Solución de problemas de reenvío de ACI dentro del fabric: reenvío de varios paquetes

Contenido

Introducción

Este documento describe los pasos para comprender y resolver problemas de un escenario de reenvío de varios grupos de dispositivos ACI.

Antecedentes

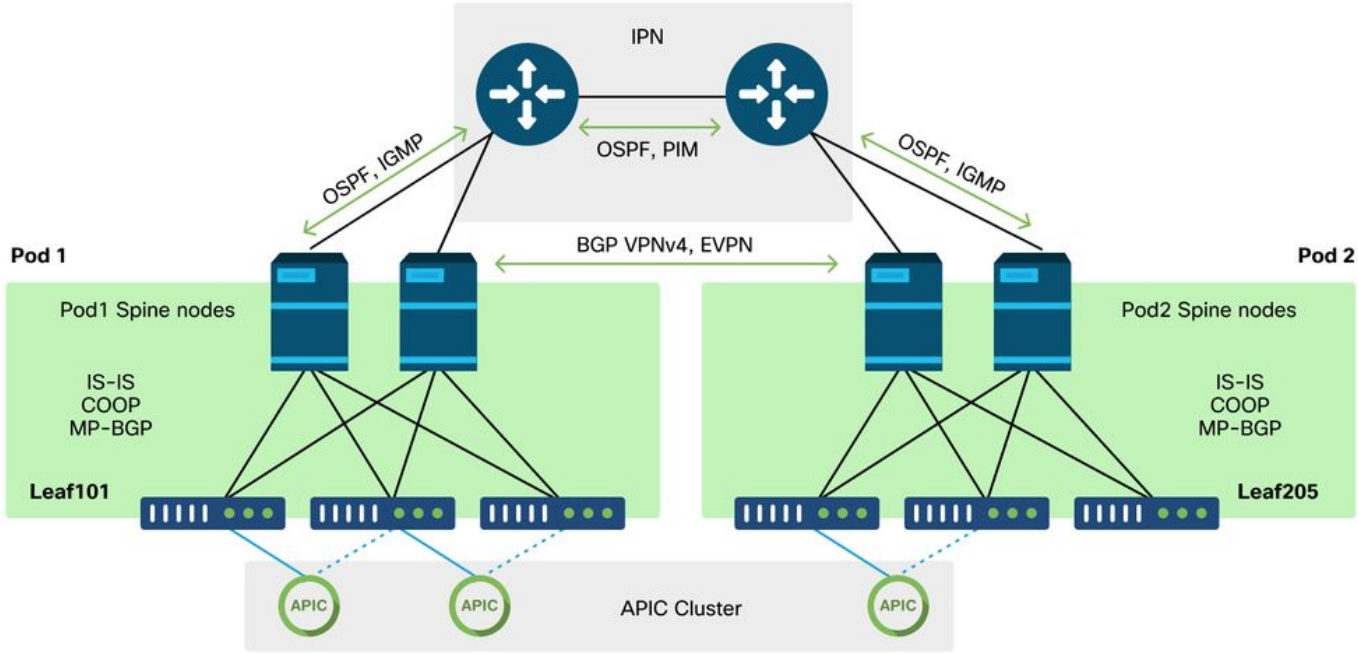
El material de este documento se ha extraído de la Solución de problemas de Cisco Application Centric Infrastructure, segunda edición libro, específicamente el **Reenvío dentro de la estructura: reenvío de varios dispositivos** capítulo.

Descripción general del reenvío de varios paquetes

En este capítulo se explica cómo solucionar problemas en situaciones en las que la conectividad no funciona correctamente entre dispositivos en un entorno de varios dispositivos

Antes de analizar ejemplos específicos de solución de problemas, es importante dedicar unos instantes a comprender los componentes de varios dispositivos a un nivel superior.

Componentes de varios dispositivos



Al igual que un fabric de ACI tradicional, un fabric de varios grupos de dispositivos sigue considerándose un único fabric de ACI y depende de un solo clúster de APIC para la gestión.

En cada POD individual, ACI aprovecha los mismos protocolos de la superposición que un fabric

tradicional. Esto incluye IS-IS para el intercambio de información de TEP, así como la selección de la interfaz de salida multidifusión (OIF), COOP para un repositorio de terminales global y BGP VPNv4 para la distribución de routers externos a través del fabric.

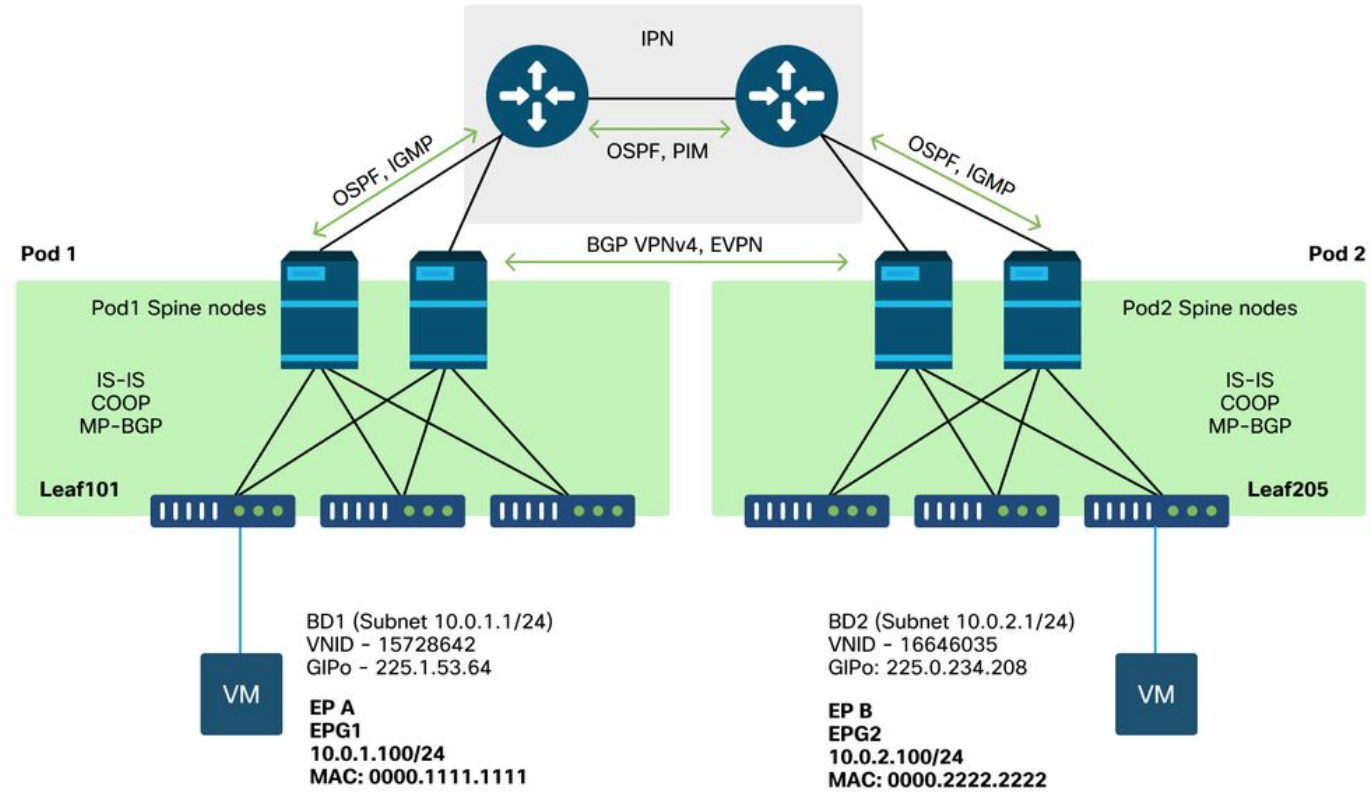
Multi-Pod se basa en estos componentes, ya que debe conectar cada Pod entre sí.

- Para intercambiar información de ruteo relacionada con los TEPs en el Pod remoto, OSPF se utiliza para anunciar el pool TEP de resumen a través del IPN.
- Para intercambiar rutas externas aprendidas de un Pod a otro, la familia de direcciones BGP VPNv4 se extiende entre los nodos de columna. Cada POD se convierte en un clúster reflector de ruta independiente.
- Para sincronizar los terminales así como otra información almacenada en COOP a través de los Pods, la familia de direcciones EVPN BGP se extiende entre los nodos de columna.
- Por último, para gestionar la saturación del tráfico de difusión, unidifusión desconocida y multidifusión (BUM) a través de los grupos de dispositivos, los nodos de columna de cada grupo de dispositivos actúan como hosts IGMP y los routers IPN intercambian información de routing multidifusión a través de PIM bidireccional.

Una gran parte de los flujos de trabajo y los escenarios de solución de problemas de varios dispositivos son similares a los fabrics de ACI de un solo grupo de dispositivos. Esta sección de varios dispositivos se centrará principalmente en las diferencias entre el envío de un solo dispositivo y de varios dispositivos.

Topología para ejemplos de varios dispositivos

Al igual que con la resolución de problemas de cualquier escenario, es importante comenzar por comprender cuál es el estado esperado. Consulte esta topología para ver los ejemplos de este capítulo.



Flujo de trabajo general para solucionar problemas de reenvío de varios paquetes

En un nivel superior, al depurar un problema de reenvío de varios grupos de dispositivos, se pueden evaluar los siguientes pasos:

1. ¿El flujo es unidifusión o multidestino? Recuerde, incluso si se espera que el flujo sea unicast en el estado de funcionamiento, si ARP no se resuelve, entonces es un flujo multidestino.
2. ¿El flujo está ruteado o puenteado? Tradicionalmente, un flujo enrutado desde una perspectiva de ACI sería cualquier flujo en el que la dirección MAC de destino es la dirección MAC del router propiedad de un gateway configurado en ACI. Además, si se inhabilita la inundación ARP, la hoja de ingreso se rutearía según la dirección IP de destino. Si la dirección MAC de destino no es propiedad de ACI, el switch se reenviará en función de la dirección MAC o seguirá el comportamiento de "unidifusión desconocida" configurado en el dominio de bridge.
3. ¿La hoja de ingreso está perdiendo el flujo? fTriage y ELAM son las mejores herramientas para confirmarlo.

Si el flujo es unidifusión de capa 3:

1. ¿La hoja de ingreso tiene un punto final que aprende para la IP de destino en el mismo VRF que el EPG de origen? Si es así, siempre tendrá prioridad sobre cualquier ruta aprendida. La hoja se reenviará directamente a la dirección del túnel o a la interfaz de salida donde se aprende el punto final.
2. Si no hay aprendizaje de terminal, ¿tiene la hoja de ingreso una ruta para el destino que

tiene el indicador 'Pervasive' configurado? Esto indica que la subred de destino está configurada como una subred de dominio de puente y que el salto siguiente debe ser el proxy de columna en el POD local.
3. Si no hay una ruta ubicua, el último recurso sería cualquier ruta que se aprenda a través de un L3Out. Esta parte es idéntica al reenvío L3Out de grupo único.

Si el flujo es unidifusión de capa 2:

1. ¿La hoja de ingreso tiene un punto final que aprende para la dirección MAC de destino en el mismo dominio de puente que el EPG de origen? Si es así, la hoja se reenviará a la IP del túnel remoto o a la interfaz local donde se aprende el punto final.
2. Si no hay aprendizaje para la dirección MAC de destino en el dominio de puente de origen, la hoja se reenviará en función del comportamiento "unicast desconocido" en el que se haya configurado BD. Si se establece en 'Flood', la hoja inundará el grupo de multidifusión GIPo asignado al dominio de puente. Los Pods locales y remotos deben obtener una copia inundada. Si se establece en 'Hardware Proxy', la trama se envía a la columna para una búsqueda de proxy y se reenvía en función de la entrada COOP de la columna.

Dado que las salidas de solución de problemas serían considerablemente diferentes para unicast en comparación con BUM, las salidas de trabajo y los escenarios para unicast se considerarán antes y luego se trasladarán a BUM.

Flujo de trabajo de resolución de problemas de unidifusión multiprocesador

Siguiendo la topología, recorra el flujo desde 10.0.2.100 en leaf205 hasta 10.0.1.100 en leaf101.

Tenga en cuenta que, antes de continuar aquí, es importante confirmar si el origen tiene ARP resuelto para la gateway (para un flujo ruteado) o la dirección MAC de destino (para un flujo puenteado)

1. Confirme que la hoja de ingreso recibe el paquete. Utilice la herramienta CLI de ELAM que se muestra en la sección "Herramientas" junto con la salida del informe disponible en 4.2. También se utiliza la aplicación ELAM Assistant.

```
module-1# debug platform internal tah elam asic 0
module-1(DBG-elam)# trigger reset
module-1(DBG-elam)# trigger init in-select 6 out-select 1
module-1(DBG-elam-insel6)# set outer ipv4 src_ip 10.0.2.100 dst_ip 10.0.1.100
module-1(DBG-elam-insel6)# start
module-1(DBG-elam-insel6)# status
ELAM STATUS
===========
Asic 0 Slice 0 Status Armed
Asic 0 Slice 1 Status Triggered
```

Observe que se activó la ELAM que confirma que el paquete se recibió en el switch de ingreso. Ahora observe un par de campos en el informe, ya que el resultado es amplio.

```
================================================================================
===========
                                                    Captured Packet
```

```
==========================================================================================
==========

------------------------------------------------------------------------------------------
----------
Outer Packet Attributes
------------------------------------------------------------------------------------------
----------
Outer Packet Attributes       : l2uc ipv4 ip ipuc ipv4uc
Opcode                        : OPCODE_UC

------------------------------------------------------------------------------------------
----------
Outer L2 Header
------------------------------------------------------------------------------------------
----------
Destination MAC               : 0022.BDF8.19FF
Source MAC                    : 0000.2222.2222
802.1Q tag is valid           : yes( 0x1 )
CoS                           : 0( 0x0 )
Access Encap VLAN             : 1021( 0x3FD )

------------------------------------------------------------------------------------------
----------
Outer L3 Header
------------------------------------------------------------------------------------------
----------
L3 Type                       : IPv4
IP Version                    : 4
DSCP                          : 0
IP Packet Length              : 84 ( = IP header(28 bytes) + IP payload )
Don't Fragment Bit            : not set
TTL                           : 255
IP Protocol Number            : ICMP
IP CheckSum                   : 10988( 0x2AEC )
Destination IP                : 10.0.1.100
Source IP                     : 10.0.2.100
```

Hay mucha más información en el informe sobre adónde va el paquete, pero la aplicación ELAM Assistant es actualmente más útil para interpretar estos datos. El resultado del asistente ELAM para este flujo se mostrará más adelante en este capítulo.

2. ¿La hoja de ingreso está aprendiendo el destino como un punto final en el VRF de ingreso? Si no es así, ¿existe alguna ruta?

```
a-leaf205# show endpoint ip 10.0.1.100 detail
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+------------------------------------+--------------+----------------+-------------+----------
---+----------------------------+
    VLAN/                           Encap           MAC Address       MAC Info/
Interface     Endpoint Group
    Domain                          VLAN            IP Address        IP Info
                  Info
+------------------------------------+--------------+----------------+-------------+----------
---+----------------------------+
```

Si no hay salida en el comando anterior significa que no se aprende la IP de destino. Luego

verifique la tabla de ruteo.

```
a-leaf205# show ip route 10.0.1.100 vrf Prod:Vrf1
IP Route Table for VRF "Prod:Vrf1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.0.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.120.34%overlay-1, [1/0], 01:55:37, static, tag 4294967294
       recursive next hop: 10.0.120.34/32%overlay-1
```

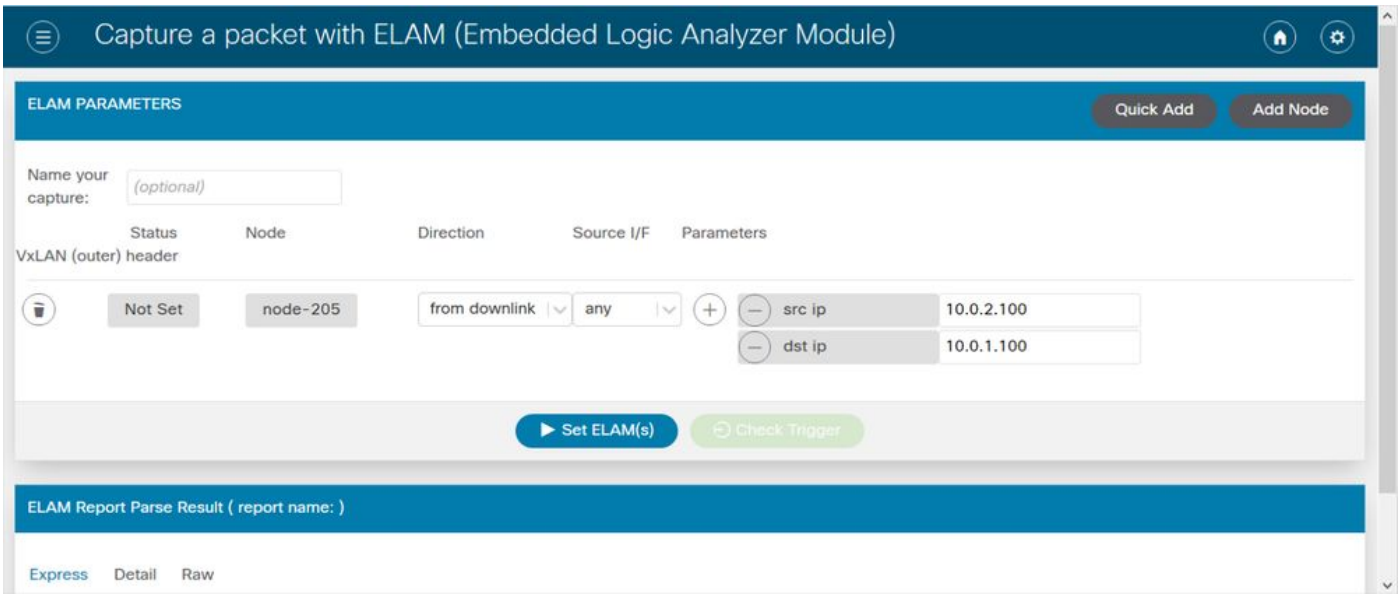
En el resultado anterior, se ve el indicador ubicuo que indica que se trata de una ruta de subred de dominio de puente. El salto siguiente debe ser una dirección proxy de difusión por proximidad en las columnas.

```
a-leaf205# show isis dtep vrf overlay-1 | grep 10.0.120.34
10.0.120.34        SPINE    N/A              PHYSICAL,PROXY-ACAST-V4
```

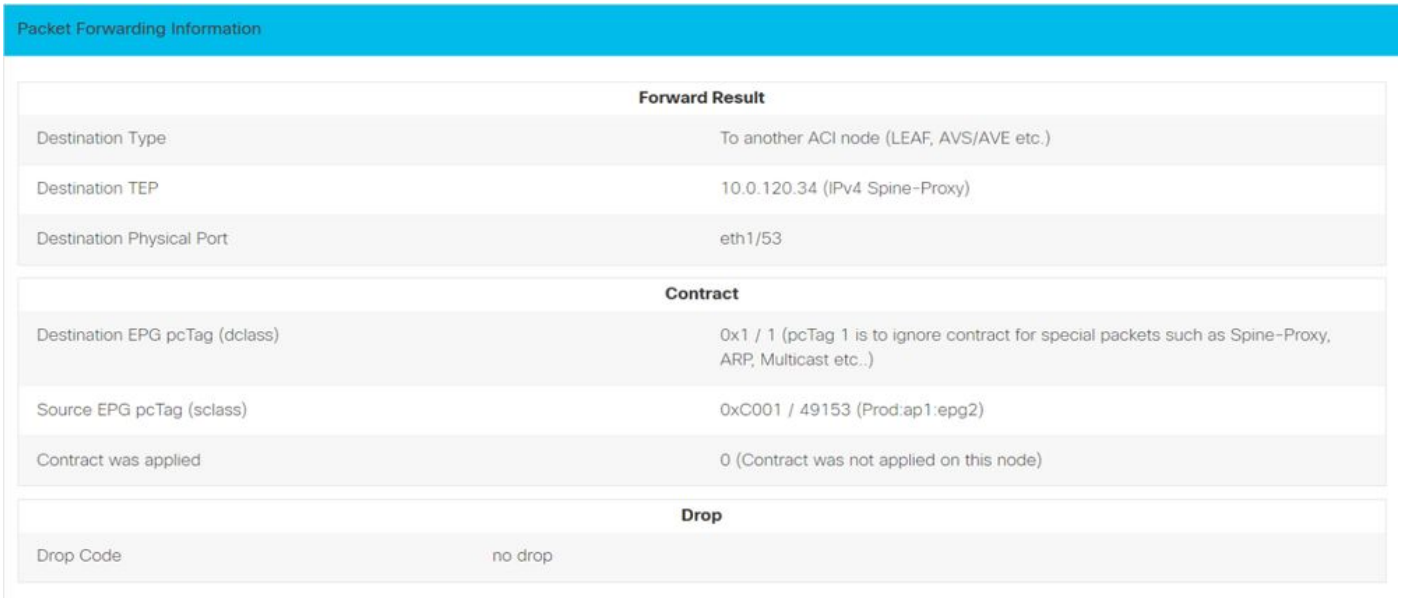
Tenga en cuenta que si el terminal se detecta en un túnel o interfaz física, esto tendrá prioridad, lo que hará que el paquete se reenvíe directamente allí. Consulte el capítulo "Reenvío externo" de este manual para obtener más información.

Utilice el asistente ELAM para confirmar las decisiones de reenvío vistas en los resultados anteriores.

Configuración de ELAM Assistant



Verificar decisiones de reenvío



El resultado anterior muestra que la hoja de ingreso está reenviando el paquete a la dirección proxy de la columna IPv4. Esto es lo que se espera que ocurra.

3. Confirme en la columna que la IP de destino está presente en COOP para que funcione la solicitud de proxy.

Hay varias maneras de obtener la salida COOP en la columna, por ejemplo, mírela con un comando 'show coop internal info ip-db':

```
a-spine4# show coop internal info ip-db | grep -B 2 -A 15 "10.0.1.100"

------------------------------
IP address : 10.0.1.100
Vrf : 2392068 <-- This vnid should correspond to vrf where the IP is learned. Check operational
tab of the tenant vrfs
Flags : 0x2
EP bd vnid : 15728642
EP mac :  00:00:11:11:11:11
Publisher Id : 192.168.1.254
Record timestamp : 12 31 1969 19:00:00 0
Publish timestamp : 12 31 1969 19:00:00 0
Seq No: 0
Remote publish timestamp: 09 30 2019 20:29:07 9900483
URIB Tunnel Info
Num tunnels : 1
      Tunnel address : 10.0.0.34 <-- When learned from a remote pod this will be an External
Proxy TEP. We'll cover this more
      Tunnel ref count : 1

------------------------------
```

Otros comandos que se ejecutarán en la columna:

Consultar COOP para entrada l2:

```
moquery -c coopEpRec -f 'coop.EpRec.mac=="00:00:11:11:22:22"
```

Consultar COOP para entrada l3 y obtener entrada l2 principal:

```
moquery -c coopEpRec -x rsp-subtree=children 'rsp-subtree-
filter=eq(coopIpv4Rec.addr,"192.168.1.1")' rsp-subtree-include=required
```

Consultar COOP sólo para entrada l3:

```
moquery -c coopIpv4Rec -f 'coop.Ipv4Rec.addr=="192.168.1.1"'
```

Lo útil de las múltiples moquery es que también se pueden ejecutar directamente en un APIC y el usuario puede ver cada columna que tiene el registro en coop.

4. Decisión de reenvío de proxy de columna de varios dispositivos

Si la entrada COOP de la columna apunta a un túnel en el POD local, el reenvío se basa en el comportamiento ACI tradicional.

Tenga en cuenta que el propietario de un TEP se puede verificar en el fabric ejecutando desde un APIC: **moquery -c ipv4Addr -f 'ipv4.Addr.addr=="<tunnel address>"'**

En el escenario proxy, el salto siguiente del túnel es 10.0.0.34. ¿Quién es el propietario de esta dirección IP?:

```
a-apic1# moquery -c ipv4Addr -f 'ipv4.Addr.addr=="10.0.0.34"' | grep dn
dn               : topology/pod-1/node-1002/sys/ipv4/inst/dom-overlay-1/if-[lo9]/addr-
[10.0.0.34/32]
dn               : topology/pod-1/node-1001/sys/ipv4/inst/dom-overlay-1/if-[lo2]/addr-
[10.0.0.34/32]
```

Esta IP pertenece a ambos nodos de columna del grupo de dispositivos 1. Se trata de una IP específica denominada dirección de proxy externo. Del mismo modo que ACI tiene direcciones proxy propiedad de los nodos de columna dentro de un POD (consulte el paso 2 de esta sección), también hay direcciones proxy asignadas al POD en sí. Este tipo de interfaz se puede verificar ejecutando:

```
a-apic1# moquery -c ipv4If -x rsp-subtree=children 'rsp-subtree-
filter=eq(ipv4Addr.addr,"10.0.0.34")' rsp-subtree-include=required

...
# ipv4.If
mode         : anycast-v4,external

# ipv4.Addr
 addr             : 10.0.0.34/32
 dn               : topology/pod-1/node-1002/sys/ipv4/inst/dom-overlay-1/if-[lo9]/addr-
[10.0.0.34/32]
```

El indicador 'external' indica que se trata de un proxy externo TEP.

5. Verifique BGP EVPN en la columna

El registro de punto final de la coop debe importarse desde BGP EVPN en la columna. El siguiente comando se puede utilizar para verificar que está en EVPN (aunque si ya está en COOP con un salto siguiente del proxy externo TEP del Pod remoto se puede asumir que proviene de EVPN):

```
a-spine4# show bgp l2vpn evpn 10.0.1.100 vrf overlay-1
Route Distinguisher: 1:16777199
BGP routing table entry for [2]:[0]:[15728642]:[48]:[0000.1111.1111]:[32]:[10.0.1.100]/272,
version 689242 dest ptr 0xaf42a4ca
Paths: (2 available, best #2)
Flags: (0x000202 00000000) on xmit-list, is not in rib/evpn, is not in HW, is locked
Multipath: eBGP iBGP

 Path type: internal 0x40000018 0x2040 ref 0 adv path ref 0, path is valid, not best reason:
Router Id, remote nh not installed
 AS-Path: NONE, path sourced internal to AS
  192.168.1.254 (metric 7) from 192.168.1.102 (192.168.1.102)
    Origin IGP, MED not set, localpref 100, weight 0
    Received label 15728642 2392068
    Received path-id 1
    Extcommunity:
        RT:5:16
        SOO:1:1
        ENCAP:8
        Router MAC:0200.0000.0000

          Advertised path-id 1
 Path type: internal 0x40000018 0x2040 ref 1 adv path ref 1, path is valid, is best path, remote
nh not installed
 AS-Path: NONE, path sourced internal to AS
  192.168.1.254 (metric 7) from 192.168.1.101 (192.168.1.101)
    Origin IGP, MED not set, localpref 100, weight 0
    Received label 15728642 2392068
    Received path-id 1
    Extcommunity:
        RT:5:16
        SOO:1:1
        ENCAP:8
        Router MAC:0200.0000.0000

          Path-id 1 not advertised to any peer
```

Tenga en cuenta que el comando anterior también se puede ejecutar para una dirección MAC.

-192.168.1.254 es el TEP del plano de datos configurado durante la configuración de varios paquetes. Sin embargo, tenga en cuenta que aunque se anuncie en BGP como NH, el salto siguiente real será el proxy externo TEP.

-192.168.1.101 y .102 son los nodos de columna Pod 1 que anuncian esta ruta.

6. Verifique COOP en las espinas del POD de destino.

Se puede utilizar el mismo comando que el anterior:

```
a-spine2# show coop internal info ip-db | grep -B 2 -A 15 "10.0.1.100"

------------------------------
IP address : 10.0.1.100
Vrf : 2392068
Flags : 0
EP bd vnid : 15728642
EP mac :  00:50:56:81:3E:E6
Publisher Id : 10.0.72.67
Record timestamp : 10 01 2019 15:46:24 502206158
Publish timestamp : 10 01 2019 15:46:24 524378376
```

```
Seq No: 0
Remote publish timestamp: 12 31 1969 19:00:00 0
URIB Tunnel Info
Num tunnels : 1
      Tunnel address : 10.0.72.67
      Tunnel ref count : 1

------------------------------
```

Verifique quién posee la dirección de túnel ejecutando el siguiente comando en un APIC:

```
a-apic1# moquery -c ipv4Addr -f 'ipv4.Addr.addr=="10.0.72.67"'
Total Objects shown: 1

# ipv4.Addr
addr             : 10.0.72.67/32
childAction      :
ctrl             :
dn               : topology/pod-1/node-101/sys/ipv4/inst/dom-overlay-1/if-[lo0]/addr-
[10.0.72.67/32]
ipv4CfgFailedBmp :
ipv4CfgFailedTs  : 00:00:00:00.000
ipv4CfgState     : 0
lcOwn            : local
modTs            : 2019-09-30T18:42:43.262-04:00
monPolDn         : uni/fabric/monfab-default
operSt           : up
operStQual       : up
pref             : 0
rn               : addr-[10.0.72.67/32]
status           :
tag              : 0
type             : primary
vpcPeer          : 0.0.0.0
```

El comando anterior muestra que el túnel de COOP apunta a leaf101. Esto significa que leaf101 debe tener el aprendizaje local para el punto final de destino.

7. Verifique que la hoja de egreso tenga el aprendizaje local.

Esto se puede hacer mediante un comando 'show endpoint':

```
a-leaf101# show endpoint ip 10.0.1.100 detail
Legend:
s - arp              H - vtep             V - vpc-attached     p - peer-aged
R - peer-attached-rl B - bounce           S - static           M - span
D - bounce-to-proxy  O - peer-attached    a - local-aged       m - svc-mgr
L - local            E - shared-service
+----------------------------------+---------------+-----------------+--------------+----------
---+----------------------------+
    VLAN/                           Encap           MAC Address       MAC Info/
Interface     Endpoint Group
    Domain                          VLAN            IP Address        IP
Info                       Info
+----------------------------------+---------------+-----------------+--------------+----------
---+----------------------------+
341                                     vlan-1075    0000.1111.1111 LV
po5                 Prod:ap1:epg1
Prod:Vrf1                               vlan-1075        10.0.1.100 LV
po5
```

Tenga en cuenta que se aprende el punto final. El paquete se debe reenviar en función del canal de puerto 5 con la etiqueta VLAN 1075 establecida.

Uso de fTriage para verificar el flujo de extremo a extremo

Tal como se describe en la sección "Herramientas" de este capítulo, fTriage se puede utilizar para trazar un flujo existente de extremo a extremo y comprender lo que cada switch de la ruta está haciendo con el paquete. Esto resulta especialmente útil en implementaciones más grandes y complejas, como las de varios dispositivos.

Tenga en cuenta que fTriage tardará algún tiempo en ejecutarse por completo (posiblemente 15 minutos).

Al ejecutar fTriage en el flujo de ejemplo:

```
a-apic1# ftriage route -ii LEAF:205 -dip 10.0.1.100 -sip 10.0.2.100
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "InProgress", "pid": "7297",
"apicId": "1", "id": "0"}}}
Starting ftriage
Log file name for the current run is: ftlog_2019-10-01-16-04-15-438.txt
2019-10-01 16:04:15,442 INFO     /controller/bin/ftriage route -ii LEAF:205 -dip 10.0.1.100 -sip
10.0.2.100
2019-10-01 16:04:38,883 INFO     ftriage:     main:1165 Invoking ftriage with default password
and default username: apic#fallback\\admin
2019-10-01 16:04:54,678 INFO     ftriage:     main:839  L3 packet Seen on a-leaf205 Ingress:
Eth1/31  Egress: Eth1/53 Vnid: 2392068
2019-10-01 16:04:54,896 INFO     ftriage:     main:242  ingress encap string vlan-1021
2019-10-01 16:04:54,899 INFO     ftriage:     main:271  Building ingress BD(s), Ctx
2019-10-01 16:04:56,778 INFO     ftriage:     main:294  Ingress BD(s) Prod:Bd2
2019-10-01 16:04:56,778 INFO     ftriage:     main:301  Ingress Ctx: Prod:Vrf1
2019-10-01 16:04:56,887 INFO     ftriage:   pktrec:490  a-leaf205: Collecting transient losses
snapshot for LC module: 1
2019-10-01 16:05:22,458 INFO     ftriage:     main:933  SIP 10.0.2.100 DIP 10.0.1.100
2019-10-01 16:05:22,459 INFO     ftriage:  unicast:973  a-leaf205: <- is ingress node
2019-10-01 16:05:25,206 INFO     ftriage:  unicast:1215 a-leaf205: Dst EP is remote
2019-10-01 16:05:26,758 INFO     ftriage:     misc:657  a-leaf205: DMAC(00:22:BD:F8:19:FF) same
as RMAC(00:22:BD:F8:19:FF)
2019-10-01 16:05:26,758 INFO     ftriage:     misc:659  a-leaf205: L3 packet getting
routed/bounced in SUG
2019-10-01 16:05:27,030 INFO     ftriage:     misc:657  a-leaf205: Dst IP is present in SUG L3
tbl
2019-10-01 16:05:27,473 INFO     ftriage:     misc:657  a-leaf205: RwDMAC DIPo(10.0.72.67) is
one of dst TEPs ['10.0.72.67']
2019-10-01 16:06:25,200 INFO     ftriage:     main:622  Found peer-node a-spine3 and IF: Eth1/31
in candidate list
2019-10-01 16:06:30,802 INFO     ftriage:     node:643  a-spine3: Extracted Internal-port GPD
Info for lc: 1
2019-10-01 16:06:30,803 INFO     ftriage:     fcls:4414 a-spine3: LC trigger ELAM with IFS:
Eth1/31 Asic :3 Slice: 1 Srcid: 24
2019-10-01 16:07:05,717 INFO     ftriage:     main:839  L3 packet Seen on a-spine3 Ingress:
Eth1/31 Egress: LC-1/3 FC-24/0 Port-1 Vnid: 2392068
2019-10-01 16:07:05,718 INFO     ftriage:   pktrec:490  a-spine3: Collecting transient losses
snapshot for LC module: 1
2019-10-01 16:07:28,043 INFO     ftriage:      fib:332  a-spine3: Transit in spine
2019-10-01 16:07:35,902 INFO     ftriage:  unicast:1252 a-spine3: Enter dbg_sub_nexthop with
Transit inst: ig infra: False glbs.dipo: 10.0.72.67
2019-10-01 16:07:36,018 INFO     ftriage:  unicast:1417 a-spine3: EP is known in COOP (DIPo =
10.0.72.67)
2019-10-01 16:07:40,422 INFO     ftriage:  unicast:1458 a-spine3: Infra route 10.0.72.67 present
```

```
in RIB
2019-10-01 16:07:40,423 INFO     ftriage:     node:1331 a-spine3: Mapped LC interface: LC-1/3
FC-24/0 Port-1 to FC interface: FC-24/0 LC-1/3 Port-1
2019-10-01 16:07:46,059 INFO     ftriage:     node:460  a-spine3: Extracted GPD Info for fc: 24
2019-10-01 16:07:46,060 INFO     ftriage:     fcls:5748 a-spine3: FC trigger ELAM with IFS: FC-
24/0 LC-1/3 Port-1 Asic :0 Slice: 1 Srcid: 40
2019-10-01 16:08:06,735 INFO     ftriage:  unicast:1774 L3 packet Seen on FC of node: a-spine3
with Ingress: FC-24/0 LC-1/3 Port-1  Egress: FC-24/0 LC-1/3 Port-1 Vnid: 2392068
2019-10-01 16:08:06,735 INFO     ftriage:   pktrec:487  a-spine3: Collecting transient losses
snapshot for FC module: 24
2019-10-01 16:08:09,123 INFO     ftriage:     node:1339 a-spine3: Mapped FC interface: FC-24/0
LC-1/3 Port-1 to LC interface: LC-1/3 FC-24/0 Port-1
2019-10-01 16:08:09,124 INFO     ftriage:  unicast:1474 a-spine3: Capturing Spine  Transit pkt-
type L3 packet on egress LC on Node: a-spine3 IFS: LC-1/3 FC-24/0 Port-1
2019-10-01 16:08:09,594 INFO     ftriage:     fcls:4414 a-spine3: LC trigger ELAM with IFS: LC-
1/3 FC-24/0 Port-1 Asic :3 Slice: 1 Srcid: 48
2019-10-01 16:08:44,447 INFO     ftriage:  unicast:1510 a-spine3: L3 packet Spine egress
Transit pkt Seen on a-spine3 Ingress: LC-1/3 FC-24/0 Port-1 Egress: Eth1/29 Vnid: 2392068
2019-10-01 16:08:44,448 INFO     ftriage:   pktrec:490  a-spine3: Collecting transient losses
snapshot for LC module: 1
2019-10-01 16:08:46,691 INFO     ftriage:  unicast:1681 a-spine3: Packet is exiting the fabric
through {a-spine3: ['Eth1/29']} Dipo 10.0.72.67 and filter SIP 10.0.2.100 DIP 10.0.1.100
2019-10-01 16:10:19,947 INFO     ftriage:     main:716  Capturing L3 packet Fex: False on node:
a-spine1 IF: Eth2/25
2019-10-01 16:10:25,752 INFO     ftriage:     node:643  a-spine1: Extracted Internal-port GPD
Info for lc: 2
2019-10-01 16:10:25,754 INFO     ftriage:     fcls:4414 a-spine1: LC trigger ELAM with IFS:
Eth2/25 Asic :3 Slice: 0 Srcid: 24
2019-10-01 16:10:51,164 INFO     ftriage:     main:716  Capturing L3 packet Fex: False on node:
a-spine2 IF: Eth1/31
2019-10-01 16:11:09,690 INFO     ftriage:     main:839  L3 packet Seen on a-spine2 Ingress:
Eth1/31 Egress: Eth1/25 Vnid: 2392068
2019-10-01 16:11:09,690 INFO     ftriage:   pktrec:490  a-spine2: Collecting transient losses
snapshot for LC module: 1
2019-10-01 16:11:24,882 INFO     ftriage:      fib:332  a-spine2: Transit in spine
2019-10-01 16:11:32,598 INFO     ftriage:  unicast:1252 a-spine2: Enter dbg_sub_nexthop with
Transit inst: ig infra: False glbs.dipo: 10.0.72.67
2019-10-01 16:11:32,714 INFO     ftriage:  unicast:1417 a-spine2: EP is known in COOP (DIPo =
10.0.72.67)
2019-10-01 16:11:36,901 INFO     ftriage:  unicast:1458 a-spine2: Infra route 10.0.72.67 present
in RIB
2019-10-01 16:11:47,106 INFO     ftriage:     main:622  Found peer-node a-leaf101 and IF:
Eth1/54 in candidate list
2019-10-01 16:12:09,836 INFO     ftriage:     main:839  L3 packet Seen on a-leaf101 Ingress:
Eth1/54 Egress: Eth1/30 (Po5) Vnid: 11470
2019-10-01 16:12:09,952 INFO     ftriage:   pktrec:490  a-leaf101: Collecting transient losses
snapshot for LC module: 1
2019-10-01 16:12:30,991 INFO     ftriage:     nxos:1404 a-leaf101: nxos matching rule id:4659
scope:84 filter:65534
2019-10-01 16:12:32,327 INFO     ftriage:     main:522  Computed egress encap string vlan-1075
2019-10-01 16:12:32,333 INFO     ftriage:     main:313  Building egress BD(s), Ctx
2019-10-01 16:12:34,559 INFO     ftriage:     main:331  Egress Ctx Prod:Vrf1
2019-10-01 16:12:34,560 INFO     ftriage:     main:332  Egress BD(s): Prod:Bd1
2019-10-01 16:12:37,704 INFO     ftriage:  unicast:1252 a-leaf101: Enter dbg_sub_nexthop with
Local inst: eg infra: False glbs.dipo: 10.0.72.67
2019-10-01 16:12:37,705 INFO     ftriage:  unicast:1257 a-leaf101: dbg_sub_nexthop invokes
dbg_sub_eg for ptep
2019-10-01 16:12:37,705 INFO     ftriage:  unicast:1784 a-leaf101: <- is egress node
2019-10-01 16:12:37,911 INFO     ftriage:  unicast:1833 a-leaf101: Dst EP is local
2019-10-01 16:12:37,912 INFO     ftriage:     misc:657  a-leaf101: EP if(Po5) same as egr
if(Po5)
2019-10-01 16:12:38,172 INFO     ftriage:     misc:657  a-leaf101: Dst IP is present in SUG L3
tbl
2019-10-01 16:12:38,564 INFO     ftriage:     misc:657  a-leaf101: RW seg_id:11470 in SUG same
```

```
as EP segid:11470
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "Idle", "pid": "0", "apicId": "0",
"id": "0"}}}
fTriage Status: {"dbgFtriage": {"attributes": {"operState": "Idle", "pid": "0", "apicId": "0",
"id": "0"}}}
```

Hay una gran cantidad de datos en el fTriage. Se resaltan algunos de los campos más importantes. Observe que la trayectoria del paquete era 'leaf205 (Pod 2) > spine3 (Pod 2) > spine2 (Pod 1) > leaf101 (Pod 1)'. También se pueden ver todas las decisiones de reenvío y las búsquedas de contratos realizadas durante el proceso.

Tenga en cuenta que si se tratase de un flujo de capa 2, la sintaxis del fTriage tendría que definirse en un valor como:

```
ftriage bridge -ii LEAF:205 -dmac 00:00:11:11:22:22
```

Solicitudes proxy cuando el PE no está en COOP

Antes de considerar escenarios de fallos específicos, hay una parte más que tratar relacionada con el reenvío de unidifusión a través de varios dispositivos. ¿Qué sucede si el punto final de destino es desconocido, la solicitud es proxy y el punto final no está en COOP?

En este escenario, el paquete/trama se envía a la columna y se genera una solicitud de búsqueda.

Cuando la columna genera una solicitud de obtención, el paquete original se conserva en la solicitud; sin embargo, el paquete recibe ethertype 0xfff2, que es un Ethertype personalizado reservado para las operaciones de obtención. Por esta razón, no será fácil interpretar estos mensajes en herramientas de captura de paquetes como Wireshark.

El destino de la capa 3 externa también está configurado en 239.255.255.240, que es un grupo multicast reservado específicamente para los mensajes de búsqueda. Estos deben inundarse a través del fabric y cualquier switch de hoja de salida que tenga implementada la subred de destino de la solicitud de obtención generará una solicitud ARP para resolver el destino. Estos ARP se envían desde la dirección IP de subred BD configurada (por lo tanto, las solicitudes de proxy no pueden resolver la ubicación de los puntos finales silenciosos/desconocidos si el ruteo unidifusión está inhabilitado en un dominio de puente).

La recepción del mensaje de búsqueda en la hoja de salida y la respuesta ARP y ARP recibida generada posteriormente se pueden verificar a través del siguiente comando:

Verificar ARP de Glean

```
a-leaf205# show ip arp internal event-history event | grep -F -B 1 192.168.21.11
...
73) Event:E_DEBUG_DSF, length:127, at 316928 usecs after Wed May  1 08:31:53 2019
Updating epm ifidx: 1a01e000 vlan: 105 ip: 192.168.21.11, ifMode: 128 mac: 8c60.4f02.88fc   <<<
Endpoint is learned
75) Event:E_DEBUG_DSF, length:152, at 316420 usecs after Wed May  1 08:31:53 2019
log_collect_arp_pkt; sip = 192.168.21.11; dip = 192.168.21.254; interface = Vlan104;info = Garp
Check adj:(nil)  <<< Response received
77) Event:E_DEBUG_DSF, length:142, at 131918 usecs after Wed May  1 08:28:36 2019
log_collect_arp_pkt; dip = 192.168.21.11; interface = Vlan104;iod = 138; Info = Internal Request
Done   <<< ARP request is generated by leaf
78) Event:E_DEBUG_DSF, length:136, at 131757 usecs after Wed May  1 08:28:36 2019   <<< Glean
```

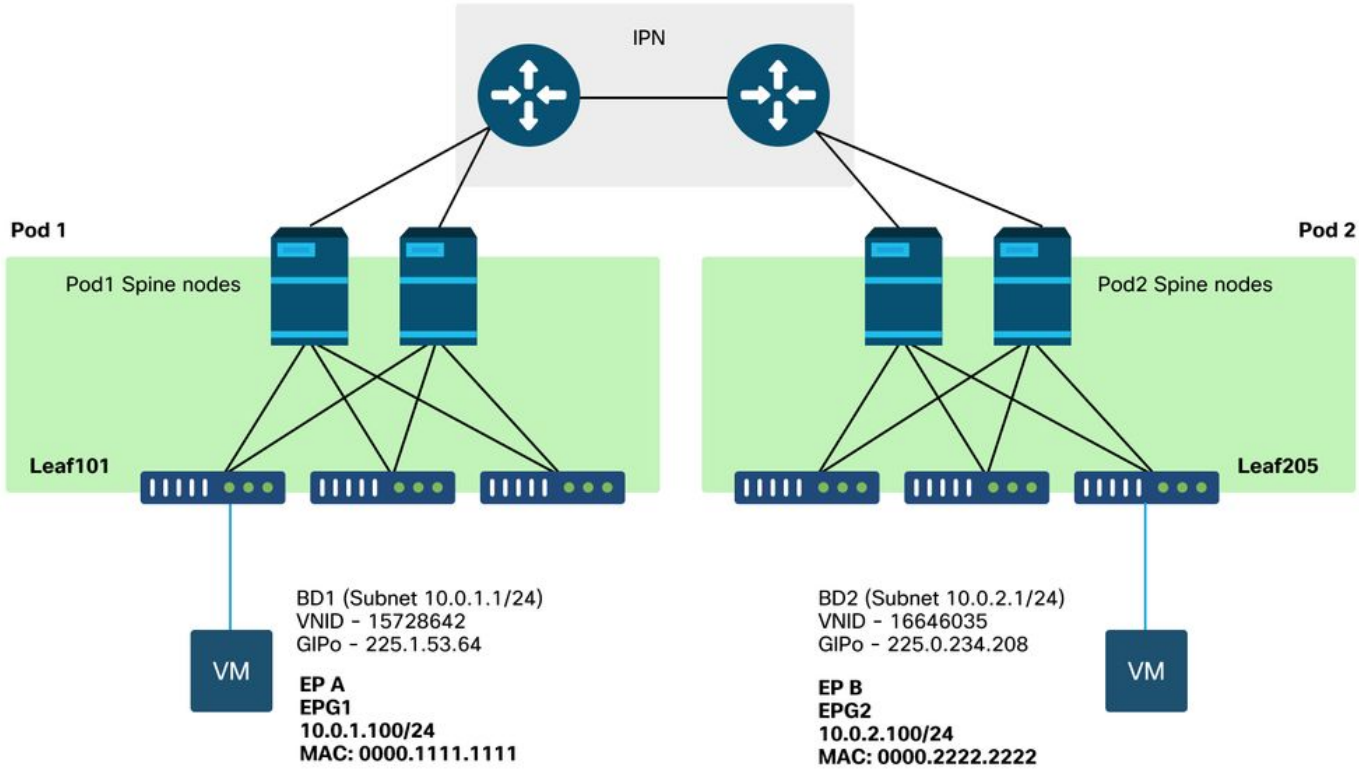
```
received, Dst IP is in BD subnet
log_collect_arp_glean;dip = 192.168.21.11;interface = Vlan104;info = Received pkt Fabric-Glean:
1
79) Event:E_DEBUG_DSF, length:174, at 131748 usecs after Wed May  1 08:28:36 2019
log_collect_arp_glean; dip = 192.168.21.11; interface = Vlan104; vrf = CiscoLive2019:vrf1; info
= Address in PSVI subnet or special VIP   <<< Glean Received, Dst IP is in BD subnet
```

Como referencia, los mensajes de búsqueda que se envían a 239.255.255.240 es la razón por la que este grupo debe incluirse en el rango del grupo PIM bidireccional en el IPN.

Situación de resolución de problemas de varios dispositivos #1 (unidifusión)

En la siguiente topología, el EP B no puede comunicarse con el EP A.

Solución de problemas de topología



Tenga en cuenta que muchos de los problemas observados para el reenvío de varios dispositivos son idénticos a los problemas observados en un único dispositivo. Por este motivo, los problemas específicos de los dispositivos multipunto se centran en.

Mientras sigue el flujo de trabajo de resolución de problemas de unidifusión descrito anteriormente, tenga en cuenta que la solicitud se procesa con proxy, pero los nodos de columna del grupo 2 no tienen la IP de destino en COOP.

Causa: Falta el terminal en COOP

Como se mencionó anteriormente, las entradas COOP para los terminales Pod remotos se llenan a partir de la información EVPN BGP. Como resultado, es importante determinar:

r.) ¿La columna vertebral de Pod (Pod 2) de origen la tiene en EVPN?

```
a-spine4# show bgp l2vpn evpn 10.0.1.100 vrf overlay-1
<no output>
```

b) ¿La columna de la Pod remota (Pod 1) la tiene en EVPN?

```
a-spine1# show bgp l2vpn evpn 10.0.1.100 vrf overlay-1
Route Distinguisher: 1:16777199    (L2VNI 1)
BGP routing table entry for [2]:[0]:[15728642]:[48]:[0050.5681.3ee6]:[32]:[10.0.1.100]/272,
version 11751 dest ptr 0xafbf8192
Paths: (1 available, best #1)
Flags: (0x00010a 00000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

 Advertised path-id 1
 Path type: local 0x4000008c 0x0 ref 0 adv path ref 1, path is valid, is best path
 AS-Path: NONE, path locally originated
  0.0.0.0 (metric 0) from 0.0.0.0 (192.168.1.101)
    Origin IGP, MED not set, localpref 100, weight 32768
    Received label 15728642 2392068
    Extcommunity:
        RT:5:16

          Path-id 1 advertised to peers:
```

La columna Pod 1 la tiene y la IP de salto siguiente es 0.0.0.0; esto significa que se exportó desde COOP localmente. Tenga en cuenta, sin embargo, que la sección "Anunciado a los pares" no incluye los nodos de columna Pod 2.

c.) ¿EVPN BGP está activo entre los Pods?

```
a-spine4# show bgp l2vpn evpn summ vrf overlay-1

Neighbor            V      AS   MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
192.168.1.101   4 65000       57380    66362            0     0      0  00:00:21  Active
192.168.1.102   4 65000       57568    66357            0     0      0  00:00:22  Active
```

Observe en el resultado anterior que los peers BGP EVPN están inactivos entre los Pods. Todo lo que no sea un valor numérico en la columna State/PfxRcd indica que la adyacencia no está activa. Los EP de Pod 1 no se aprenden a través de EVPN y no se importan a COOP.

Si se observa este problema, verifique lo siguiente:

1. ¿OSPF está activo entre los nodos de columna y los IP conectados?
2. ¿Los nodos de columna tienen rutas aprendidas a través de OSPF para las IP de columna remotas?
3. ¿La ruta completa a través del IPN admite MTU jumbo?
4. ¿Son estables todas las adyacencias de protocolo?

Otras posibles causas

Si el terminal no está en la base de datos COOP de ningún POD y el dispositivo de destino es un host silencioso (no detectado en ningún switch de hoja del fabric), compruebe que el proceso de limpieza del fabric funciona correctamente. Para que esto funcione:

- El Unicast Routing debe estar habilitado en el BD.
- El destino debe estar en una subred BD.
- El IPN debe proporcionar el servicio de ruteo multicast para el grupo 239.255.255.240.

La parte de multidifusión se trata con más detalle en la siguiente sección.

Descripción general del reenvío de multidifusión, unidifusión desconocida y multidifusión (BUM)

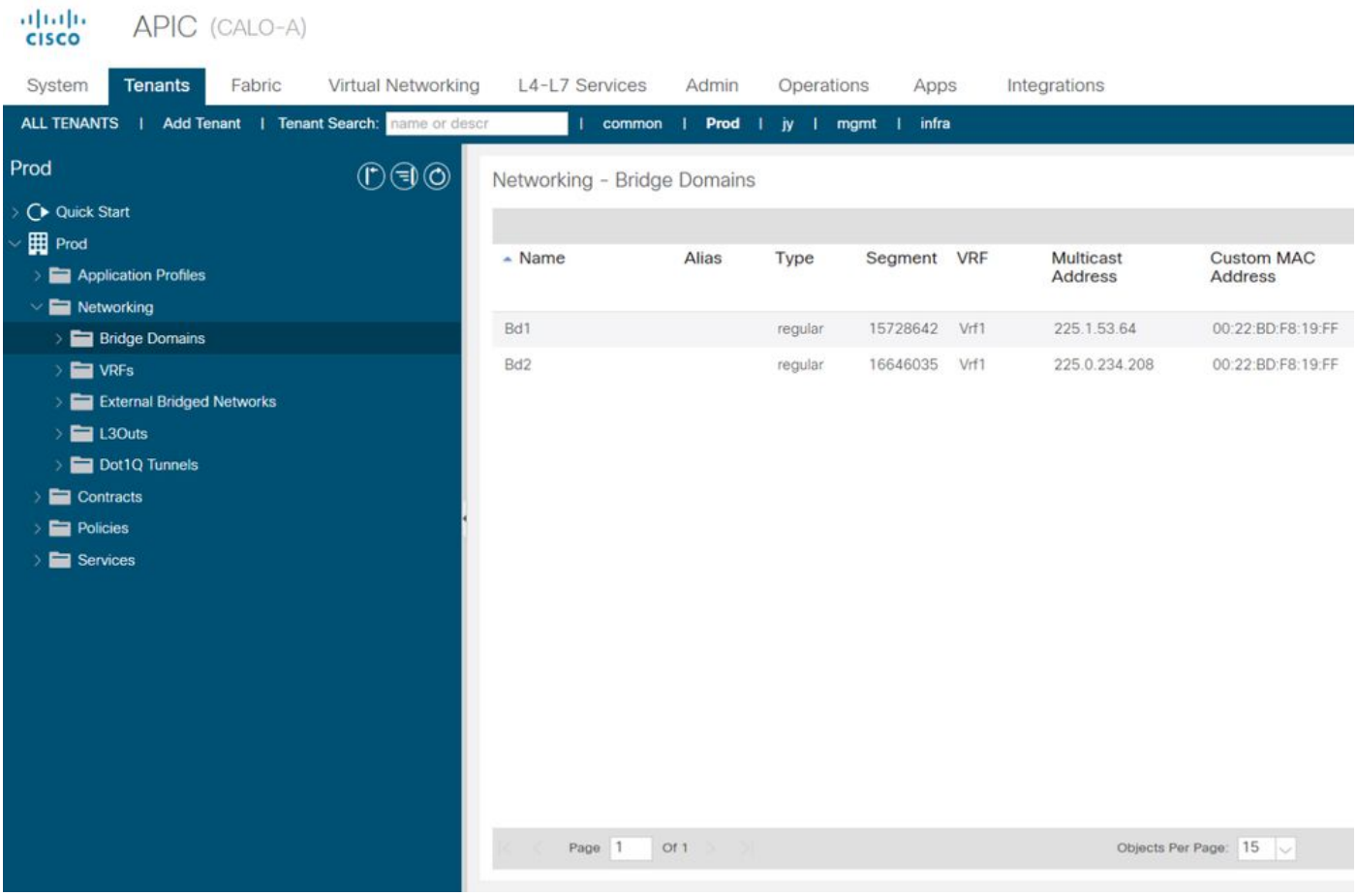
En ACI, el tráfico se inunda a través de grupos de multidifusión superpuestos en muchos escenarios diferentes. Por ejemplo, la inundación ocurre para:

- Tráfico de multidifusión y difusión.
- Unidifusión desconocida que debe saturarse.
- Mensajes de limpieza ARP de fabric.
- Mensajes de anuncio de EP.

Muchas funciones y funcionalidades dependen del reenvío de BUM.

En ACI, a todos los dominios de puente se les asigna una dirección de multidifusión conocida como dirección IP externa (o GIPo) de grupo. Todo el tráfico que debe inundarse dentro de un dominio de puente se inunda en este GIPo.

BD GIPo en GUI



El objeto se puede consultar directamente en uno de los APIC.

BD GIPo en Moquery

```
a-apic1# moquery -c fvBD -f 'fv.BD.name=="Bd1"'
Total Objects shown: 1

# fv.BD
name                     : Bd1
OptimizeWanBandwidth     : no
annotation               :
arpFlood                 : yes
bcastP                   : 225.1.53.64
childAction              :
configIssues             :
descr                    :
dn                       : uni/tn-Prod/BD-Bd1
epClear                  : no
epMoveDetectMode         :
extMngdBy                :
hostBasedRouting         : no
intersiteBumTrafficAllow : no
intersiteL2Stretch       : no
ipLearning               : yes
ipv6McastAllow           : no
lcOwn                    : local
limitIpLearnToSubnets    : yes
llAddr                   : ::
mac                      : 00:22:BD:F8:19:FF
mcastAllow               : no
modTs                    : 2019-09-30T20:12:01.339-04:00
monPolDn                 : uni/tn-common/monepg-default
mtu                      : inherit
multiDstPktAct           : bd-flood
nameAlias                :
ownerKey                 :
ownerTag                 :
pcTag                    : 16387
rn                       : BD-Bd1
scope                    : 2392068
seg                      : 15728642
status                   :
type                     : regular
uid                      : 16011
unicastRoute             : yes
unkMacUcastAct           : proxy
unkMcastAct              : flood
v6unkMcastAct            : flood
vmac                     : not-applicable
```

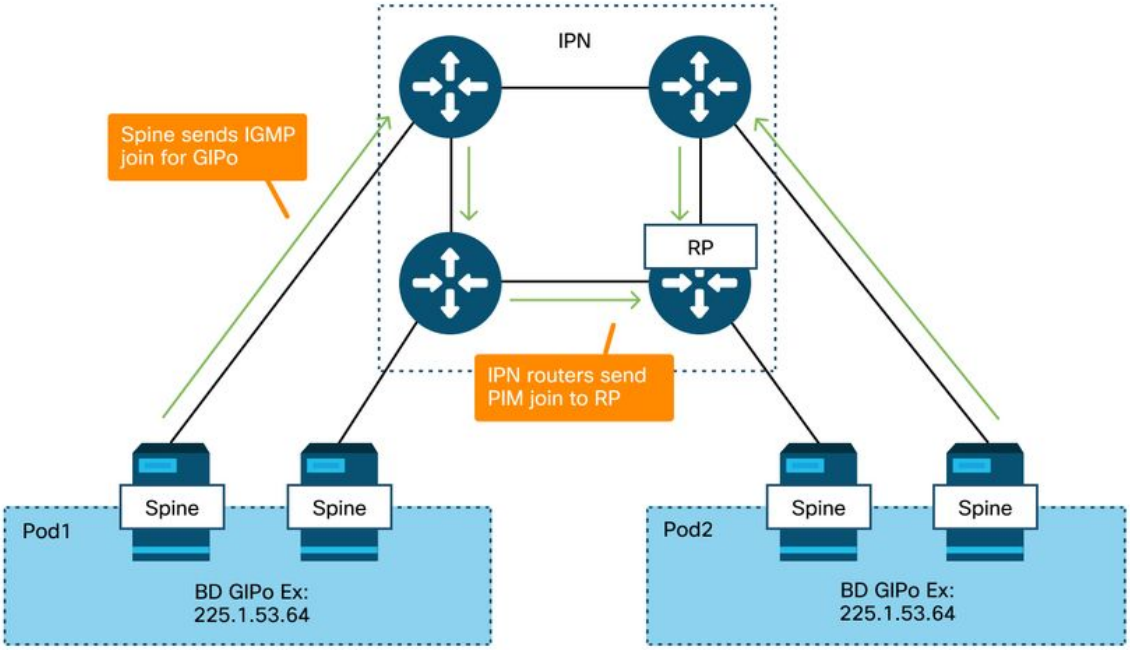
La información anterior sobre la inundación GIPo es verdadera independientemente de si Multi-Pod se utiliza o no. La parte adicional de esto que pertenece a Multi-Pod es el ruteo multicast en el IPN.

El ruteo de multidifusión IPN implica lo siguiente:

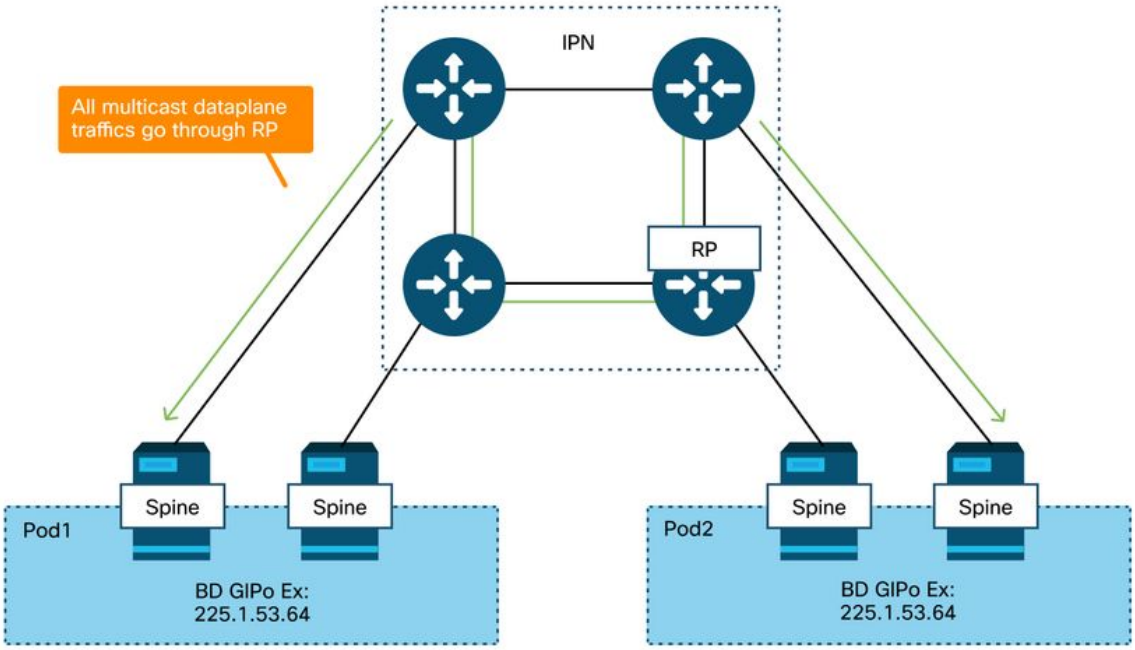
- Los nodos de columna actúan como hosts multidifusión (solo IGMP). No ejecutan PIM.
- Si se implementa un BD en un Pod, una columna de ese POD enviará una unión IGMP en una de sus interfaces de cara a IPN. Esta funcionalidad se divide en bandas en todos los nodos de columna y la interfaz orientada a IPN en muchos grupos.
- Los IPN reciben estas uniones y envían uniones PIM hacia el RP PIM bidireccional.
- Como se utiliza PIM Bidir, no hay árboles (S,G). Sólo se utilizan árboles (*,G) en PIM Bidir.

- Todo el tráfico del plano de datos enviado al GIPo pasa a través del RP.

Plano de control de multidifusión IPN



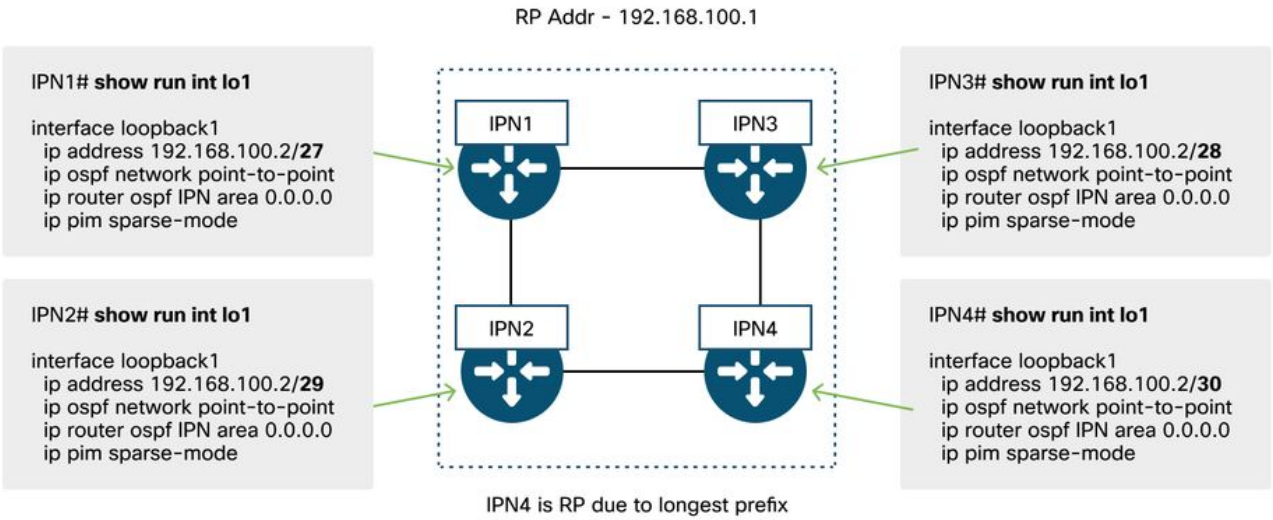
plano de datos multidifusión IPN



El único medio de redundancia RP con PIM Bidir es utilizar Phantom. Esto se trata en detalle en la parte de este libro dedicada a la detección de varios dispositivos. Como resumen rápido, tenga en cuenta que con Phantom RP:

- Todos los IPN deben configurarse con la misma dirección RP.
- La dirección RP exacta no debe existir en ningún dispositivo.
- Varios dispositivos anuncian la disponibilidad a la subred que contiene la dirección IP de RP fantasma. Las subredes anunciadas deben variar en longitud de subred para que todos los routers se pongan de acuerdo sobre quién anuncia la mejor trayectoria para el RP. Si se pierde esta trayectoria, la convergencia depende del IGP.

Configuración de RP fantasma



Flujo de trabajo de solución de problemas de difusión de varios dispositivos, unidifusión desconocida y multidifusión (BUM)

1. Primero confirme si el flujo está siendo tratado realmente como multidestino por el entramado.

El flujo se inundará en el BD en estos ejemplos comunes:

- La trama es una transmisión ARP y la inundación ARP está habilitada en el BD.
- La trama está destinada a un grupo multicast. Tenga en cuenta que incluso si se habilita la indagación IGMP, el tráfico siempre se inunda en el entramado del GIPo.
- El tráfico está destinado a un grupo de multidifusión para el que ACI proporciona servicios de routing de multidifusión.
- El flujo es de Capa 2 (flujo puenteado) y la dirección MAC de destino es desconocida y el comportamiento de unidifusión desconocido en el BD se establece en 'Inundación'.

La manera más fácil de determinar qué decisión de reenvío se tomará es con un ELAM.

2. Identifique el GIPo BD.

Consulte la sección anterior de este capítulo que trata sobre este tema. Los ELAM de columna

también se pueden ejecutar a través de la aplicación ELAM Assistant para verificar que se reciba el tráfico inundado.

3. Verifique las tablas de ruteo multicast en el IPN para ese GIPo.

Los resultados para hacer esto variarían dependiendo de la plataforma IPN en uso, pero a un nivel alto:

- Todos los routers IPN deben coincidir en el RP y el RPF para este GIPo debe apuntar a este árbol.
- Un router IPN conectado a cada POD debe obtener una unión IGMP para el grupo.

Situación de resolución de problemas de varios dispositivos #2 (BUM Flow)

Este escenario cubriría cualquier escenario que implique que ARP no se resuelva en los escenarios Multi-Pod o BUM (unidifusión desconocida, etc.).

Hay varias causas posibles comunes aquí.

Posible causa 1: Varios routers poseen la dirección RP de PIM

Con este escenario, la hoja de ingreso inunda el tráfico (verificar con ELAM), el POD de origen recibe e inunda el tráfico, pero el POD remoto no lo obtiene. Para algunos BD, la inundación funciona, pero para otros no.

En el IPN, ejecute 'show ip mroute <GIPo address>' para que el GIPo vea que el árbol RPF apunta a varios routers diferentes.

En este caso, compruebe lo siguiente:

- Verifique que la dirección RP PIM real no esté configurada en ninguna parte. Cualquier dispositivo que posea esa dirección RP real verá una ruta /32 local para ella.
- Verifique que varios routers IPN no anuncien la misma longitud de prefijo para el RP en el escenario de RP fantasma.

Posible causa 2: Los routers IPN no están aprendiendo rutas para la dirección RP

De la misma manera que la primera causa posible, aquí el tráfico inundado no logra salir del IPN. La salida de 'show ip route <rp address>' en cada router IPN mostraría solamente la longitud de prefijo configurada localmente en lugar de lo que anuncian los otros routers.

El resultado de esto es que cada dispositivo piensa que es el RP aunque la dirección IP real del RP no esté configurada en ninguna parte.

Si este es el caso. verifique lo siguiente:

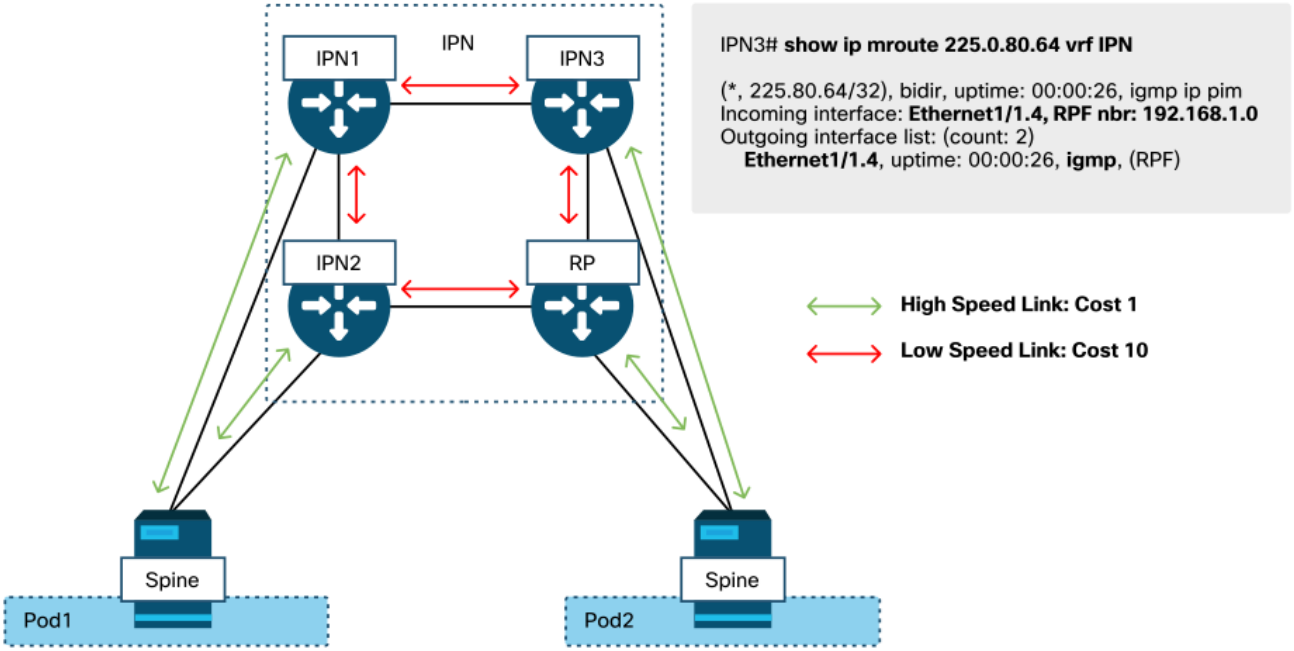
- Verifique que las adyacencias de ruteo estén activas entre los routers IPN. Verifique que la ruta esté en la base de datos del protocolo real (como la base de datos OSPF).
- Verifique que todos los loopbacks que se supone que son RP candidatos estén configurados como tipos de red punto a punto OSPF. Si este tipo de red no está configurado, cada router

siempre anunciará una longitud de prefijo /32 independientemente de lo que esté realmente configurado.

Posible causa 3: Los routers IPN no están instalando la ruta GIPo o los puntos RPF en ACI

Como se ha mencionado anteriormente, ACI no ejecuta PIM en sus enlaces orientados a IPN. Esto significa que la mejor trayectoria del IPN hacia el RP nunca debe apuntar a ACI. La situación en la que esto podría suceder sería si se conectan varios routers IPN a la misma columna y se observa una mejor métrica OSPF a través de la columna que directamente entre routers IPN.

Interfaz RPF hacia ACI



Para resolver este problema:

- Asegúrese de que las adyacencias del protocolo de ruteo entre los routers IPN estén activas.
- Aumente las métricas de costo OSPF para los links orientados a IPN en los nodos de columna a un valor que hará que esa métrica sea menos preferible que los links IPN a IPN.

Otras referencias

Antes de la versión 4.0 del software ACI, se experimentaban algunos retos con respecto al uso de COS 6 por parte de dispositivos externos. La mayoría de estos problemas se han resuelto mediante mejoras de la versión 4.0, pero para obtener más información, consulte la sesión de CiscoLive "BRKACI-2934 - Resolución de problemas de varios dispositivos" y la sección "Calidad del servicio".

Acerca de esta traducción
Cisco ha traducido este documento combinando la traducción automática y los recursos humanos a fin de ofrecer a nuestros usuarios en todo el mundo contenido en su propio idioma.

Tenga en cuenta que incluso la mejor traducción automática podría no ser tan precisa como la proporcionada por un traductor profesional.

Cisco Systems, Inc. no asume ninguna responsabilidad por la precisión de estas traducciones y recomienda remitirse siempre al documento original escrito en inglés (insertar vínculo URL).