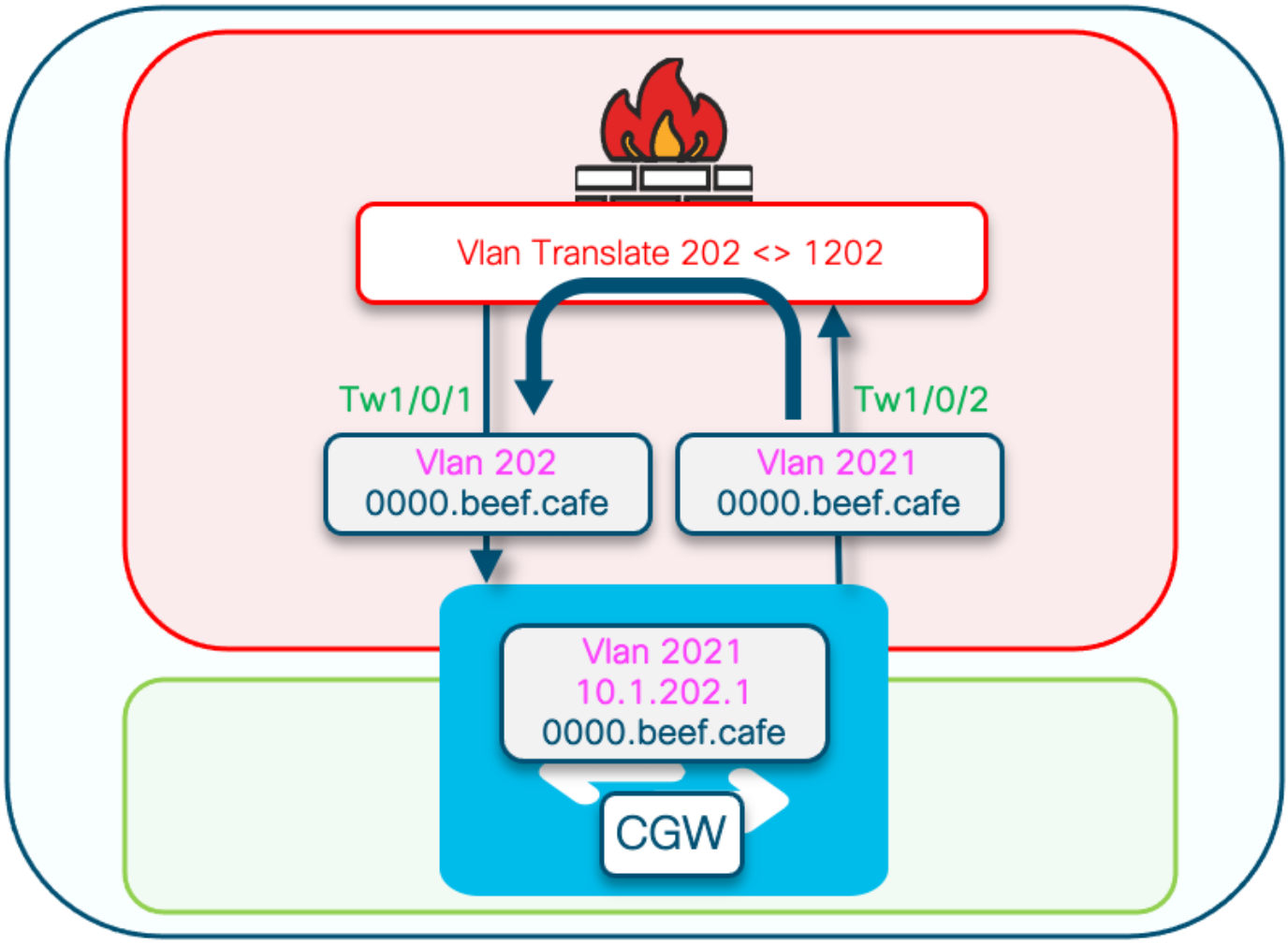


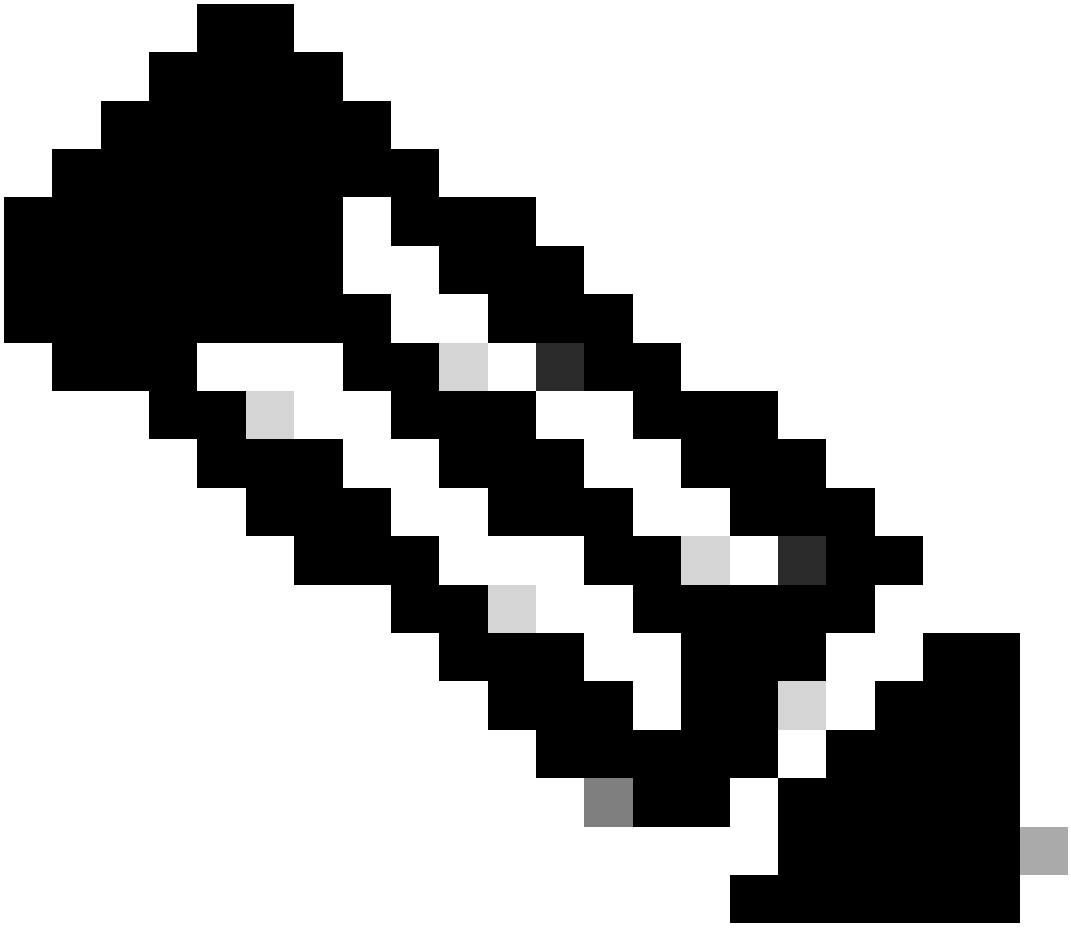
注意：您也可以使用“show platform software fed switch active matm macTable vlan 201 detail”，它将此命令与FED命令链接到一个结果中

配置（部分隔离）

网络图


NORTH
UNTRUSTED
Non-Fabric
Vlan 1202
L2 Inspection
Vlan Translate 202 <> 1202
Services
DHCP, DNS, etc.
Vlan 202
Vlan 2021
Tw1/0/1
Tw1/0/2
Vlan 2021
10.1.202.1
CGW
Lo1:172.16.254.6
SOUTH
TRUSTED
Fabric
VRF pink
VNI 20201
Vlan 202
Spine-01
Lo1:172.16.254.1
DN
Spine-02
Lo1:172.16.254.2
DN
Lo1:172.16.254.3
Leaf-01
Lo1:172.16.254.4
Leaf-02
Lo1:172.16.254.5
Leaf-03
Gi1/0/1
Gi1/0/1
Te1/0/3
Gi1/0/1
Tw1/0/1
Gi1/0/45
Vl 202: 10.1.202.10
Vl 202: 10.1.202.11
Vl 202: 10.1.202.12


Vlan Translate 202 <> 1202
Tw1/0/1
Tw1/0/2
Vlan 202
0000.beef.cafe
Vlan 2021
0000.beef.cafe
Vlan 2021
10.1.202.1
0000.beef.cafe
CGW



注意：本部分仅介绍与完全隔离网段的区别。

- Routing-policy，用DEF GW属性标记GCW网关MAC IP
- 需要自定义设备跟踪策略来防止MAC摆动
- GW MAC IP的静态设备跟踪绑定

枝叶01（基本EVPN配置）

<#root>

```
Leaf-01#

show run | sec l2vpn
l2vpn evpn

 replication-type static

 flooding-suppression address-resolution disable  <-- Disables ARP caching so ARP is always sent up to t

 router-id Loopback1
```

```
l2vpn evpn
instance 202
 vlan-based
 encapsulation vxlan

replication-type ingress
 multicast advertise enable
```

<#root>

```
Leaf01#
show run | sec vlan config

vlan configuration 202
 member evpn-instance 202 vni 20201
protected <-- protected keyword added
```

CGW（基本配置）

在nve下设置复制模式

<#root>

```
CGW#
show run int nve 1

Building configuration...

Current configuration : 313 bytes
!
interface nve1
 no ip address
 source-interface Loopback1
 host-reachability protocol bgp
 member vni 20201 ingress-replication local-routing <-- 'ingress-replication' (Unicast all BUM traffic)
end
```

配置外部网关SVI

<#root>

```
CGW#
show run interface vlan 2021
```

```
Building configuration...

Current configuration : 231 bytes
!
interface Vlan2021

mac-address 0000.beef.cafe          <-- MAC is static in this example for viewing simplicity. This is no

 vrf forwarding pink                  <-- SVI is in VRF pink

 ip address 10.1.202.1 255.255.255.0
 no ip redirects

 ip local-proxy-arp                 <-- Sets CGW to Proxy reply even for local subnet ARP requests

 ip pim sparse-mode

 ip route-cache same-interface      <-- This is auto added when local-proxy-arp is configured. However,

 ip igmp version 3
 no autostate
end
```

创建禁用收集功能的策略

<#root>

```
device-tracking policy dt-no-glean

 <-- Configure device tracking policy to prevent MAC-IP flapping

 security-level glean
 no protocol ndp
 no protocol dhcp6
 no protocol arp
 no protocol dhcp4
```

连接到externalgatewayevi/vlan

<#root>

```
CGW#

show running-config | sec vlan config

vlan configuration 202
 member evpn-instance 202 vni 20201

 device-tracking attach-policy dt-no-glean <-- apply the new device tracking policy to the vlan configur
```

为externalgateway mac-ip添加静态条目到设备跟踪表中

<#root>

```
device-tracking binding vlan 202 10.1.202.1 interface TwentyFiveGigE1/0/1 0000.beef.cafe
```

<-- All static entries in device tracking table should be for external gateway mac-ip's.
If there is any other static entry in device tracking table, match ip/ipv6 configurations in route map

创建BGP路由映射以匹配RT2 MAC-IP前缀并设置默认网关extendedcommunity

<#root>

```
route-map CGW_DEF_GW permit 10
```

match evpn route-type 2-mac-ip <-- match RT2 type MAC-IP

set extcommunity default-gw <-- Set Default-gateway (DEF GW 0:0) extended community

```
route-map CGW_DEF_GW permit 20
```

将路由映射应用到BGP路由反射器邻居

<#root>

CGW#

sh run | s r bgp

```
address-family l2vpn evpn
 neighbor 172.16.255.1 activate
 neighbor 172.16.255.1 send-community both
 neighbor 172.16.255.1
```

route-map CGW_DEF_GW out <-- Sets the DEF GW Community when it advertises MAC-IP type RT2 to the RR

```
 neighbor 172.16.255.2 activate
 neighbor 172.16.255.2 send-community both
 neighbor 172.16.255.2
```

route-map CGW_DEF_GW out <-- Sets the DEF GW Community when it advertises MAC-IP type RT2 to the RR

验证（部分隔离）

EVI详细信息

<#root>

```
Leaf01#

show l2vpn evpn evi 202 detail

EVPN instance:       202 (VLAN Based)
  RD:                172.16.254.3:202 (auto)
  Import-RTs:        65001:202
  Export-RTs:        65001:202
  Per-EVI Label:     none
  State:             Established
  Replication Type:  Ingress
  Encapsulation:     vxlan
  IP Local Learn:    Enabled (global)
  Adv. Def. Gateway: Enabled (global)
  Re-originate RT5:  Disabled
  Adv. Multicast:    Enabled

  Vlan:              202
    Protected:       True (local access p2p blocked)  <-- Vlan 202 is in protected mode

<...snip...>
```

本地RT2生成（本地主机到RT2）

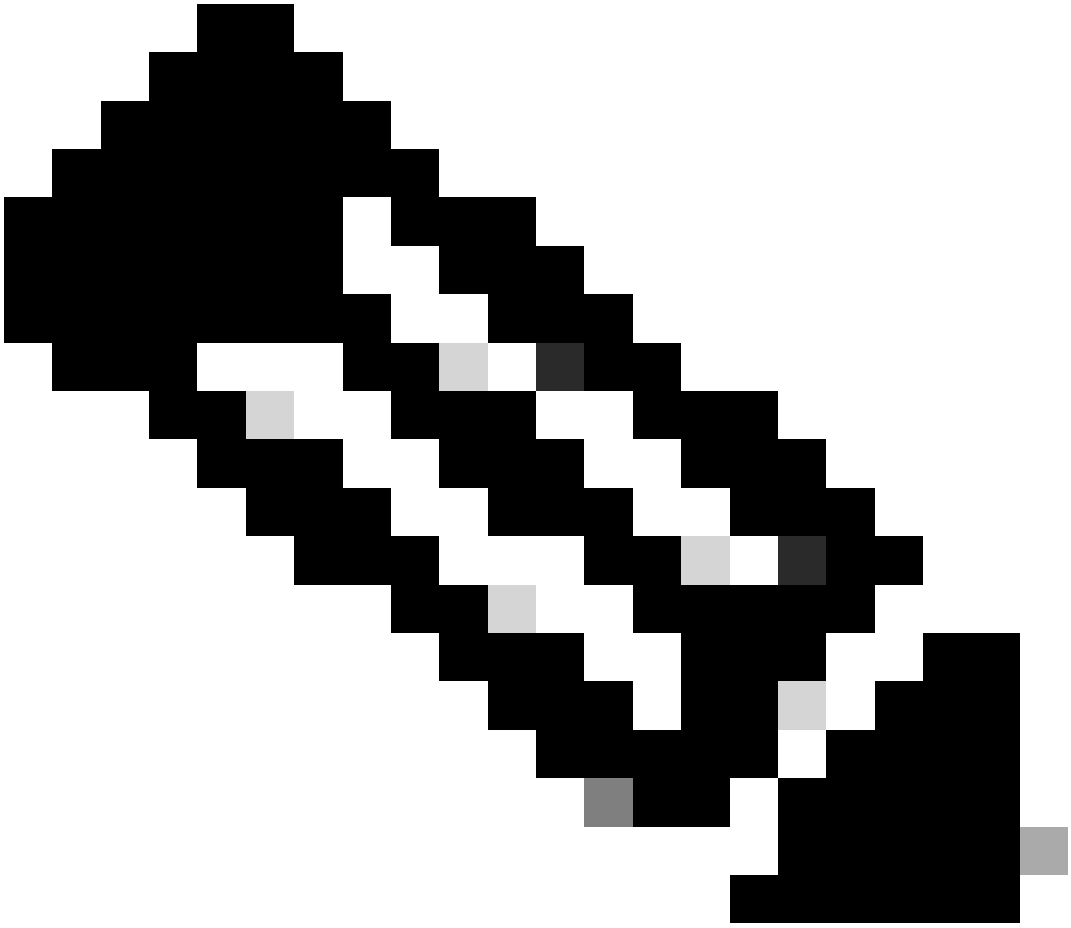
在前面的完全隔离的示例中介绍

远程RT2学习（默认网关RT2）

涵盖与完全隔离的区别

CGW默认网关前缀（枝叶）

检查前缀是否具有适当的属性，以便有资格安装到硬件中



注意：这对DHCP L2中继正常运行至关重要

<#root>

Leaf01#

show bgp l2vpn evpn route-type 2 0 0000.beef.cafe 10.1.202.1

BGP routing table entry for [2][172.16.254.3:202][0][48][0000BEEFCAFE][32][10.1.202.1]/24, version 1846
Paths: (1 available, best #1,

table evi_202

)

<-- the EVI context of 202 which matches the Vlan/EVI we are concerned about

  Not advertised to any peer
  Refresh Epoch 2
  Local, imported path from [2][172.16.254.6:202][0][48][0000BEEFCAFE][32][10.1.202.1]/24 (global)
    172.16.254.6 (metric 3) (via default) from 172.16.255.1 (172.16.255.1)
      Origin incomplete, metric 0, localpref 100, valid, internal, best

```
      EVPN ESI: 00000000000000000000,

Label1 20201              <-- Correct Segment ID

      Extended Community: RT:65001:202 ENCAP:8

EVPN DEF GW:0:0  <-- prefix has the Default GW attribute added

      Originator: 172.16.255.6, Cluster list: 172.16.255.1
      rx pathid: 0, tx pathid: 0x0
      Updated on Sep 7 2023 19:56:43 UTC
```

FED MATM (枝叶)

<#root>

```
F241.03.23-9300-Leaf01#

show platform software fed active matm macTable vlan 202 mac 0000.beef.cafe

VLAN   MAC                  Type  Seq#    EC_Bi  Flags  machandle           siHandle            riHandl
-------------------------------------------------------------------------------------------------------

202    0000.beef.cafe

  0x5000001       0      0     64  0x71e058da7858      0x71e05916c0d8      0x71e059171678      0x0

VTEP 172.16.254.6

 adj_id 651

No
<-- MAC of Default GW is installed in FED
```

SISF (CGW)

<#root>

```
CGW#

sh device-tracking database vlanid 202

vlanDB has 1 entries for vlan 202, 0 dynamic
Codes: L - Local, S - Static, ND - Neighbor Discovery, ARP - Address Resolution Protocol, DH4 - IPv4 DH
Preflevel flags (prlvl):
0001:MAC and LLA match     0002:Orig trunk            0004:Orig access
0008:Orig trusted trunk    0010:Orig trusted access   0020:DHCP assigned
0040:Cga authenticated     0080:Cert authenticated    0100:Statically assigned


    Network Layer Address                  Link Layer Address     Interface  vlan       prlvl      ag
S   10.1.202.1                             0000.beef.cafe         Twe1/0/1   202        0100       13
```

IOS MATM (CGW)

<#root>

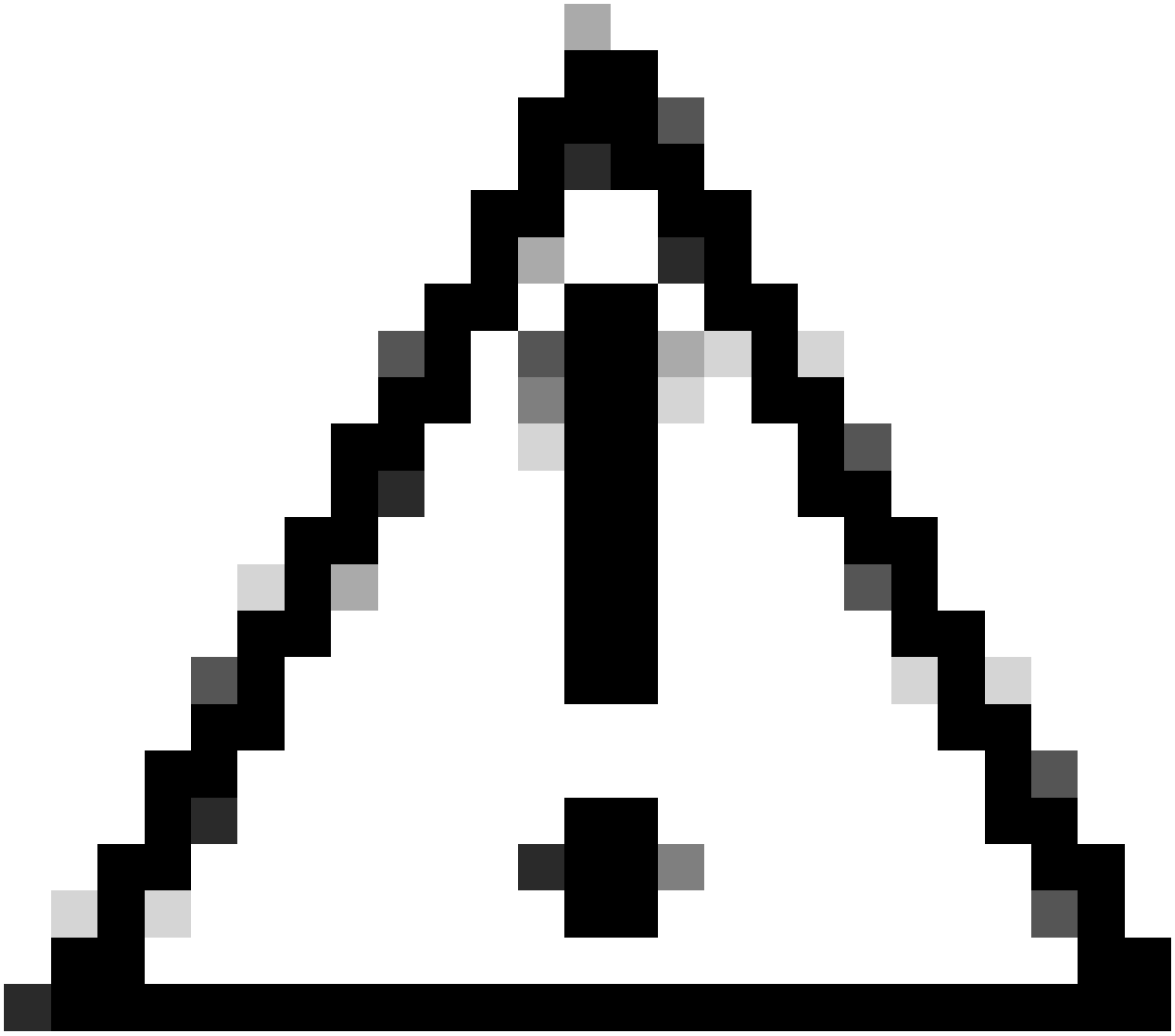
```
CGW#

show mac address-table address 0000.beef.cafe

          Mac Address Table
-------------------------------------------

Vlan    Mac Address       Type        Ports
----    -----------       --------    -----
 201    0000.beef.cafe    STATIC      Vl201

2021    0000.beef.cafe    STATIC      Vl2021   <-- The Vlan 2021 SVI MAC advertised out Tw1/0/1

 202    0000.beef.cafe    DYNAMIC     Twe1/0/1 <-- The Vlan 2021 SVI MAC learned dynamically after passi
```

故障排除

地址解析(ARP)

隔离ARP问题的常规步骤

- 确认IMET隧道已就绪
- 在CGW上行链路上捕获，以验证从枝叶交换机封装的ARP是否收到
- 如果看不到ARP到达上行链路上的封装
  - 验证枝叶和CGW上的IMET隧道是否已就绪
  - 在枝叶上行链路上进行捕获，以确认ARP已封装并发送
  - 排除中间路径故障
- 如果ARP到达边界IMET隧道捕获，但未在VRF ARP表中编程。
  - 排除CPU/CoPP传送路径故障，确认ARP传送到CPU
  - 确认IP地址/客户端信息是否正确。
  - 调试VRF中的ARP以查看可能影响ARP过程的内容
- 验证作为下一跳/目标mac安装在主机上的CGW MAC
- 确认CGW具有包含实际主机MAC的两个ARP条目
- 验证防火墙策略是否允许此类流量



注意：启用调试时请小心！

确保您已禁用泛洪抑制

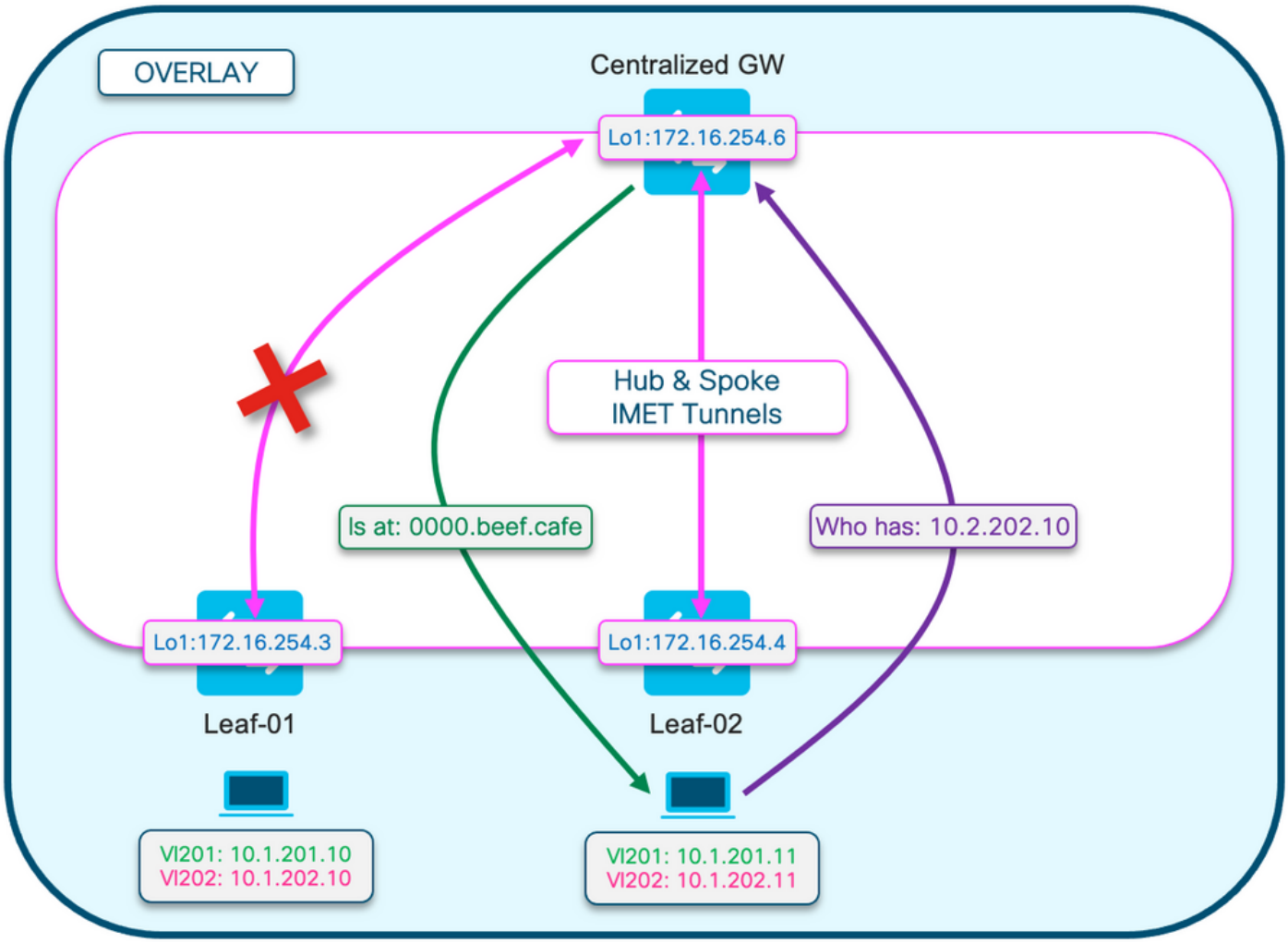
<#root>

```
Leaf-01#

show run | sec l2vpn
l2vpn evpn

 replication-type static

 flooding-suppression address-resolution disable <-- This CLI prevents a VTEP from trying to unicast oth
```

当枝叶02上的主机解析枝叶01上的主机的ARP时，不会将ARP请求直接广播到枝叶01

- 相反，ARP会向上传递在Leaf-02上编程的指向CGW的唯一BUM隧道

- CGW不会将此转发到Leaf-01，而是使用自己的MAC进行应答
- 这会导致所有通信向上传递到CGW，然后路由到主机之间
- CGW路由数据包，即使它们位于同一本地子网中



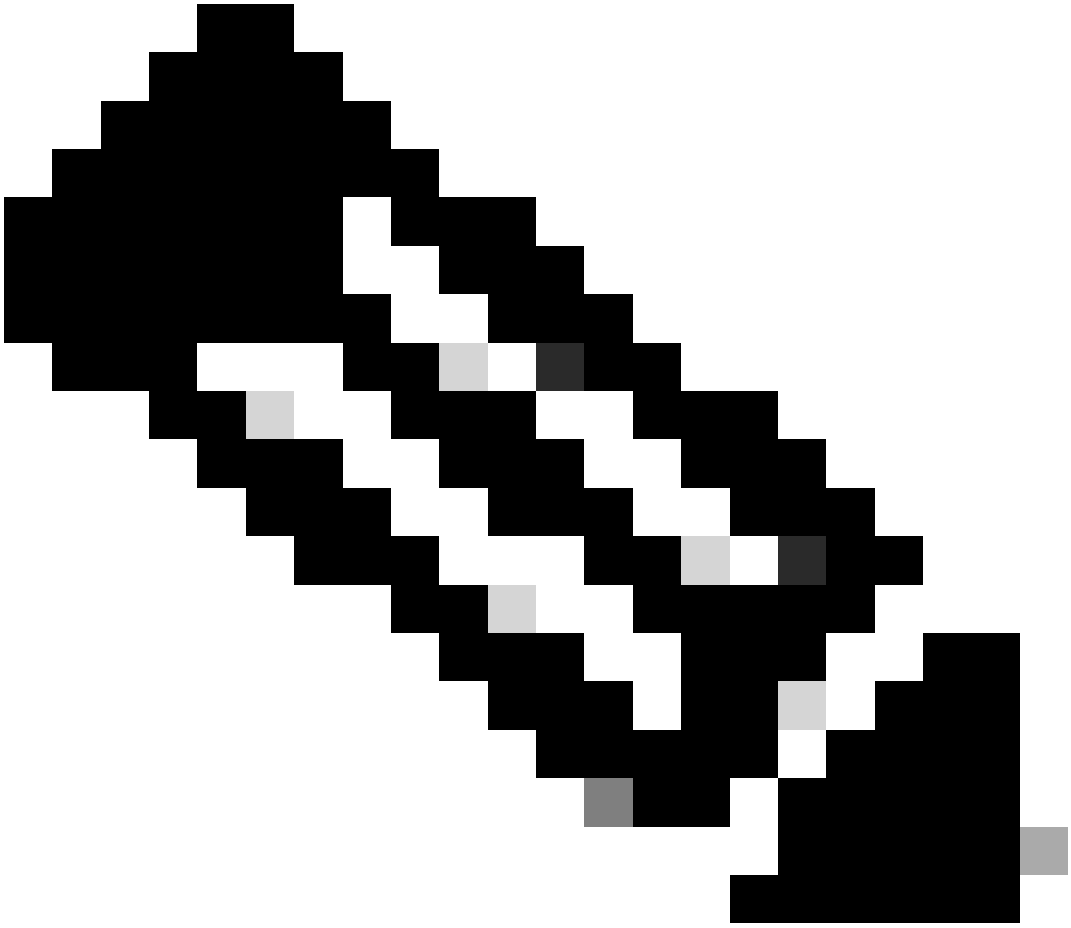
此图有助于直观显示本部分所描述的ARP解析过程。

ARP请求显示为紫色

- 此ARP请求用于解析主机10.1.202.10的MAC地址Leaf-01
- 请注意，紫色线路在CGW终止，并且未到达枝叶01

ARP应答显示为绿色

- 回复包含Vlan 202的CGW SVI的MAC
- 请注意，绿线来自CGW，而不是来自实际主机



注意：红色X表示此通信不涉及向枝叶01发送流量。

观察每台相应主机上的ARP条目

<#root>

Leaf02-HOST#

sh ip arp 10.1.202.10

Protocol Address   Age (min) Hardware Addr   Type   Interface
Internet 10.1.202.10   1

0000.beef.cafe

  ARPA   Vlan202

<-- MAC address for Leaf01 host is CGW MAC

Leaf01-HOST#

sh ip arp 10.1.202.11

```
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  10.1.202.11             7
```

0000.beef.cafe

```
  ARPA   Vlan202
```

<-- MAC address for Leaf02 host is CGW MAC

观察在CGW上如何获知RT2前缀。CGW路由数据包时需要此步骤

<#root>

CGW#

sh bgp l2vpn evpn route-type 2 0 0006.f617.eec4 * <-- Leaf02 actual MAC

BGP routing table entry for [2][172.16.254.6:202][0][48][0006F617EEC4][0][*]/20, version 235458
Paths: (1 available, best #1,

table evi_202

)

```
  Not advertised to any peer
  Refresh Epoch 2
  Local, imported path from [2][172.16.254.4:202][0][48][0006F617EEC4][0][*]/20 (global)
    172.16.254.4 (metric 3) (via default) from 172.16.255.1 (172.16.255.1)
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      EVPN ESI: 00000000000000000000,
```

Label1 20201 <-- correct segment identifier

```
      Extended Community: RT:65001:202 ENCAP:8
```

EVPN E-Tree:flag:1

,label:0

<-- prefix contains the Leaf flag indicating this is a normal host

```
      Originator: 172.16.255.4, Cluster list: 172.16.255.1
      rx pathid: 0, tx pathid: 0x0
      Updated on Apr 9 2025 17:11:22 UTC
```

CGW#

sh bgp l2vpn evpn route-type 2 0 0006.f601.cd44 * <-- Leaf01 actual MAC

BGP routing table entry for [2][172.16.254.6:202][0][48][0006F601CD44][0][*]/20, version 235521
Paths: (1 available, best #1,

table evi_202)

```
  Not advertised to any peer
  Refresh Epoch 2
  Local, imported path from [2][172.16.254.3:202][0][48][0006F601CD44][0][*]/20 (global)
    172.16.254.3 (metric 3) (via default) from 172.16.255.1 (172.16.255.1)
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      EVPN ESI: 00000000000000000000,
```

```
Label1 20201                          <-- correct segment identifier

      Extended Community: RT:65001:202 ENCAP:8

EVPN E-Tree:flag:1

,label:0

<-- prefix contains the Leaf flag indicating this is a normal host


      Originator: 172.16.255.3, Cluster list: 172.16.255.1
      rx pathid: 0, tx pathid: 0x0
      Updated on Apr 9 2025 17:17:06 UTC
```

捕获上行链路上的ARP交换以确认双向通信

- 您可以在交换矩阵上行链路上使用嵌入式数据包捕获(EPC)
- 此场景显示枝叶01上行链路上的EPC。如有必要，在CGW上重复此相同过程

配置EPC

```
<#root>

Leaf01#

monitor capture 1 interface range te 1/1/2 , te 1/1/4 both match any buffer size 100


<-- both Uplinks toward fabric included
```

开始捕获

```
<#root>

Leaf01#

monitor capture 1 start
```

启动ping以触发ARP请求（在本例中，ping是从Leaf01主机10.1.201.10到Leaf02主机10.1.201.11）

```
<#root>

Leaf01-HOST#

ping vrf red 10.1.201.11


Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 10.1.201.11, timeout is 2 seconds:
```

```
...!!
Success rate is 40 percent (2/5), round-trip min/avg/max = 1/1/1 ms
```

停止捕获并检查ARP帧

<#root>

```
Leaf01#
mon cap 1 stop

F241.03.23-9300-Leaf01#
show mon cap 1 buff br | i ARP

11
 8.153510 00:06:f6:01:cd:42 -> ff:ff:ff:ff:ff:ff ARP 110
Who has 10.1.201.11? Tell 10.1.201.10  <-- .10 requests .11 MAC (this is Frame 11)

12 8.154030 00:00:be:ef:ca:fe -> 00:06:f6:01:cd:42 ARP 110 10.1.201.11
is at 00:00:be:ef:ca:fe    <-- CGW replies with its MAC
```

查看捕获数据包的详细信息。如果要查看有关数据包的详细信息，请使用EPC的detail选项

- 请注意，为了简洁起见，此输出会在不同位置进行剪切

<#root>

```
Leaf01#
show mon cap 1 buffer detailed | beg Frame 11   <-- begin detail result from Frame 11 (ARP Request)
Frame 11: 110 bytes on wire (880 bits), 110 bytes captured (880 bits) on interface /tmp/epc_ws/wif_to_t
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
    Destination: 00:00:00:00:00:00 (00:00:00:00:00:00)
        Address: 00:00:00:00:00:00 (00:00:00:00:00:00)
        .... ..0. .... .... .... .... = LG bit: Globally unique address (factory default)
        .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
    Source: 00:00:00:00:00:00 (00:00:00:00:00:00)
        Address: 00:00:00:00:00:00 (00:00:00:00:00:00)
        .... ..0. .... .... .... .... = LG bit: Globally unique address (factory default)
        .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
    Type: IPv4 (0x0800)
Internet Protocol Version 4, Src: 172.16.254.3, Dst: 172.16.254.6     <--- Outer tunnel IP header
```

```
    Source: 172.16.254.3
    Destination: 172.16.254.6
User Datagram Protocol, Src Port: 65483,
Dst Port: 4789  <-- VXLAN Dest port


Virtual eXtensible Local Area Network
    VXLAN Network Identifier
(VNI): 20101               <-- Verify the VNI for the segment you are investigating

    Reserved: 0
Ethernet II, Src: 00:06:f6:01:cd:42 (00:06:f6:01:cd:42), Dst: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)  <--



  Type: ARP (0x0806)

    Trailer: 00000000000000000000000000000000000000
Address Resolution Protocol (
request
)
 <-- is an ARP request


    Hardware type: Ethernet (1)
    Protocol type: IPv4 (0x0800)
    Hardware size: 6
    Protocol size: 4
    Opcode: request (1)


Sender MAC address: 00:06:f6:01:cd:42 (00:06:f6:01:cd:42)   <-- Sending host
    Sender IP address: 10.1.201.10
    Target MAC address: 00:00:00:00:00:00 (00:00:00:00:00:00)   <-- Trying to resolve MAC for host
    Target IP address: 10.1.201.11



Frame 12:
 110 bytes on wire (880 bits), 110 bytes captured (880 bits) on interface /tmp/epc_ws/wif_to_ts_pipe, i
<-- ARP reply


Ethernet II,
Src: dc:77:4c:8a:6d:7f
 (dc:77:4c:8a:6d:7f),
Dst: 68:2c:7b:f8:87:48
 (68:2c:7b:f8:87:48)
<-- Underlay MACs



Internet Protocol Version 4, Src: 172.16.254.6, Dst: 172.16.254.3
User Datagram Protocol, Src Port: 65410, Dst Port: 4789
Virtual eXtensible Local Area Network
```

```
    VXLAN Network Identifier (VNI): 20101
    Reserved: 0
Ethernet II,

Src: 00:00:be:ef:ca:fe

 (00:00:be:ef:ca:fe),

Dst: 00:06:f6:01:cd:42

 (00:06:f6:01:cd:42)

 <-- Start of payload



 Type: ARP

 (0x0806)
    Trailer: 00000000000000000000000000000000000000
Address Resolution Protocol (

reply

)

<-- is an ARP reply


    Hardware type: Ethernet (1)
    Protocol type: IPv4 (0x0800)
    Hardware size: 6
    Protocol size: 4
    Opcode: reply (2)


 Sender MAC address: 00:00:be:ef:ca:fe (00:00:be:ef:ca:fe)   <-- Reply is that of the CGW MAC due to loc

    Sender IP address: 10.1.201.11
    Target MAC address: 00:06:f6:01:cd:42 (00:06:f6:01:cd:42)
    Target IP address: 10.1.201.10
```

CGW RT2网关前缀

缺少网关前缀

如前面有关部分隔离网段部分所述，需要在交换矩阵VLAN中学习MAC

- 如果没有超过MAC老化计时器的流量流向网关，则可能会出现此问题。
- 如果CGW网关前缀缺失，您需要确认MAC存在

```
<#root>

CGW#

show bgp l2vpn evpn route-type 2 0 0000.beef.cafe 10.1.202.1

% Network not in table <-- RT2 not generated on CGW
```

```
CGW#

show mac address-table address 0000.beef.cafe

          Mac Address Table
------------------------------------------

Vlan    Mac Address       Type        Ports
----    -----------       --------    -----
 201    0000.beef.cafe    STATIC      Vl201
2021    0000.beef.cafe    STATIC      Vl2021

<-- MAC is not learned in Fabric Vlan 202

Total Mac Addresses for this criterion: 2
```

网关前缀缺失补救

在大多数生产网络中，可能始终有一些流量。但是，如果您遇到此问题，可以使用以下选项之一来修复此问题：

- 添加静态MAC条目，例如“mac address-table static 0000.beef.cafe vlan 202 interface TwentyFiveGigE1/0/1”
- 使用“mac address-table aging-time <seconds>”增加MAC老化计时器。（请记住，这会增加所有MAC地址的老化时间，因此首选静态MAC选项）

缺少DEF GW属性

对于部分隔离网段，有许多其他配置可添加此属性。

缺少DEF GW属性补救

确认以下详细信息：

- 您运行的是17.12.1或更高版本
- 配置中存在SISF（设备跟踪）CLI
- 配置route-map match & set命令并将路由映射应用到BGP邻居
- 您已刷新BGP通告(必须清除BGP才能使用新属性重新通告前缀

无线漫游

频繁漫游可能导致BGP更新过于频繁，并且应在交换机声明它拥有MAC并发送RT2更新之前增加每个时间间隔的漫游量

- 当主机在不同交换机上的两个AP之间移动时，就会发生这种情况。
- 漫游的默认限制为每180秒5次

```
<#root>

Leaf01#

sh run | sec l2vpn
```

```
l2vpn evpn
 replication-type static
 flooding-suppression address-resolution disable

 ip duplication limit 10 time 180          <--- You can adjust this default in the global l2vpn section
 mac duplication limit 10 time 180


Leaf01#

sh l2vpn evpn summary

L2VPN EVPN
  EVPN Instances (excluding point-to-point): 4
    VLAN Based:   4
  Vlans: 4
  BGP: ASN 65001, address-family l2vpn evpn configured
  Router ID: 172.16.254.3
  Global Replication Type: Static
  ARP/ND Flooding Suppression: Disabled
  Connectivity to Core: UP

  MAC Duplication: seconds 180 limit 10

  MAC Addresses: 13
    Local:     6
    Remote:    7

    Duplicate: 0
  IP Duplication: seconds 180 limit 10

  IP Addresses: 7
    Local:     4
    Remote:    3

    Duplicate: 0

<...snip...>
```

为TAC收集的命令

如果此指南未能解决您的问题，请收集显示的命令列表，并将它们附加到您的TAC服务请求。

要收集的最低信息

（在重新加载/恢复操作之前收集数据的时间有限）

- Show tech evpn
- Show tech
- Show tech sisf

要收集的详细信息

(如果有时间收集更完整的数据，则首选)

- show tech
- show tech evpn
- show tech platform evpn_vxlan switch <number>
- show tech platform
- show tech resource
- show tech sisf
- show tech isis
- show tech bgp
- show monitor event-trace evpn event all
- show monitor event-trace evpn error all
- request platform software trace archive

相关信息

- 在Catalyst 9000系列交换机上实施BGP EVPN路由策略
- DHCP第2层中继（即将推出）

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）。