

# Solucionar problemas de encaminhamento de estrutura interna da ACI - encaminhamento de vários pods

## Contents

[Introduction](#)

[Informações de Apoio](#)

[Visão geral do encaminhamento de vários pods](#)

[Componentes do Multi-Pod](#)

[Topologia para Exemplos de Multipods](#)

[Fluxo de trabalho geral para solução de problemas de encaminhamento de vários pods](#)

[Fluxo de trabalho de solução de problemas de unicast de vários pods](#)

[1. Confirme se a folha de entrada recebe o pacote. Use a ferramenta CLI do ELAM mostrada na seção "Ferramentas" junto com a saída de relatório disponível no 4.2. O aplicativo Assistente do ELAM também é usado.](#)

[2. A folha de entrada está aprendendo o destino como um endpoint no VRF de entrada? Em caso negativo, existe uma rota?](#)

[Configuração do ELAM Assistant](#)

[Verificar decisões de encaminhamento](#)

[3. Confirme no spine que o IP de destino está presente no COOP para que a solicitação proxy funcione.](#)

[4. Decisão de encaminhamento de proxy de spine de vários pods](#)

[5. Verificar o BGP EVPN na coluna](#)

[6. Verifique o COOP nos spines do Pod de destino.](#)

[7. Verifique se a folha de saída tem o aprendizado local.](#)

[Usando Triagem para verificar o fluxo fim-a-fim](#)

[Solicitações com proxy em que o EP não está no COOP](#)

[Verificação Glean ARP](#)

[Cenário #1 de Troubleshooting de Multipods \(Unicast\)](#)

[Topologia de solução de problemas](#)

[Causa: Ponto de Extremidade Ausente no COOP](#)

[Outras causas possíveis](#)

[Visão geral do encaminhamento de broadcast de vários pods, unicast desconhecido e multicast \(BUM\)](#)

[BD GIPo na GUI](#)

[Plano de controle multicast IPN](#)

[Painel de dados multicast IPN](#)

[Configuração do RP fantasma](#)

[Fluxo de trabalho de solução de problemas de transmissão multipods, unicast desconhecido e multicast \(BUM\)](#)

[1. Primeiro, confirme se o fluxo está realmente sendo tratado como multidestino pela malha.](#)

[2. Identifique o BD GIPo.](#)

[3. Verifique as tabelas de roteamento multicast no IPN para esse GIPo.](#)

## [Cenário #2 de Troubleshooting de Multipods \(Fluxo BUM\)](#)

[Causa possível 1: Vários roteadores possuem o endereço PIM RP](#)

[Causa possível 2: Os roteadores IPN não estão aprendendo rotas para o endereço RP](#)

[Causa possível 3: Os roteadores IPN não estão instalando a rota GIPO ou os pontos RPF para a ACI](#)

[Outras referências](#)

## Introduction

Este documento descreve as etapas para entender e solucionar problemas de um cenário de encaminhamento de vários pods da ACI.

## Informações de Apoio

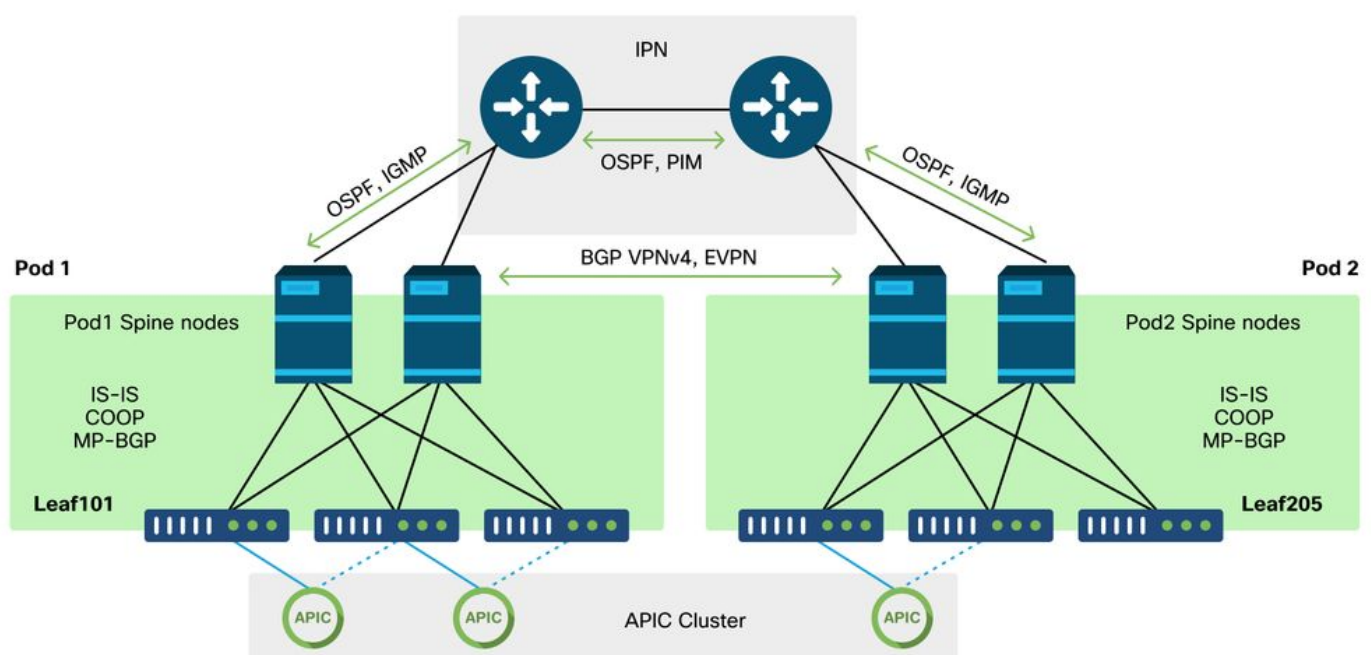
O material deste documento foi extraído do [Solução de problemas da Cisco Application Centric Infrastructure, segunda edição](#) livro, especificamente o **Encaminhamento dentro da estrutura - Encaminhamento de vários pods** capítulo.

## Visão geral do encaminhamento de vários pods

Este capítulo abordará como solucionar problemas em cenários nos quais a conectividade não está funcionando corretamente em pods em um ambiente de vários pods

Antes de analisar exemplos específicos de solução de problemas, é importante reservar um tempo para entender os componentes do Multi-Pod em um alto nível.

## Componentes do Multi-Pod



Semelhante a uma estrutura de ACI tradicional, uma estrutura de vários pods ainda é considerada

uma única estrutura de ACI e depende de um único cluster de APIC para gerenciamento.

Dentro de cada Pod individual, a ACI aproveita os mesmos protocolos na camada que uma estrutura tradicional. Isso inclui IS-IS para troca de informações TEP, bem como seleção de Interface de Saída (OIF - Outgoing Interface) multicast, COOP para um repositório de endpoint global e BGP VPNv4 para a distribuição de roteadores externos através da estrutura.

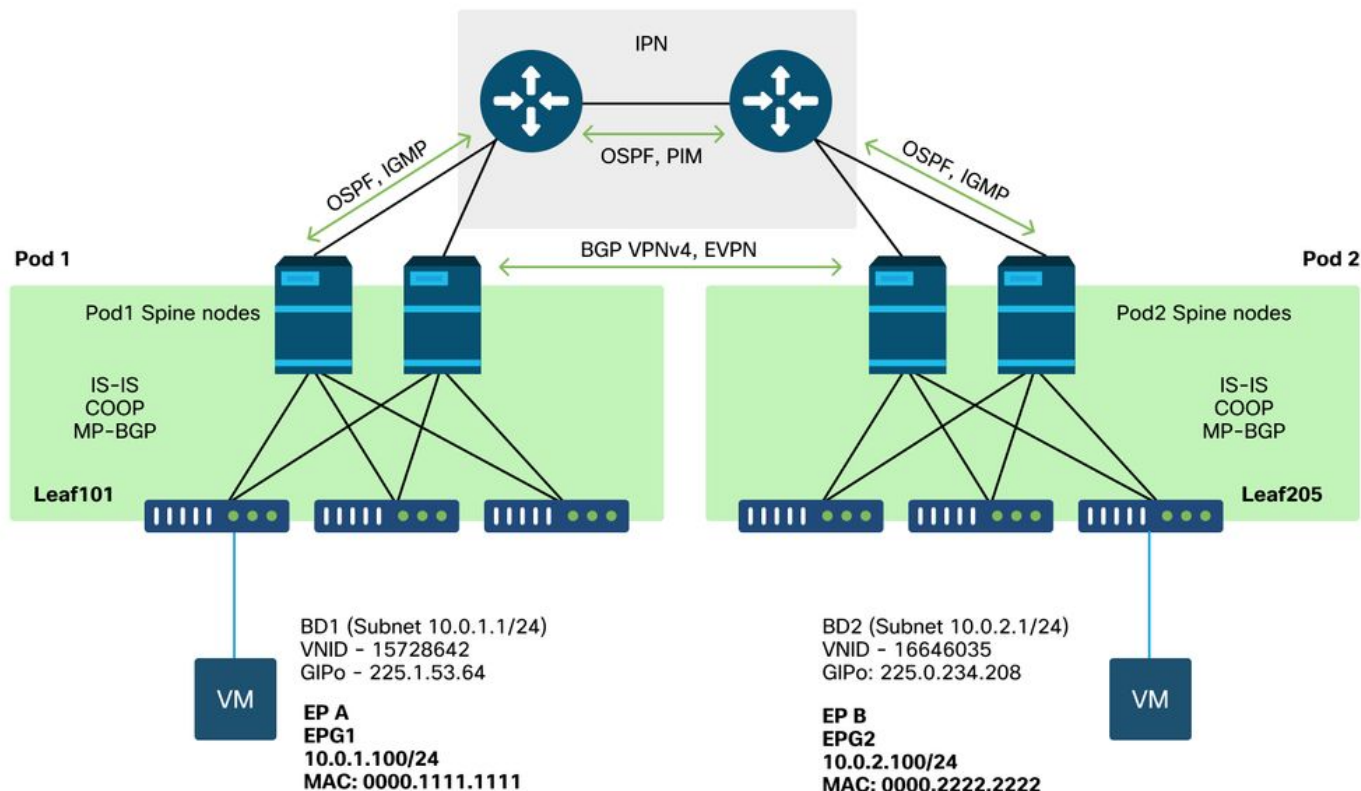
O Multi-Pod se baseia nesses componentes, pois deve conectar cada Pod.

- Para trocar informações de roteamento sobre TEPs no Pod remoto, o OSPF é usado para anunciar o pool TEP de resumo através do IPN.
- Para trocar rotas externas aprendidas de um Pod para outro, a família de endereços BGP VPNv4 é estendida entre nós spine. Cada Pod se torna um cluster de refletor de rota separado.
- Para sincronizar endpoints, bem como outras informações armazenadas no COOP através de Pods, a família de endereços BGP EVPN é estendida entre nós spine.
- Por fim, para lidar com a inundação de tráfego de Broadcast, Unknown-Unicast e Multicast (BUM) entre os Pods, os nós spine em cada Pod atuam como hosts IGMP e os roteadores IPN trocam informações de roteamento multicast através do PIM Bidirecional.

Grande parte dos cenários de solução de problemas e fluxos de trabalho de vários pods é semelhante às estruturas da ACI de pods únicos. Esta seção de vários pods se concentrará principalmente nas diferenças entre o encaminhamento de um único pod e de vários pods.

## Topologia para Exemplos de Multipods

Como ocorre com a solução de problemas de qualquer cenário, é importante começar compreendendo qual é o estado esperado. Consulte esta topologia para obter os exemplos deste capítulo.



## Fluxo de trabalho geral para solução de problemas de encaminhamento de vários pods

Em um alto nível, ao depurar um problema de encaminhamento de vários pods, as seguintes etapas podem ser avaliadas:

1. O fluxo é unicast ou multidestino? Lembre-se, mesmo que se espere que o fluxo seja unicast no estado de funcionamento, se o ARP não for resolvido, é um fluxo de vários destinos.
2. O fluxo é roteado ou interligado? Tradicionalmente, um fluxo roteado de uma perspectiva da ACI seria qualquer fluxo em que o endereço MAC destino é o endereço MAC do roteador que pertence a um gateway configurado na ACI. Além disso, se a inundação ARP estiver desativada, a folha de entrada será roteada com base no endereço IP de destino. Se o endereço MAC destino não pertencer à ACI, o switch encaminharia com base no endereço MAC ou seguiria o comportamento 'unicast desconhecido' configurado no domínio da bridge.
3. A folha de entrada está diminuindo o fluxo? A triagem e o ELAM são as melhores ferramentas para confirmar isso.

### Se o fluxo for unicast da camada 3:

1. A folha de entrada tem um ponto final aprendido para o IP de destino no mesmo VRF que o EPG de origem? Em caso afirmativo, isso sempre terá precedência sobre qualquer rota aprendida. O leaf encaminhará diretamente ao endereço do túnel ou à interface de saída onde o endpoint é aprendido.
2. Se não houver nenhum ponto final aprendido, a folha de ingresso tem uma rota para o destino que tem o sinalizador 'Pervasivo' definido? Isso indica que a sub-rede de destino está configurada como uma sub-rede de Domínio de Bridge e que o próximo salto deve ser



-----  
-----  
Outer Packet Attributes  
-----  
-----

Outer Packet Attributes : 12uc 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

Há muito mais informações no relatório sobre para onde o pacote está indo, mas o aplicativo Assistente do ELAM atualmente é mais útil para interpretar esses dados. A saída do Assistente do ELAM para esse fluxo será mostrada posteriormente neste capítulo.

## 2. A folha de entrada está aprendendo o destino como um endpoint no VRF de entrada? Em caso negativo, existe uma rota?

```
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  
+-----+-----+-----+-----+-----+  
+-----+  
+-----+
```

Nenhuma saída no comando acima significa que o IP de destino não foi aprendido. Em seguida, verifique a tabela de roteamento.

```

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

```

Na saída acima, o flag pervasivo é visto indicando que essa é uma rota de sub-rede de domínio de bridge. O próximo salto deve ser um endereço proxy anycast nos spines.

```

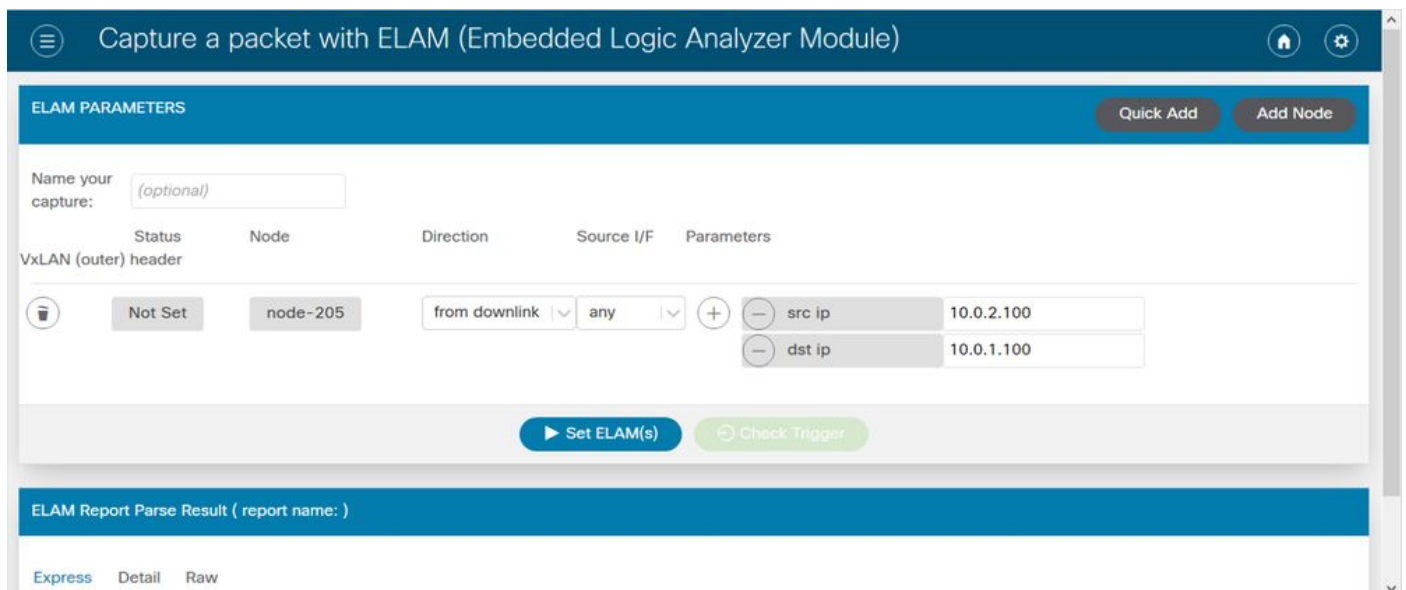
a-leaf205# show isis dtep vrf overlay-1 | grep 10.0.120.34
10.0.120.34      SPINE    N/A      PHYSICAL,PROXY-ACAST-V4

```

Observe que se o ponto final for aprendido em um túnel ou interface física, isso terá precedência, fazendo com que o pacote seja encaminhado diretamente para lá. Consulte o capítulo "Encaminhamento externo" deste manual para obter mais detalhes.

Use o ELAM Assistant para confirmar as decisões de encaminhamento vistas nas saídas acima.

## Configuração do ELAM Assistant



## Verificar decisões de encaminhamento

Forward Result	
Destination Type	To another ACI node (LEAF, AVS/AVE etc.)
Destination TEP	10.0.120.34 (IPv4 Spine-Proxy)
Destination Physical Port	eth1/53
Contract	
Destination EPG pcTag (dclass)	0x1 / 1 (pcTag 1 is to ignore contract for special packets such as Spine-Proxy, ARP, Multicast etc..)
Source EPG pcTag (sclass)	0xC001 / 49153 (Prod.ap1:epg2)
Contract was applied	0 (Contract was not applied on this node)
Drop	
Drop Code	no drop

A saída acima mostra que o leaf de entrada está encaminhando o pacote para o endereço proxy spine IPv4. É o que se espera que aconteça.

### 3. Confirme no spine que o IP de destino está presente no COOP para que a solicitação proxy funcione.

Há várias maneiras de obter a saída COOP na coluna, por exemplo, examiná-la com um 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
-----
```

Outros comandos a serem executados no spine:

#### Consultar COOP para entrada I2:

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

#### Consulte COOP para entrada I3 e obtenha entrada I2 pai:



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

### Consultar COOP somente para entrada I3:

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

O que é útil sobre a consulta múltipla é que eles também podem ser executados diretamente em um APIC e o usuário pode ver cada coluna que tem o registro em coop.

## 4. Decisão de encaminhamento de proxy de spine de vários pods

Se a entrada COOP da coluna aponta para um túnel no Pod local, o encaminhamento é baseado no comportamento tradicional da ACI.

Observe que o proprietário de um TEP pode ser verificado na estrutura executando-se de um APIC: `moquery -c ipv4Addr -f 'ipv4.Addr.addr=="<tunnel address>"'`

No cenário proxy, o próximo salto do túnel é 10.0.0.34. Quem é o proprietário desse endereço 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]
```

Esse IP pertence aos dois nós spine no Pod 1. Esse é um IP específico chamado endereço de proxy externo. Da mesma forma que a ACI tem endereços proxy de propriedade dos nós spine dentro de um Pod (consulte a etapa 2 desta seção), também há endereços proxy atribuídos ao próprio Pod. Esse tipo de interface pode ser verificado executando-se:

```
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]
```

O sinalizador 'external' indica que este é um TEP de proxy externo.

## 5. Verificar o BGP EVPN na coluna

O registro de ponto final de cooperação deve ser importado do BGP EVPN na coluna. O comando a seguir pode ser usado para verificar se está em EVPN (embora se já estiver em COOP com um próximo salto do TEP de proxy externo do Pod remoto, pode-se supor que veio 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
```

Observe que o comando acima também pode ser executado para um endereço MAC.

-192.168.1.254 é o TEP do plano de dados configurado durante a configuração do Multi-Pod. Observe, no entanto, que mesmo que seja anunciado no BGP como o NH, o próximo salto real será o TEP de proxy externo.

-192.168.1.101 e .102 são os nós spine do Pod 1 que anunciam esse caminho.

## 6. Verifique o COOP nos spines do Pod de destino.

O mesmo comando que o anterior pode ser usado:

```
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 quem é o proprietário do endereço do túnel executando o seguinte comando em um 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

```

O comando acima mostra que o túnel do COOP aponta para o leaf101. Isso significa que o leaf101 deve ter o aprendizado local para o endpoint de destino.

## 7. Verifique se a folha de saída tem o aprendizado local.

Isso pode ser feito por meio de um 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/ Interface Domain Info	Endpoint Group Info	Encap VLAN	MAC Address IP Address	MAC Info/ IP
341 po5	Prod:apl:epg1	vlan-1075	0000.1111.1111	LV
Prod:Vrf1 po5		vlan-1075	10.0.1.100	LV

Observe que o endpoint é aprendido. O pacote deve ser encaminhado com base no canal de

porta 5 com a marca de VLAN 1075 definida.

## Usando Triagem para verificar o fluxo fim-a-fim

Conforme discutido na seção "Ferramentas" deste capítulo, a Triagem pode ser usada para mapear um fluxo existente de ponta a ponta e entender o que cada switch no caminho está fazendo com o pacote. Isso é particularmente útil em implantações maiores e mais complexas, como o Multi-Pod.

Observe que a triagem levará algum tempo para ser totalmente executada (possivelmente 15 minutos).

Ao executar Triagem no fluxo de exemplo:

```
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:Vrfl
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\_nextthop 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 encaps 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:Vrfl

2019-10-01 16:12:34,560 INFO ftriage: main:332 Egress BD(s): Prod:Bdl

2019-10-01 16:12:37,704 INFO ftriage: unicast:1252 a-leaf101: Enter dbg\_sub\_nextthop 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\_nextthop 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"}}}
```

Há uma grande quantidade de dados na Triagem. Alguns dos campos mais importantes são destacados. Observe que o caminho do pacote foi 'leaf205 (Pod 2) > spine3 (Pod 2) > spine2 (Pod 1) > leaf101 (Pod 1)'. Todas as decisões de encaminhamento e pesquisas de contrato feitas ao longo do caminho também são visíveis.

Observe que, se esse fosse um fluxo de Camada 2, a sintaxe da Triagem precisaria ser definida como:

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

## Solicitações com proxy em que o EP não está no COOP

Antes de considerar cenários de falha específicos, há mais uma parte a ser discutida relacionada ao encaminhamento unicast sobre Multi-Pod. O que acontece se o endpoint de destino for desconhecido, a solicitação for submetida a proxy e o endpoint não estiver em COOP?

Neste cenário, o pacote/quadro é enviado para o spine e uma solicitação de glean é gerada.

Quando o spine gera uma solicitação glean, o pacote original ainda é preservado na solicitação, no entanto, o pacote recebe o ethertype 0xfff2, que é um Ethertype personalizado reservado para gleans. Por esse motivo, não será fácil interpretar essas mensagens em ferramentas de captura de pacotes, como o Wireshark.

O destino da camada 3 externa também é definido como 239.255.255.240, que é um grupo multicast reservado especificamente para mensagens glean. Eles devem ser despejados na estrutura e qualquer switch leaf de saída que tenha a sub-rede de destino da solicitação glean implantada gerará uma solicitação ARP para resolver o destino. Esses ARPs são enviados do endereço IP de sub-rede de BD configurado (portanto, as solicitações de proxy não podem resolver o local de pontos finais silenciosos/desconhecidos se o roteamento unicast estiver desativado em um domínio de ponte).

A recepção da mensagem de glean na folha de saída e o ARP gerado subsequentemente e a resposta ARP recebida podem ser verificadas através do seguinte comando:

## Verificação Glean ARP

```
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

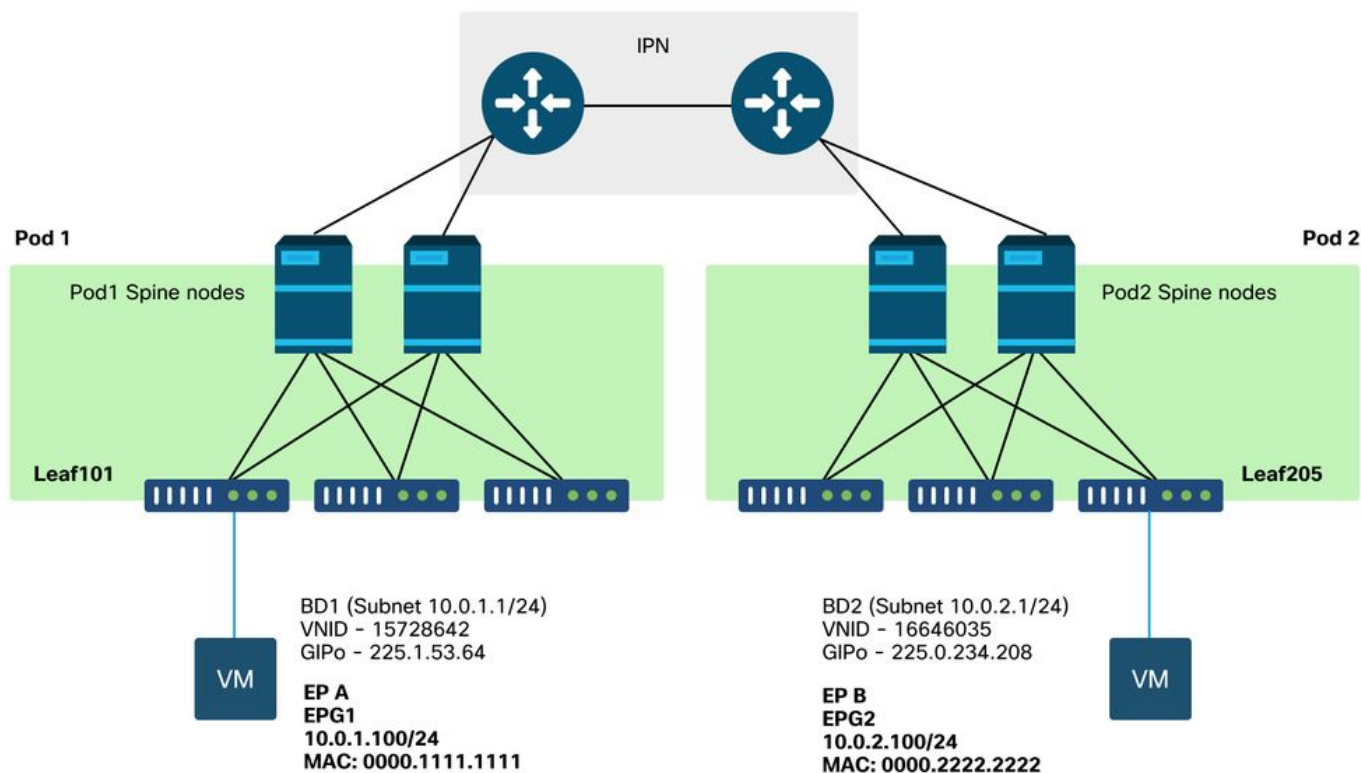
```
79) Event:E_DEBUG_DSF, length:174, at 131748 usecs after Wed May 1 08:28:36 2019  
log_collect_arp_glean; dip = 192.168.21.11; interface = Vlan104; vrf = CiscoLive2019:vrf1; info  
= Address in PSVI subnet or special VIP <<< Glean Received, Dst IP is in BD subnet
```

Para referência, mensagens glean sendo enviadas para 239.255.255.240 é o motivo pelo qual esse grupo precisa ser incluído no intervalo de grupo PIM bidirecional no IPN.

## Cenário #1 de Troubleshooting de Multipods (Unicast)

Na topologia a seguir, o EP B não pode se comunicar com o EP A.

### Topologia de solução de problemas



Observe que muitos dos problemas observados no encaminhamento de vários pods são idênticos aos problemas observados em um único pod. Por esse motivo, os problemas específicos do Multi-Pod são focados.

Ao seguir o fluxo de trabalho de solução de problemas de unicast descrito anteriormente, observe que a solicitação tem proxy, mas os nós spine no Pod 2 não têm o IP de destino no COOP.

### Causa: Ponto de Extremidade Ausente no COOP

Conforme discutido anteriormente, as entradas COOP para endpoints Pod remotos são preenchidas a partir de informações BGP EVPN. Consequentemente, é importante determinar:

r.) O spine do Pod (Pod 2) de origem o tem no EVPN?

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

b.) A coluna do Pod (Pod 1) remoto a tem no 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:

O spine do Pod 1 tem isso e o IP do próximo salto é 0.0.0.0; isso significa que ele foi exportado do COOP localmente. Observe, no entanto, que a seção "Anunciado aos peers" não inclui os nós spine do Pod 2.

c.) O BGP EVPN está ativo entre os 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 na saída acima que os peers BGP EVPN estão desativados entre os pods. Qualquer coisa além de um valor numérico na coluna State/PfxRcd indica que a adjacência não está ativa. Os EPs do Pod 1 não são aprendidos através do EVPN e não são importados para o COOP.

Se esse problema for observado, verifique o seguinte:

1. O OSPF está ativo entre os nós spine e os IPNs conectados?
2. Os nós spine têm rotas aprendidas através do OSPF para os IPs spine remotos?
3. O caminho completo através do IPN suporta MTU jumbo?
4. Todas as adjacências de protocolo são estáveis?

## Outras causas possíveis

Se o endpoint não estiver no banco de dados COOP de qualquer Pod e o dispositivo de destino for um host silencioso (não aprendido em nenhum switch de folha na malha), verifique se o processo de limpeza de malha está funcionando corretamente. Para que isso funcione:

- O roteamento unicast deve ser ativado no BD.
- O destino deve estar em uma sub-rede BD.



- O IPN deve fornecer serviço de roteamento multicast para o grupo 239.255.255.240. A parte do multicast é abordada mais na próxima seção.

## Visão geral do encaminhamento de broadcast de vários pods, unicast desconhecido e multicast (BUM)

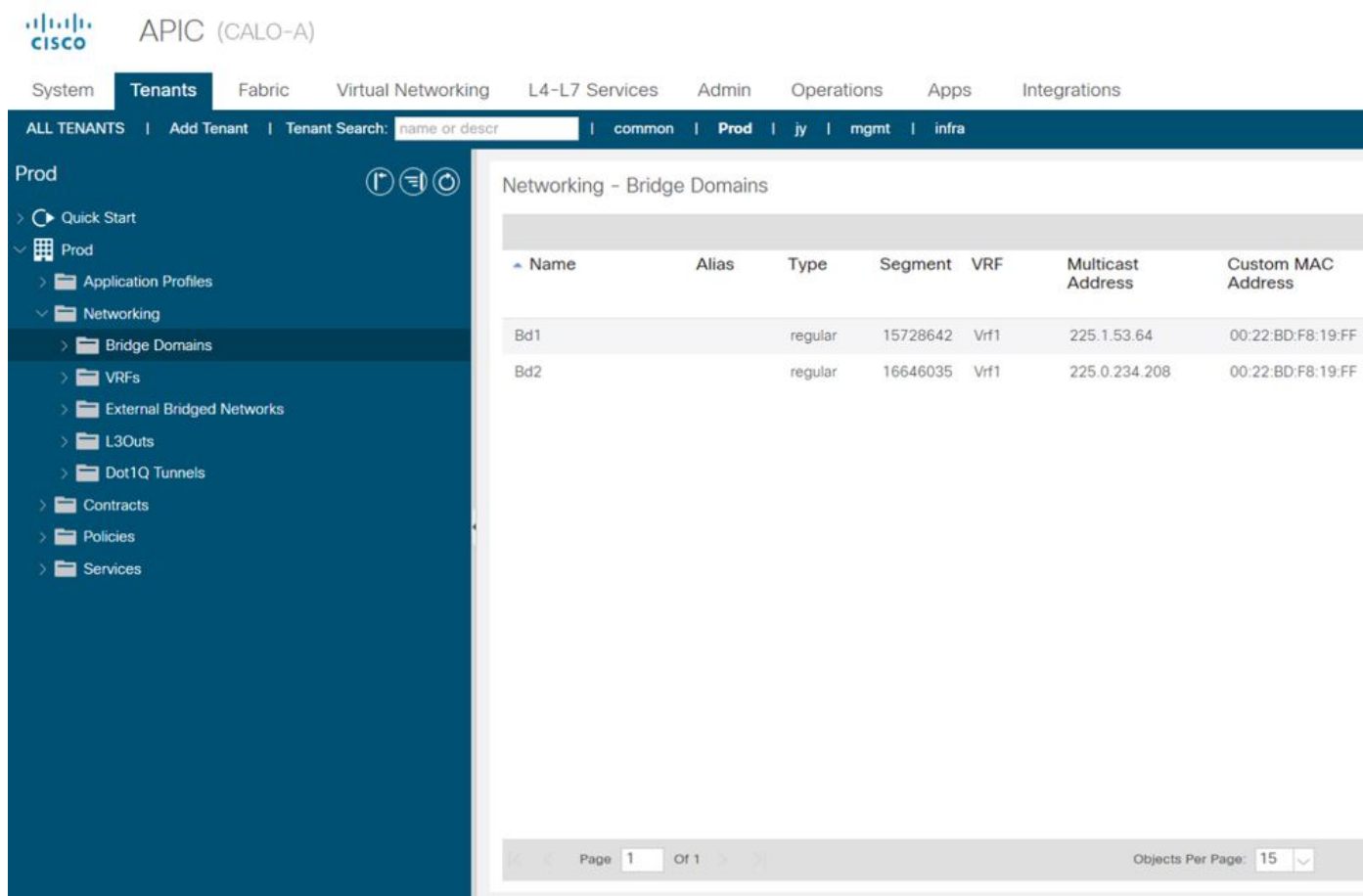
Na ACI, o tráfego é inundado por grupos multicast de sobreposição em vários cenários diferentes. Por exemplo, a inundação ocorre para:

- Multicast e tráfego de broadcast.
- Unicast desconhecido que deve ser inundado.
- Mensagens de limpeza ARP de malha.
- O PE anuncia mensagens.

Muitos recursos e funcionalidades dependem do encaminhamento BUM.

Dentro da ACI, todos os domínios de bridge recebem um endereço multicast conhecido como endereço GIPo (Group IP Outer). Todo o tráfego que deve ser despejado dentro de um Domínio de Bridge é despejado nesse GIPo.

## BD GIPo na GUI



The screenshot shows the Cisco APIC (CALO-A) GUI. The navigation menu on the left includes 'Prod' > 'Networking' > 'Bridge Domains'. The main content area displays a table titled 'Networking - Bridge Domains' with the following data:

Name	Alias	Type	Segment	VRF	Multicast Address	Custom MAC Address
Bd1		regular	15728642	Vrf1	225.1.53.64	00:22:BD:F8:19:FF
Bd2		regular	16646035	Vrf1	225.0.234.208	00:22:BD:F8:19:FF

At the bottom of the table, there is a pagination bar showing 'Page 1 Of 1' and 'Objects Per Page: 15'.

O objeto pode ser consultado diretamente em um dos APICs.

## BD GIPo em 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
```

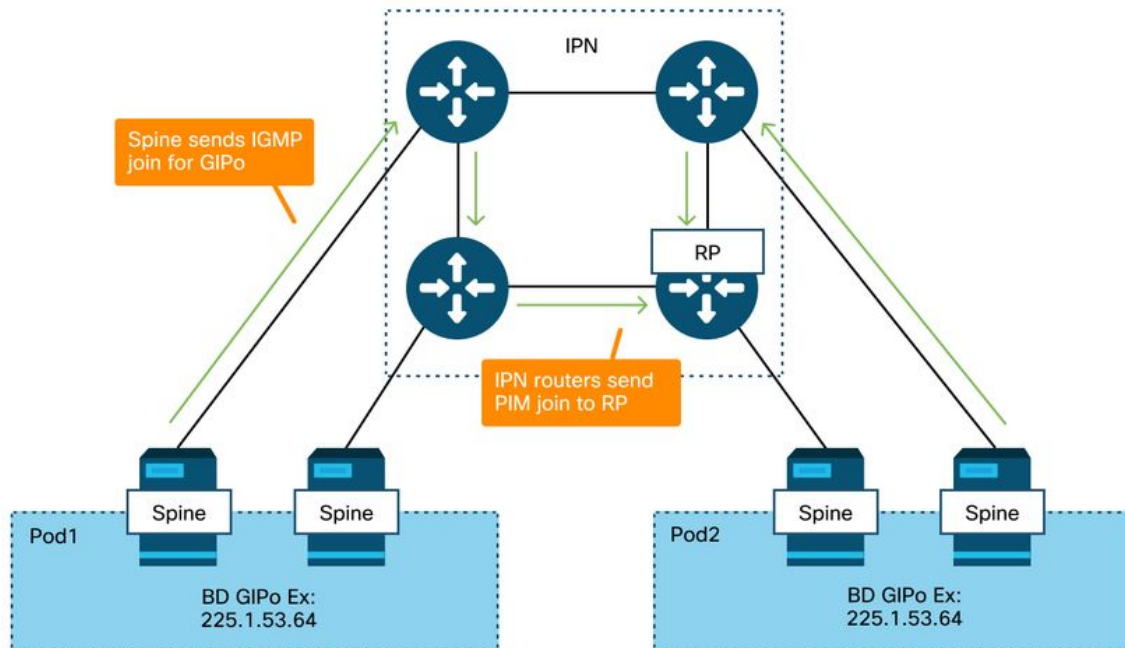
As informações acima sobre a inundação de GIPo são verdadeiras, independentemente do Multi-Pod ser usado ou não. A parte adicional disso que pertence ao Multi-Pod é o roteamento multicast no IPN.

O roteamento multicast IPN envolve o seguinte:

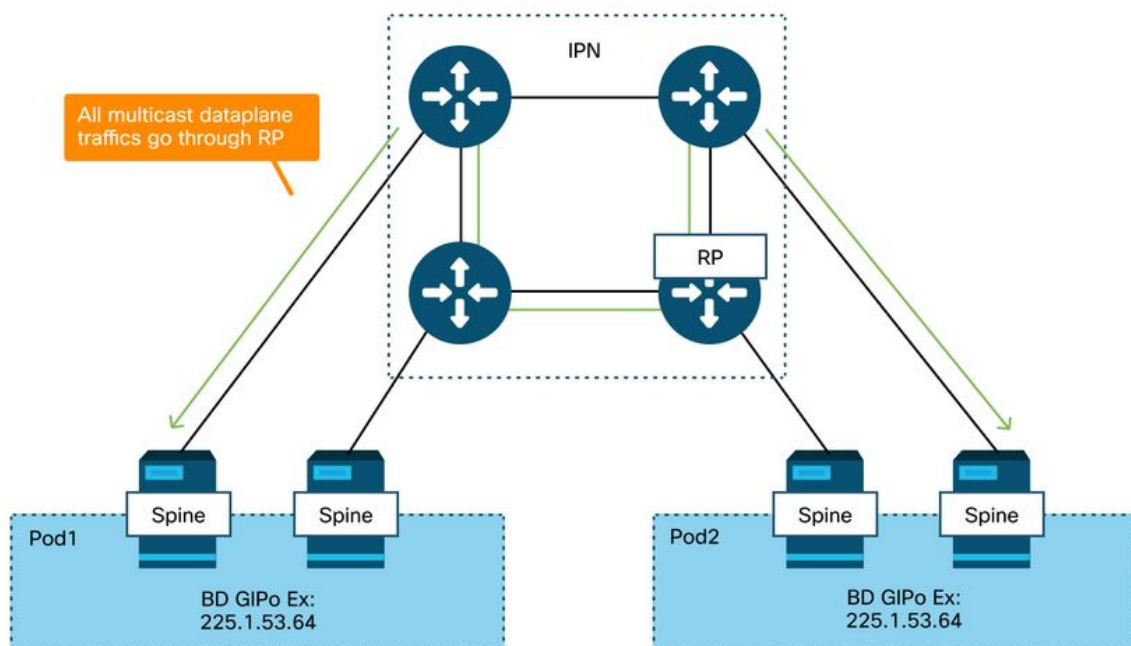
- Os nós spine atuam como hosts multicast (somente IGMP). Eles não executam PIM.
- Se um BD for implantado em um Pod, um spine desse pod enviará uma união IGMP em uma de suas interfaces para IPN. Essa funcionalidade é distribuída por todos os nós spine e interface para IPN em muitos grupos.
- Os IPNs recebem essas junções e enviam junções PIM em direção ao RP PIM bidirecional.
- Como PIM Bidir é usado, não há árvores (S,G). Somente (\*,G) árvores são usadas no PIM Bidir.

- Todo o tráfego do plano de dados enviado ao GIPO passa pelo RP.

## Plano de controle multicast IPN



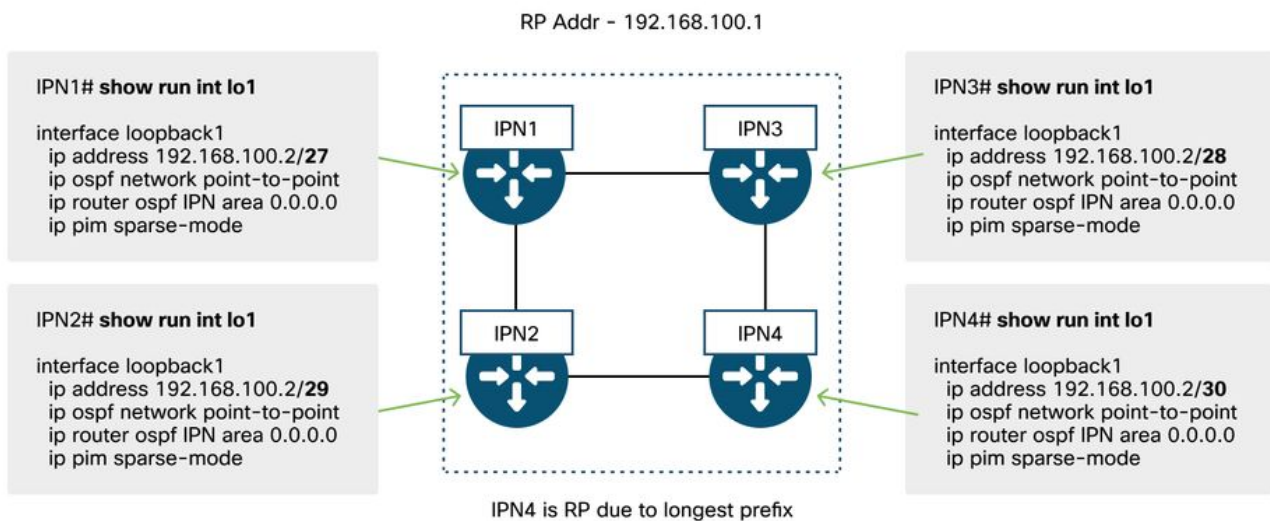
## Painel de dados multicast IPN



O único meio de redundância de RP com PIM Bidir é usar o Phantom. Isso é abordado em detalhes na parte do Multi-Pod Discovery deste livro. Como resumo rápido, observe que com o RP Fantasma:

- Todos os IPNs devem ser configurados com o mesmo endereço RP.
- O endereço RP exato não deve existir em nenhum dispositivo.
- Vários dispositivos anunciam a acessibilidade à sub-rede que contém o endereço IP do RP Fantasma. As sub-redes anunciadas devem variar no comprimento da sub-rede para que todos os roteadores concordem sobre quem está anunciando o melhor caminho para o RP. Se esse caminho for perdido, a convergência dependerá do IGP.

## Configuração do RP fantasma



## Fluxo de trabalho de solução de problemas de transmissão multipods, unicast desconhecido e multicast (BUM)

1. Primeiro, confirme se o fluxo está realmente sendo tratado como multidestino pela malha.

O fluxo será inundado no BD nestes exemplos comuns:

- O quadro é um broadcast ARP e a inundação ARP é ativada no BD.
- O quadro é destinado a um grupo multicast. Observe que mesmo se o rastreamento de IGMP estiver habilitado, o tráfego ainda será sempre inundado na estrutura do GIPo.
- O tráfego é destinado a um grupo multicast para o qual a ACI está fornecendo serviços de roteamento multicast.
- O fluxo é uma Camada 2 (fluxo interligado) e o endereço MAC destino é desconhecido e o comportamento unicast desconhecido no BD é definido como 'Flood'.

A maneira mais fácil de determinar qual decisão de encaminhamento será tomada é com um ELAM.

2. Identifique o BD GIPo.

Consulte a seção anterior deste capítulo que trata disso. Os ELAMs spine também podem ser executados através do aplicativo ELAM Assistant para verificar se o tráfego inundado está sendo recebido.

### 3. Verifique as tabelas de roteamento multicast no IPN para esse GIPO.

As saídas para fazer isso variam dependendo da plataforma IPN em uso, mas em um alto nível:

- Todos os roteadores IPN devem concordar com o RP e o RPF para esse GIPO deve apontar para essa árvore.
- Um roteador IPN conectado a cada Pod deve receber uma união IGMP para o grupo.

### Cenário #2 de Troubleshooting de Multipods (Fluxo BUM)

Este cenário cobriria qualquer cenário que envolva o ARP não sendo resolvido nos cenários de Multi-Pod ou BUM (unicast desconhecido, etc.).

Há várias causas possíveis comuns aqui.

#### Causa possível 1: Vários roteadores possuem o endereço PIM RP

Com esse cenário, a folha de entrada inunda o tráfego (verifique com ELAM), o Pod de origem recebe e inunda o tráfego, mas o Pod remoto não o recebe. Para alguns BDs, a inundação funciona, mas para outros não.

No IPN, execute 'show ip mroute <GIPO address>' para o GIPO para ver se a árvore RPF aponta para vários roteadores diferentes.

Se esse for o caso, verifique o seguinte:

- Verifique se o endereço RP PIM real não está configurado em nenhum lugar. Qualquer dispositivo que possua esse endereço RP real veria uma rota local /32 para ele.
- Verifique se vários roteadores IPN não estão anunciando o mesmo comprimento de prefixo para o RP no cenário RP Fantasma.

#### Causa possível 2: Os roteadores IPN não estão aprendendo rotas para o endereço RP

Assim como a primeira causa possível, aqui o tráfego inundado não consegue sair do IPN. A saída de 'show ip route <rp address>' em cada roteador IPN mostraria apenas o comprimento do prefixo configurado localmente em vez do que os outros roteadores estão anunciando.

O resultado disso é que cada dispositivo pensa que é o RP mesmo que o endereço IP RP real não esteja configurado em nenhum lugar.

Se for esse o caso, verifique o seguinte:

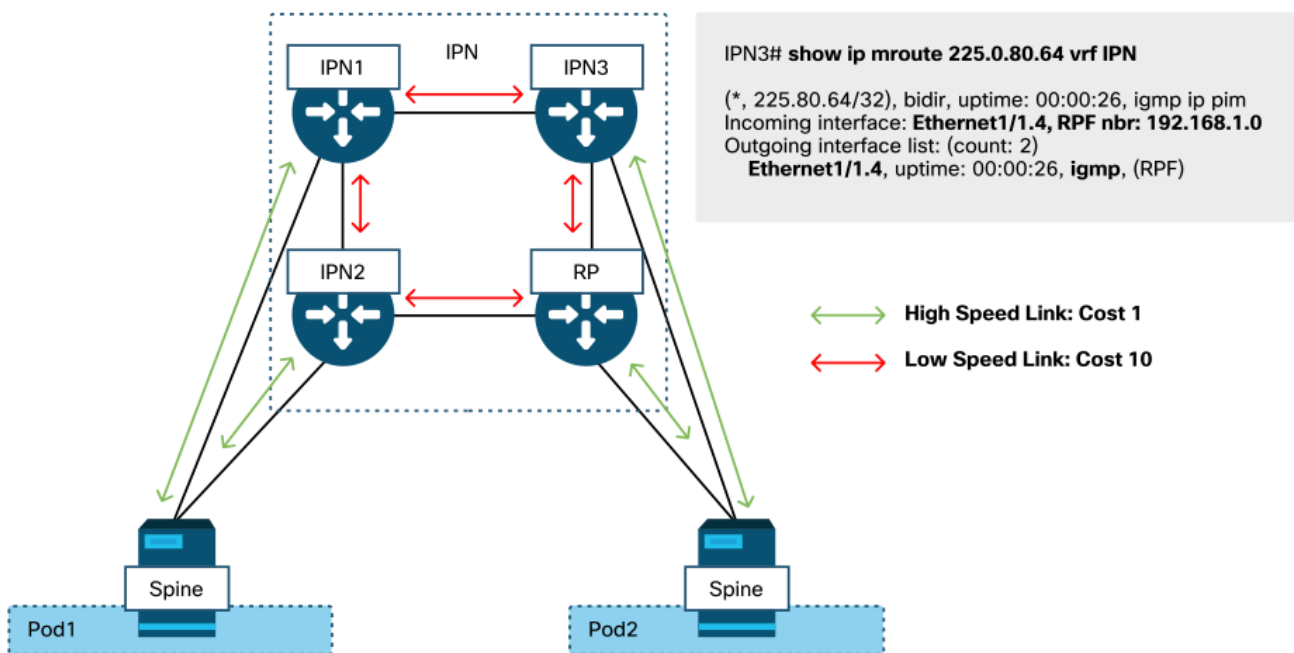
- Verifique se as adjacências de roteamento estão ativas entre os roteadores IPN. Verifique se a rota está no banco de dados do protocolo real (como o banco de dados OSPF).
- Verifique se todos os loopbacks que devem ser RPs candidatos estão configurados como tipos de rede ponto a ponto OSPF. Se esse tipo de rede não estiver configurado, cada

roteador sempre anunciará um tamanho de prefixo /32, independentemente do que estiver realmente configurado.

### Causa possível 3: Os roteadores IPN não estão instalando a rota GIPO ou os pontos RPF para a ACI

Como mencionado anteriormente, a ACI não executa o PIM em seus links para IPN. Isso significa que o melhor caminho do IPN em direção ao RP nunca deve apontar para a ACI. O cenário onde isso poderia acontecer seria se vários roteadores IPN fossem conectados ao mesmo spine e uma métrica OSPF melhor fosse vista através do spine do que diretamente entre os roteadores IPN.

#### Interface RPF para ACI



Para resolver esse problema:

- Certifique-se de que as adjacências do protocolo de roteamento entre os roteadores IPN estejam ativas.
- Aumente as métricas de custo do OSPF para os links para IPN nos nós spine para um valor que tornará essa métrica menos preferível do que os links IPN para IPN.

## Outras referências

Antes do software ACI 4.0, havia alguns desafios relacionados ao uso do COS 6 por dispositivos externos. A maioria desses problemas foram resolvidos através de aprimoramentos da versão 4.0, mas para obter mais informações, consulte a sessão do CiscoLive "BRKACI-2934 - Troubleshooting do Multi-Pod" e a seção "Qualidade de Serviço".

Sobre esta tradução

A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês ([link fornecido](#)) seja sempre consultado.