
cisco Live!
ALL IN

The bridge to possible

ACI Troubleshooting: A deep dive into PBR

Roland Ducomble – Technical Leader ACI TAC EMEAR – CCIE 3745

BRKDCN-3815

Cisco Webex App

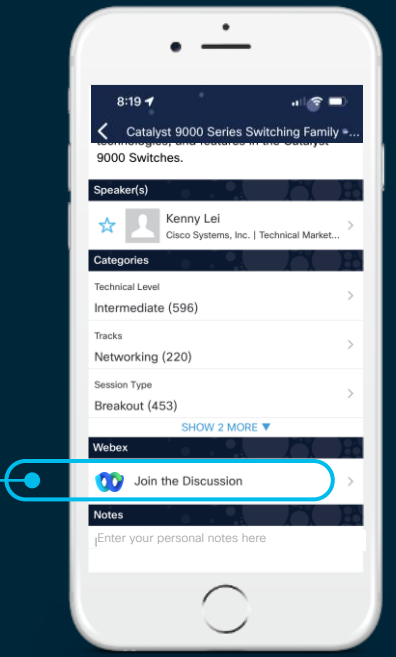
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 17, 2022.



https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-3815

Agenda

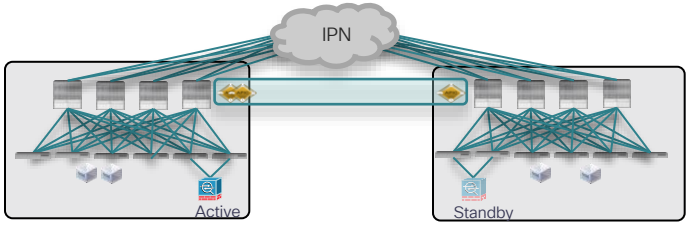
- Introduction
- PBR Firewall insertion in ACI Multipod
  - East-West
  - North-South
- PBR Firewall insertion in ACI Multisite
  - East-West
  - North-South (in Reference)
- Unidirection PBR – Load Balancer with no SNAT
- Multinode PBR – Firewall + Load Balancer

Reference Slide Icon ➔ 

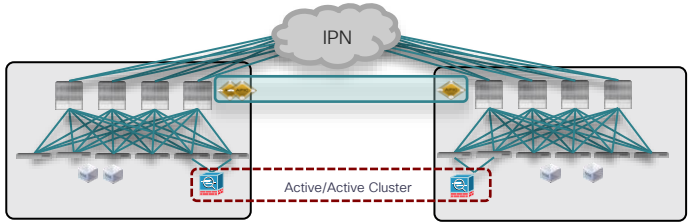
Acronyms/Definitions

Acronyms	Definitions	Acronyms	Definitions
EPG and EP	Endpoint Group and Endpoint	BD	Bridge Domain
FW	Firewall	Zoning-rule	Refer to a permit/deny/redirect rule between two pcTag on a leaf
LB	Load Balancer	Redir-info	Redirect info – refers to relevant info to apply redirect including VMAC to redirect, VIP and Service BD
PBR	Policy Based Redirect	SNAT	Source NAT
L3out	Layer 3 out		
North-South	Refer to traffic between EPG and L3out		
East-West	Refer to traffic between EPG or within EPG		
Ext EPG	External EPG aka EPG part of a L3out		
pcTag	Policy Tag		
sclass	Source class or pcTag of source		
dclass	Destination class of pcTag of destination		
VNID	VXLAN network identifier – refer to either a BD or a VRF in ACI		

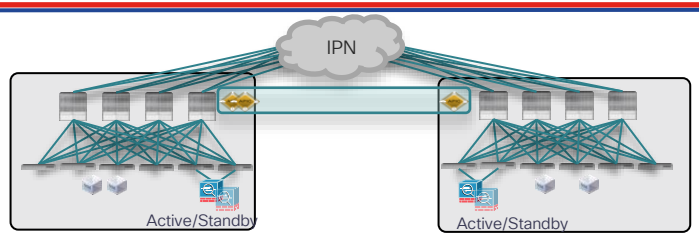
Multi-Pod and Network Services Integration Models



- Active and Standby pair deployed across Pods
- No issues with asymmetric flows
- All traffic redirected to same site – not ideal for active/active DC



- Active/Active FW cluster nodes stretched across Sites (single logical FW)
- Each device in the cluster share same VIP/VMAC
- Requires the ability of discovering the same MAC/IP info in separate sites at the same time
- Supported from ACI release 3.2(4d) with the use of Service-Graph with PBR (anycast PBR)



- Independent Active/Standby pairs deployed in separate Pods
- Each active/standby pair has a VIP and VMAC
- Use of Symmetric PBR to avoid the creation of asymmetric paths crossing different active FW nodes

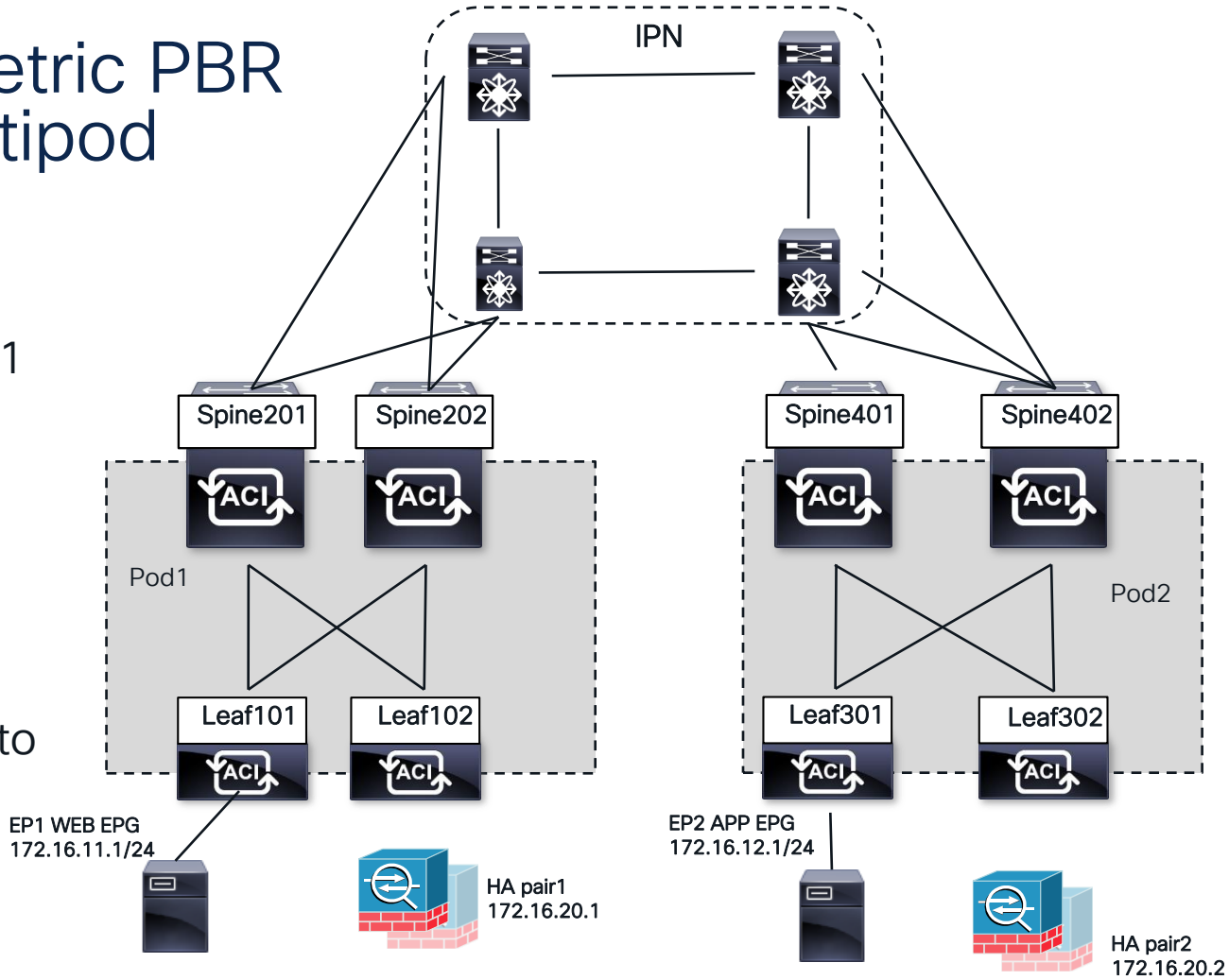
Multipod East-West Symmetric PBR

Setup – symmetric PBR East-West Multipod

Routed flow between 172.16.11.1 to 172.16.12.1
Default gateway for both is ACI 172.16.[1-2].254

Redirected to one the Firewall HA pair

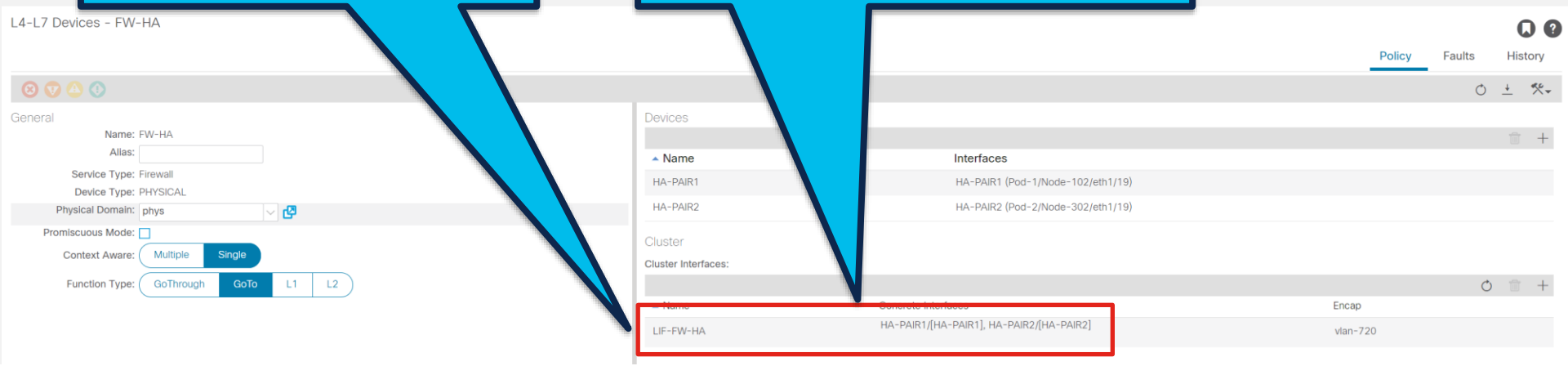
FW are one-arm attached to ACI



Config Gotcha – L4/L7 devices for Symmetric PBR)

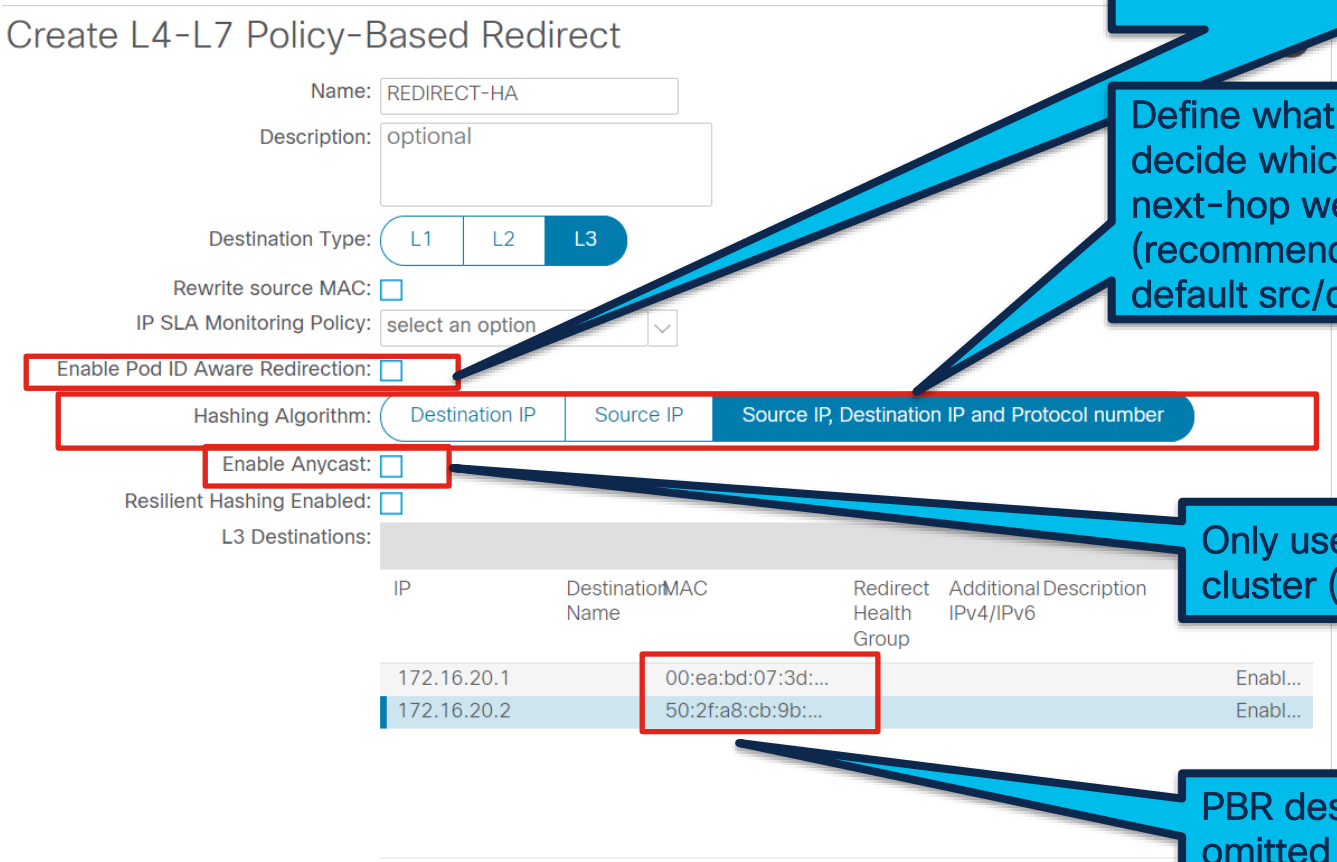
Cluster interface contains path to both HA Pair of firewall

Here we use one-arm
Hence only one Cluster Interface



Note a single L4/L7 cluster interface is representing both HA pair
Symmetric PBR will select HA Pair1 or HA Pair2 based on hashing

Config Gotcha - Redirect policy



Should only be considered in North-South PBR scenario

Define what we will hash to decide which of the PBR next-hop we will use (recommended to keep default src/dst/proto_

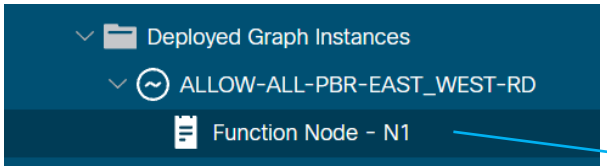
Only used for Active/Active cluster (Anycast VIP/VMAC)

PBR dest MAC can be omitted in 5.2 with PBR tracking

Check 1 – check the graph is deployed

Once Config is completed (Contract, Serv Graph Template, device selection policies.,)

Class id (pcTag) for the service EPG (“shadow” EPG). Created between Service node and ACI Leaf



If not deployed : usually it is Contract related (no consumer Or provider, ..)

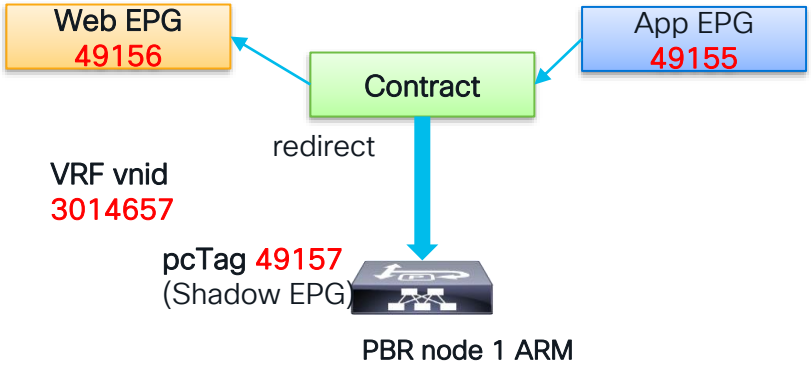
Check 2 – Service EPG and service VLAN is deployed on service leaf

```
S1P1-Leaf102# show vlan encap-id 720

 VLAN Name                              Status    Ports
 ---- --------------------------------- --------- -----------
 15   RD-MPOD:FW-HActxRD:LIF-FW-HA:     active    Eth1/19
..

S1P1-Leaf102# show system internal epm vlan 15  det

VLAN 15
VLAN type : FD vlan
hw id : 32 ::: sclass : 49157
access enc : (802.1Q, 720)
fabric enc : (VXLAN, 8912)
Object store EP db version : 4
BD vlan id : 14 ::: BD vnid : 14843887 ::: VRF vnid :
3014657
Valid : Yes ::: Incomplete : No  ::: Learn Enable : Yes
pol_ctrl_flags:  ::: dom ctrl : ep-service-enabled |
Endpoint count : 1 ::: Local Endpoint count : 1 On Peer
Endpoint count 0
```

- FW cluster interface is using encap vlan-720.
- This is deployed on service leaf and is using service EPG pcTag (sclass 49157).
- It is also mark as service epg

Check 3 - Zoning-rule

- Make note all all vnid and sclass involved



Application Profile Name	EPG Name	Class ID	Scope
App	App	49156	3014657
App	Web	49155	3014657

- Check expected zoning-rule

1. Cons to Prov : 49156 to 49155 : REDIRECT
2. Shadow to Prov : 49157 to 49155 : PERMIT
3. Prov to Cons : 49155 to 49156 : REDIRECT
4. Shadow to Cons : 49157 to 49156 : PERMIT

Note it may be all rules are not on the same leaf

```
S1P1-Leaf101# show zoning-rule scope 3014657
+---------+--------+--------+----------+---------+---------+----------------+---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Action    | Priority      |
+---------+--------+--------+----------+---------+---------+----------------+---------------+
|  4128   | 49157  | 49156  |    11    | enabled | 3014657 |      permit    | src_dst_any(9)|
|  4191   | 49157  | 49155  |    11    | enabled | 3014657 |      permit    | src_dst_any(9)|
|  4190   | 49155  | 49156  |    11    | enabled | 3014657 |redir(destgrp-1)| src_dst_any(9)|
|  4189   | 49156  | 49155  |    11    | enabled | 3014657 |redir(destgrp-1)| src_dst_any(9)|
+---------+--------+--------+----------+---------+---------+----------------+---------------+
```

Check 3 - Redirect info

Redir group should have the VIP/VMAC of each HA pair

Vxlan VNID and vMac will be used for COOP MAC lookup on spine

```
S1P1-Leaf102# show service redir info group 1
================================================================================
GrpID Name            destination                    operSt

1     destgrp-1       dest-[172.16.20.2]-[vxlan-3014657]   enabled       per-grp   sym
                      dest-[172.16.20.1]-[vxlan-3014657]


S1P1-Leaf102# show service redir info destination ip 172.16.20.2 vnid 3014657
================================================================================
Name                                    bdVnid          vMac               vrf
====                                    ======          ====               ====
dest-[172.16.20.2]-[vxlan-3014657]      vxlan-14843887  50:2F:A8:CB:9B:3C  RD-MPOD:RD

S1P1-Leaf102# show service redir info destination ip 172.16.20.1 vnid 3014657
Name                                    bdVnid          vMac               vrf
====                                    ======          ====               ====
dest-[172.16.20.1]-[vxlan-3014657]      vxlan-14843887  00:EA:BD:07:3D:7C  RD-MPOD:RD
```

Check 4 – Check load balancing
for a given flow (in vsh_lc mode)

- 172.16.11.1 to 172.16.12.1 ICMP (proto 0x1) will go to 172.16.20.1 with mac 00:ea:bd:07:3d:7c (FW pair in Pod1)

```
module-1# show platform internal hal policy redirdst  group_id 1 ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
protocol 0x6
Group Id                        : 0x1
Src IP                          : 172.16.11.1/32
Dst IP                          : 172.16.12.1/32
Protocol                        : 0x6
Rewrite MAC                     : 50:2f:a8:cb:9b:3c
Rewrite VNID                    : 0xe27fef
Redirect Dst's IP               : 172.16.20.2/32
Redirect Dst's vrf              : 0x2e0001
```

- 172.16.11.1 to 172.16.12.1 TCP (proto 0x6) will go to 172.16.20.2 with mac 50:2f:a8:cb:9b:3c (FW pair in Pod2)

```
module-1# show platform internal hal policy redirdst  group_id 1 ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
protocol 0x1
Group Id                        : 0x1
Src IP                          : 172.16.11.1/32
Dst IP                          : 172.16.12.1/32
Protocol                        : 0x1
Rewrite MAC                     : 00:ea:bd:07:3d:7c
Rewrite VNID                    : 0xe27fef
Redirect Dst's IP               : 172.16.20.1/32
Redirect Dst's vrf              : 0x2e0001
```

Packet – symmetric PBR

Packet path Consumer Web to Provider APP Epg

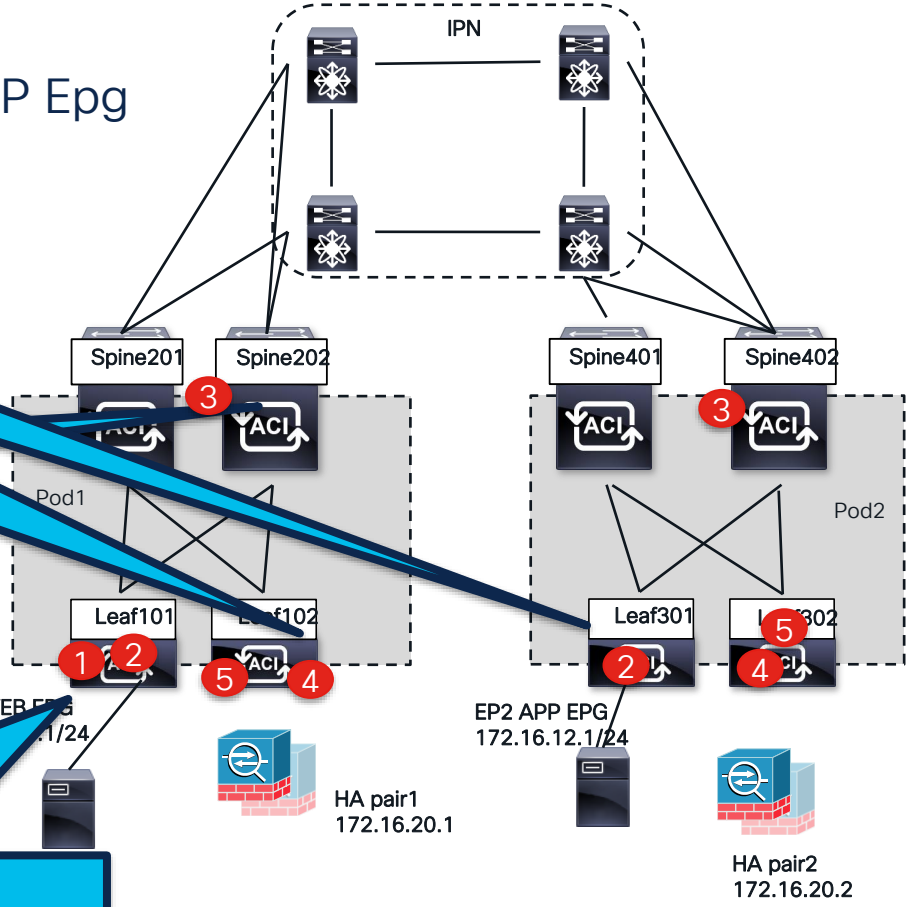
5 Returning from firewall on service leaf
permit rule to egress leaf 301

4 On service leaf, it is a pure Layer 2 packet to the firewall

3 Spine : COOP lookup in BD VNID for Redirect
and will send it toward service leaf (here 102)

2 Leaf doing redirect (101 or 301)
DMAC is rewritten to Firewall Pair1 or 2 . Say we hash to Pair 1 here
No Mac lookup happening on leaf.
Packet is encapsulated to Service BD VNID and send to vxlan tunnel to anycast-mac on spine.

1 Ingress leaf
if EP is known → redirect
if EP is unknown redirect will happen on egress leaf (301)



CLI check ingress leaf 1

```
S1P1-Leaf101# show system internal epm endpoint ip 172.16.11.1
MAC : 0050.568f.96b7 ::: Num IPs : 1
IP# 0 : 172.16.11.1 ::: IP# 0 flags :  ::: l3-sw-hit: No
Interface : Ethernet1/11
Flags : 0x80004c04 ::: sclass : 49155 ::: Ref count : 5

S1P1-Leaf101# show system internal epm endpoint ip 172.16.12.1
MAC : 0000.0000.0000 ::: Num IPs : 1
IP# 0 : 172.16.12.1 ::: IP# 0 flags :  ::: l3-sw-hit: No
Interface : Tunnel16
Flags : 0x80004400 ::: sclass : 49156 ::: Ref count : 3

S1P1-Leaf101# show zoning-rule scope 3014657 src-epg 49155 dst-epg 49156
+---------+--------+--------+----------+--------+---------+---------+------+-----
| Rule ID | SrcEPG | DstEPG | FilterID |  Dir   |  operSt |  Scope  | Name |      Action      |    Priority    |
+---------+--------+--------+----------+--------+---------+---------+------+------------------+----------------+
|   4128  | 49155  | 49156  | default  | bi-dir | enabled | 3014657 |      | redir(destgrp-1) | src_dst_any(9) |
+---------+--------+--------+----------+--------+---------+---------+------+------------------+----------------+
S1P1-Leaf101# show service redir info group 1
========================================================================================================================
GrpID Name      destination                          operSt     operStQual
===== ====      ===========                          =======    ==========
destgrp-1       dest-[172.16.20.2]-[vxlan-3014657]      enabled    no-oper-grp      sym
                dest-[172.16.20.1]-[vxlan-3014657]
```

Local EP is known
In sclass 49155

Destination EP is known in sclass

Zoning-rule from src epg to dst epg
Gives redirect group 1

Redirect group 1 is symmetric PBR
To two service IP/MAC

Spine COOP DB if hashing gives FW MAC local to Pod1

```
S1P1-Spine201# show coop internal info  repo ep key 14843887 00:EA:BD:07:3D:7C
Repo Hdr Checksum : 46240
Repo Hdr record timestamp : 10 12 2021 14:37:13 505028097
Repo Hdr last pub timestamp : 10 12 2021 14:37:13 507060173
Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Repo Hdr dampen penalty : 0
Repo Hdr flags : IN_OBJ EXPORT ACTIVE
EP bd vnid : 14843887
EP mac :  00:EA:BD:07:3D:7C
flags : 0x80
repo flags : 0x122
Vrf vnid : 3014657
PcTag : 0x1008004
EVPN Seq no : 0
Remote publish timestamp: 01 01 1970 00:00:00 0
Snapshot timestamp: 10 12 2021 14:37:13 505028097
Tunnel nh : 10.0.0.67
MAC Tunnel  : 10.0.0.67
TX Status: COOP_TX_DONE
Damp penalty: 30
Damp status: NORMAL
Leaf 0 Info :
IPv4 Repo Hdr flags : IN_OBJ EXPORT
Real IPv4 EP : 172.16.20.1
```

VMAC in service BD VNID

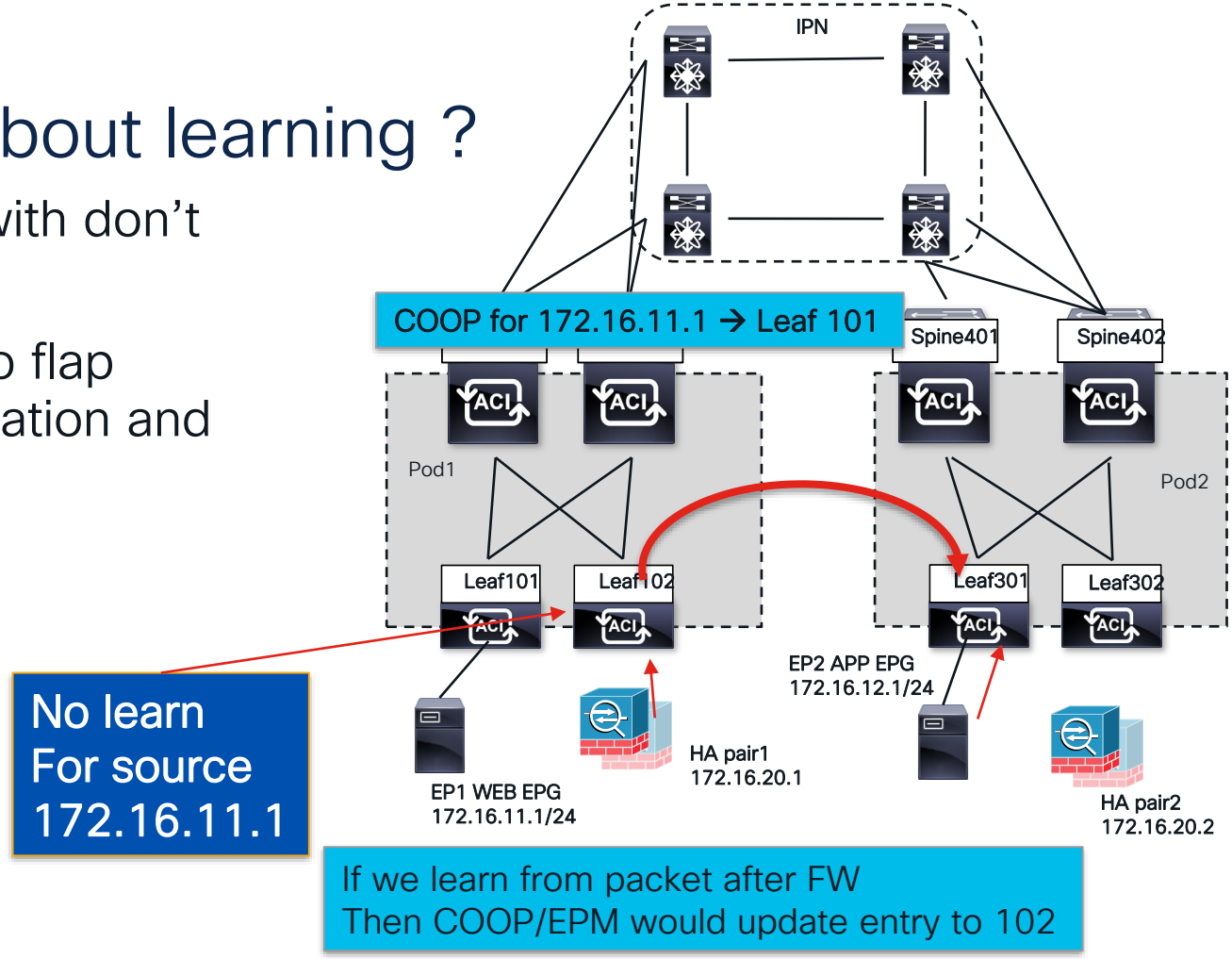
Healthy

BD Name	BD Alias	Class ID	Segment ID
BD1		49153	14909416
BD2		49154	15105997
Service-BD		32771	14843887

```
S1P1-Spine201# acidiag fnvread | egrep 10.0.0.67
      102       1        S1P1-Leaf102      FDO223007G7       10.0.0.67/32     leaf         active   0
```

PBR – What about learning ?

- Service EPG are set with don't learn from dataplane.
- Needed to avoid EP to flap between their real location and Firewall



Datapath Troubleshooting Tool:
ftriage from APIC CLI (Here 5.2(3))

Before service device

```
bdsol-aci37-apic1# ftriage route -ii LEAF:101 -sip 172.16.11.2 -dip 172.16.12.2
2021-10-27 08:28:41,179 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf101 Ingress: Eth1/11 Egress: Eth1/49  Vnid: 14909416
2021-10-27 08:29:27,042 INFO     ftriage:  unicast:1543 S1P1-Leaf101: traffic is redirected to vnid:14843887 mac:00:EA:BD:07:3D:7C via tenant:RD-
    MPOD graph:EAST_WEST contract: ALLOW-ALL-PBR
2021-10-27 08:30:18,974 INFO     ftriage:     main:1333 S1P1-Spine201: Incoming Packet captured with Outer [SIP:10.0.0.67, DIP:10.0.72.65] ....
    Inner [SIP:172.16.11.2, DIP:172.16.12.2]
2021-10-27 08:31:28,056 INFO     ftriage:  unicast:2196 S1P1-Spine201: EP is known in COOP (DIPo = 10.0.0.67)
2021-10-27 08:31:41,494 INFO     ftriage:     main:958  Found peer-node S1P1-Leaf102 and IF: Eth1/49 in candidate list
2021-10-27 08:31:51,918 INFO     ftriage:       ep:128  S1P1-Leaf102: pbr traffic with dmac: 00:EA:BD:07:3D:7C
2021-10-27 08:32:06,748 INFO     ftriage:     main:1796 Packet is Exiting fabric with peer-device: POD1-router1 and peer-port: Ethernet1/19
2021-10-27 08:32:06,753 INFO     ftriage: acigraph:646   found matching devicenode:N1 ldev:FW-HA dev:HA-PAIR1HA-PAIR1uni/tn-RD-MPOD/lDevVip-FW-
    HA/cDev-HA-PAIR1/cIf-[HA-PAIR1]
2021-10-27 08:32:06,754 INFO     ftriage:  unicast:2739 S1P1-Leaf102:  PBR first pass is done and trafic is sent to service device: node:N1
    ldev:FW-HA dev:HA-PAIR1
2021-10-27 08:32:06,754 INFO     ftriage:  unicast:2741 S1P1-Leaf102:  expected traffic to return from: topology/pod-1/paths-102/pathep-[eth1/19]
    encap:720
```

After service device

```
2021-10-27 08:32:21,224 INFO     ftriage:     main:1821 pbr return path, nxt_nifs {S1P1-Leaf102: ['Eth1/19']}, nxt_dbg_f_n ig, nxt_inst ig, eg_ifs
    Eth1/19, Vnid: 720
2021-10-27 08:32:33,581 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf102 Ingress: Eth1/19 Egress: Eth1/49  Vnid: 3014657
2021-10-27 08:33:14,060 INFO     ftriage:     main:958  Found peer-node S1P1-Spine201 and IF: Eth1/2 in candidate list
```

Multipod North-South symmetric PBR and optimization

North-South PBR in multipod – Optimization

How to avoid hair pinning Across IPN ?

Host based routing (HBR) (4.0 and plus)
Location Aware PBR (3.1 and plus)

- If we have multiple PBR service nodes, it's load-balanced based on Source IP, Destination IP and Protocol Type by default. Hash tuple is configurable, but we don't have capability to select local PBR service node. In 3.1, we have option to prefer local pod PBR node (multipod fabric only)
- It is recommended (not mandatory) that Location aware PBR be used for North-South firewall integration with GOLF/HBR host route advertisement.
- **Location aware PBR CANNOT be used for EAST-WEST** traffic or it will lead to asymmetric flow (using different FW pair in both direction)
- It can't be used neither for L3out to L3out.

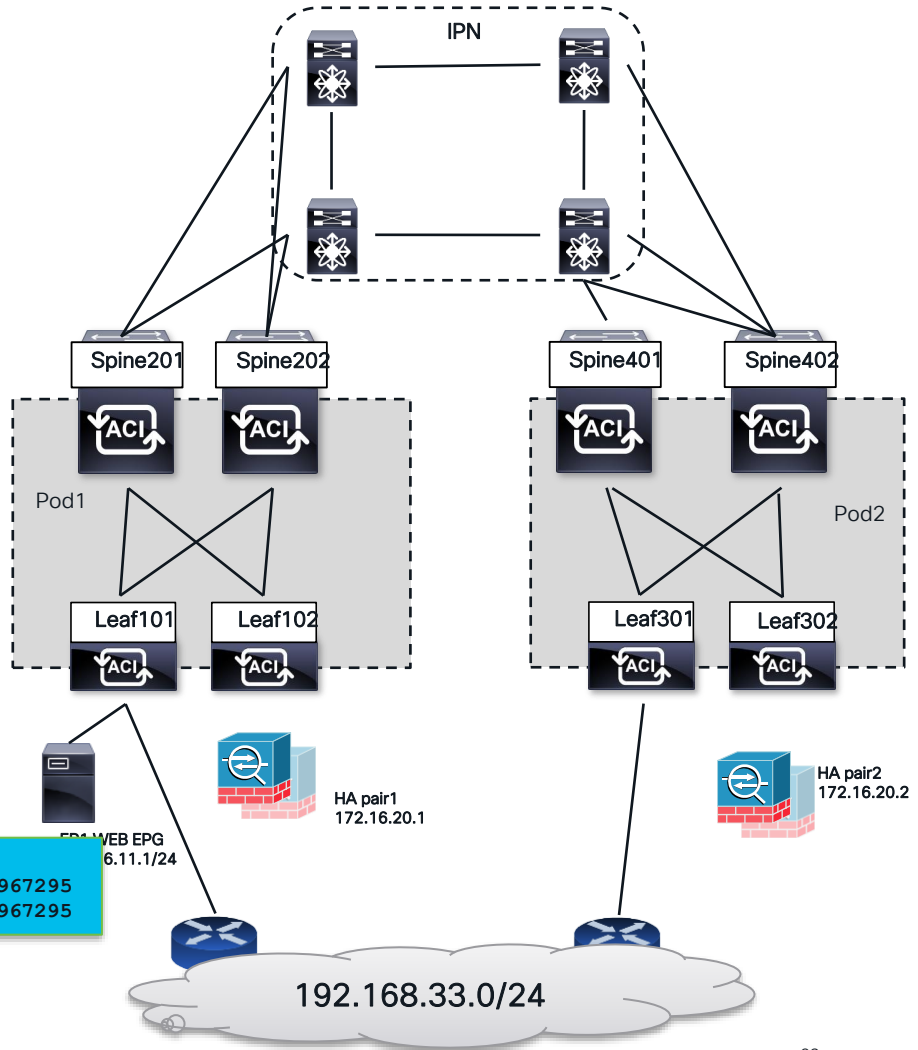
Symmetric PBR
North-South PBR

Between internal endpoint and Layer 3 out.
Similar symmetric PBR as for East-West works
However there may be a lot of hairpinning across IPN

- Hashing across Pod
- L3 out entering other Pod than EP

Routing table in external router will get route for BD subnet either through leaf 101 or leaf 103 (or ECMP) depending where you are in the L3 out network

```
172.16.11.0/24, ubest/mbest: 2/0
    *via 192.168.1.1, Vlan920, [110/20], 00:01:44, ospf-1, type-2, tag 4294967295
    *via 192.168.1.3, Vlan920, [110/20], 00:01:44, ospf-1, type-2, tag 4294967295
```

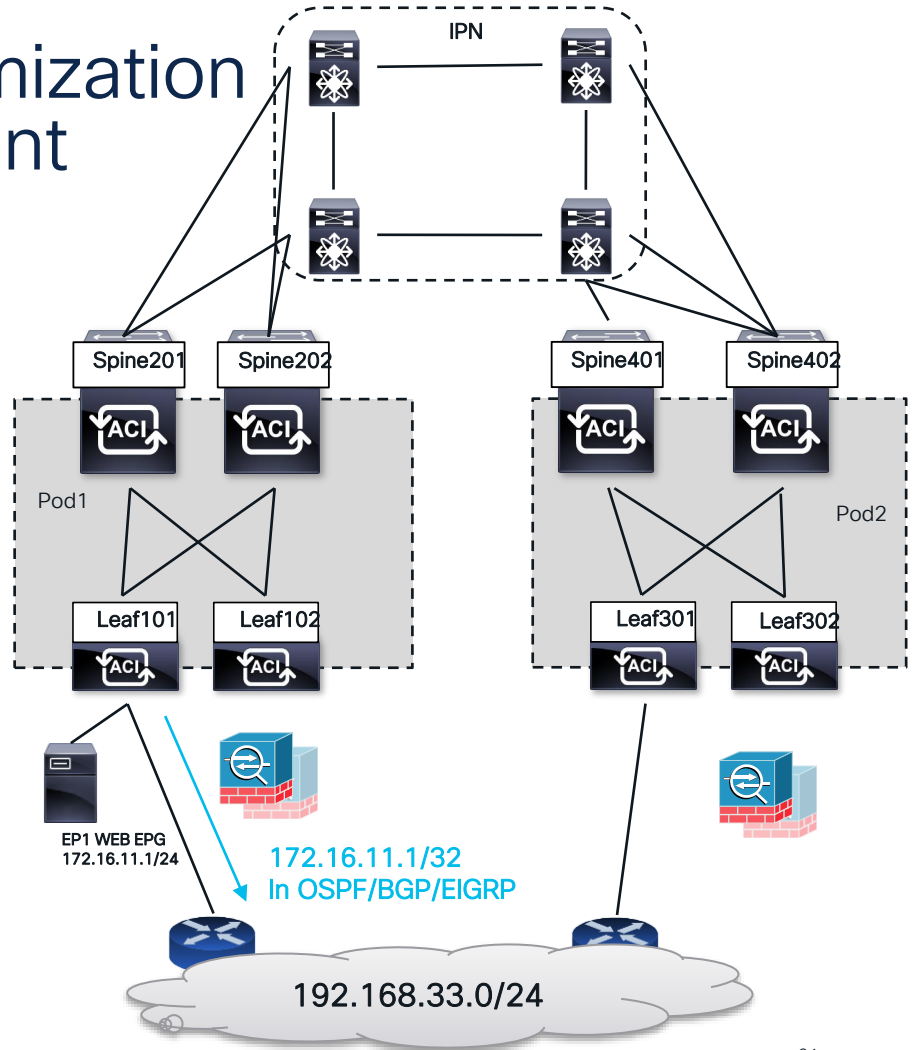


Multipod N-S PBR - Optimization 1. Host route advertisement

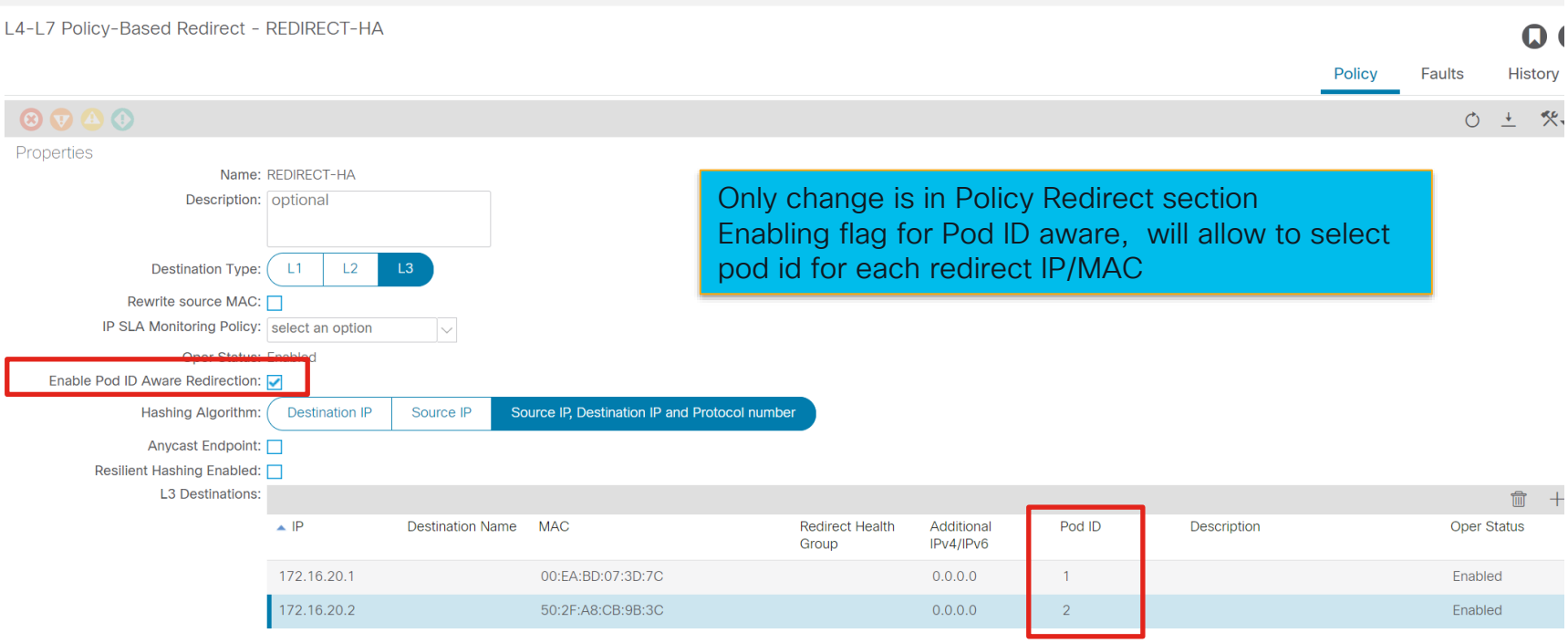
Starting 4.x you can enable on regular L3 out a BD to advertise /32 host route for each of the EP. Only Pod local Border leaf will advertise the EP.

```
172.16.11.0/24, ubest/mbest: 2/0
    *via 192.168.1.1, Vlan920, [110/20], 00:01:44, ospf-1, type-2
    *via 192.168.1.3, Vlan920, [110/20], 00:01:44, ospf-1, type-2
172.16.11.1/32, ubest/mbest: 1/0
    *via 192.168.1.1, Vlan920, [110/1], 01:43:18, ospf-1, type-2
```

Ensure L3 out to Endpoint path enters ACI in the target Pod



Multipod N-S PBR – Optimization
2. Enable location aware PBR



Only change is in Policy Redirect section
Enabling flag for Pod ID aware, will allow to select pod id for each redirect IP/MAC

Note that there are no visible changes in service redir info . However hardware HAL shows the changes

Changes in hardware (leaf pod1 shown)

Before enabling Pod aware
On leaf 101 we see both redirect
(group id comes from zoning-rule)

```
module-1# show platform internal hal objects policy dstgrp group_id 1
## Get Objects for policy dstgrp for Asic 0

  OBJECT 0:
Handle                                                  : 81469
group_id                                                : 0x1
hash_prof                                               : symmetric
resilienthash                                           : Disabled
sortbyname                                              : Disabled
up                                                      : Enabled
backuponly                                              : Disabled
backup_group_id                                         : 0x0
svctotaldests                                           : 0x2
dstips                                                  :
 Element 0 : 172.16.20.1/32
 Element 1 : 172.16.20.2/32
dstindices                                              :
  Element 0 : 0
  Element 1 : 1
destsbehindl3out                                        : Disabled
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle               : 81497
  rel-dstgrptodst-policy-redir_dst-group_id             : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                   : 172.16.20.1/32
  rel-dstgrptodst-policy-redir_dst-vrf                  : 0x2e0001
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle               : 100480
  rel-dstgrptodst-policy-redir_dst-group_id             : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                   : 172.16.20.2/32
  rel-dstgrptodst-policy-redir_dst-vrf                  : 0x2e0001
```

After enabling Pod aware
On leaf 101 we only see local 172.16.20.1
In the hash list

```
module-1# show platform internal hal objects policy dstgrp group_id 1
## Get Objects for policy dstgrp for Asic 0

  OBJECT 0:
Handle                                                  : 81469
group_id                                                : 0x1
hash_prof                                               : symmetric
resilienthash                                           : Disabled
sortbyname                                              : Disabled
up                                                      : Enabled
backuponly                                              : Disabled
backup_group_id                                         : 0x0
svctotaldests                                           : 0x2
dstips                                                  :
 Element 0 : 172.16.20.1/32
 Element 1 : 172.16.20.2/32
dstindices                                              :
  Element 0 : 0
  Element 1 : 1
destsbehindl3out                                        : Disabled
Relation Object dstgrptodst :
  rel-dstgrptodst-policy-redir_dst-handle               : 81497
  rel-dstgrptodst-policy-redir_dst-group_id             : 0x1
  rel-dstgrptodst-policy-redir_dst-ip                   : 172.16.20.1/32
  rel-dstgrptodst-policy-redir_dst-vrf                  : 0x2e0001
```

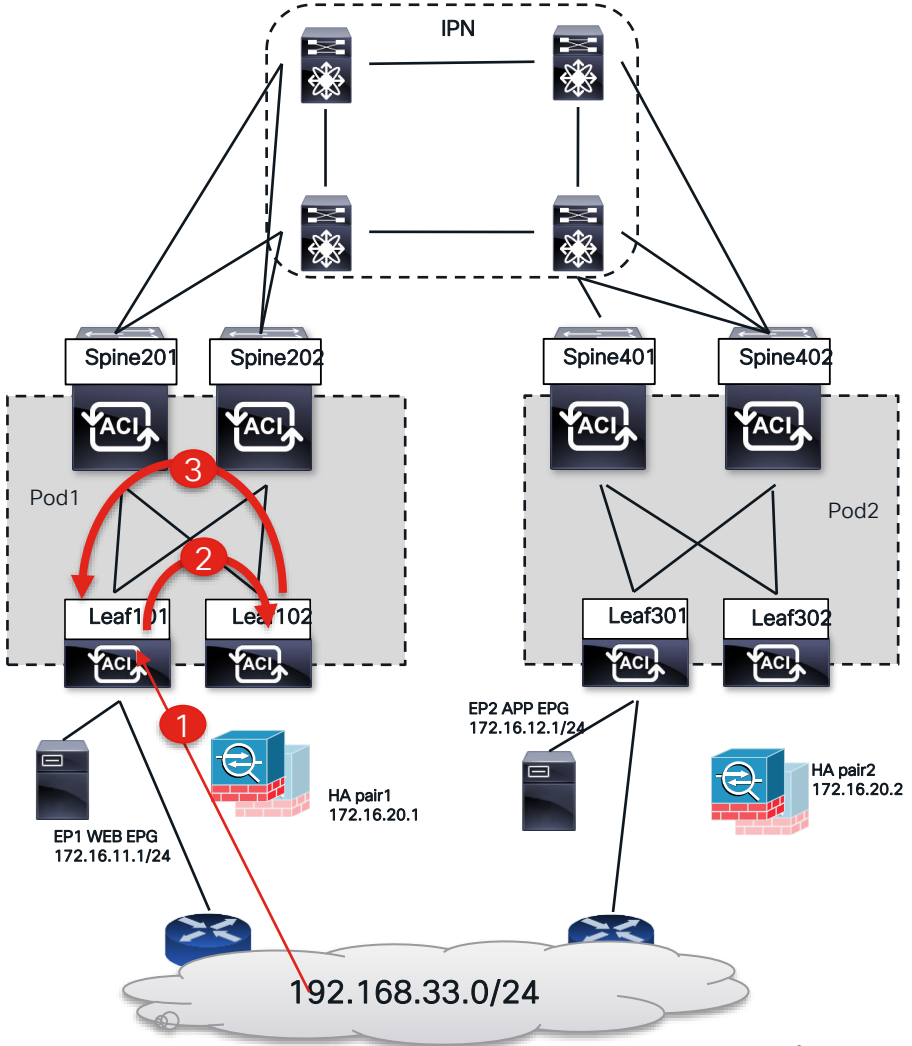
North-South PBR

Packet path from External to 172.16.11.1

1. From L3 network, Entering ACI in Pod1 BL Always (Host based routing).
   In ACI, for packet coming in from L3 out, policy is always applied in server leaf (here it is the same aka leaf 101)

2. Leaf 101 will Redirect and it WILL ALWAYS BE to HA pair1 (per pod aware feature)
No IPN crossing

3. Back from HA pair 1 it goes to EP1 on leaf 101 (permit rule)

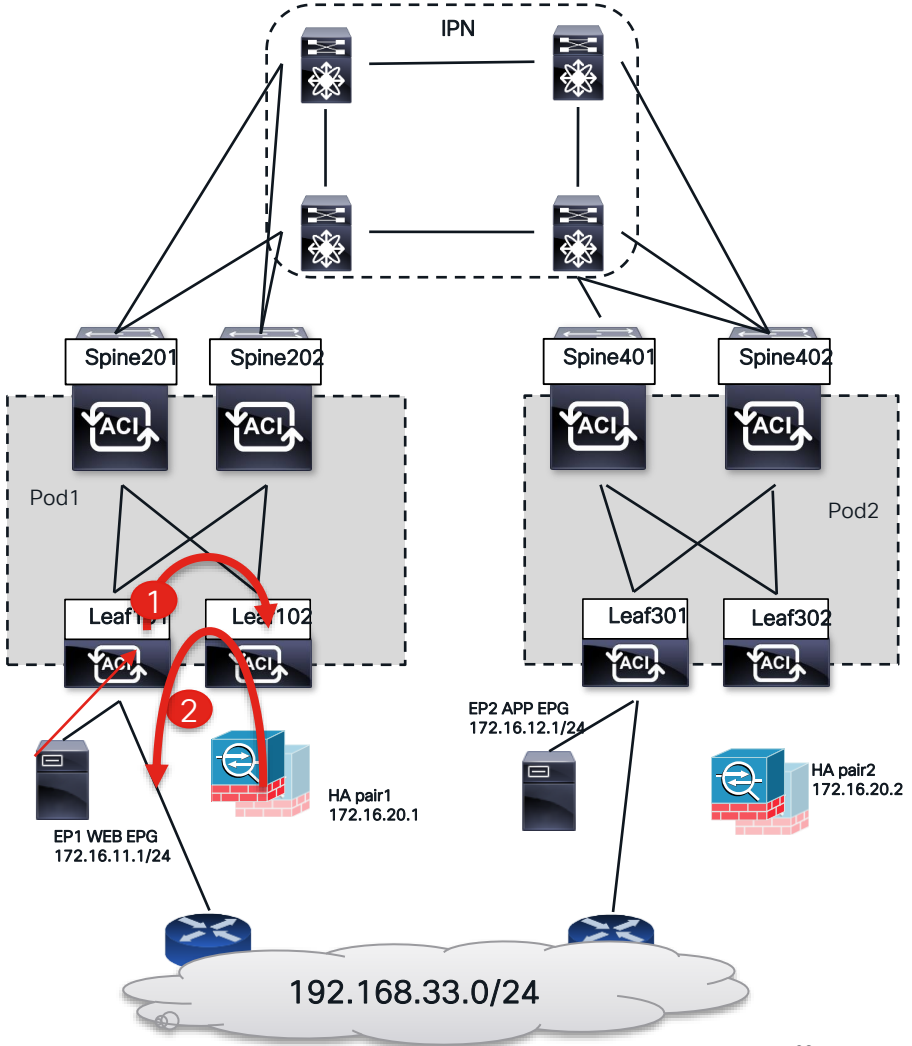


North-South PBR

Packet path from EP 172.16.11.1 to L3 out

Return from EP

1. Ingress Server leaf (101 here) Redirect to Local HA Pair (pod aware)
2. Back from Firewall we will use regular routing.
   - Most likely will exit through l3 out in pod1 (unless routing metric to destination mandate it differently)



Multisite PBR
ACI code post 4.0

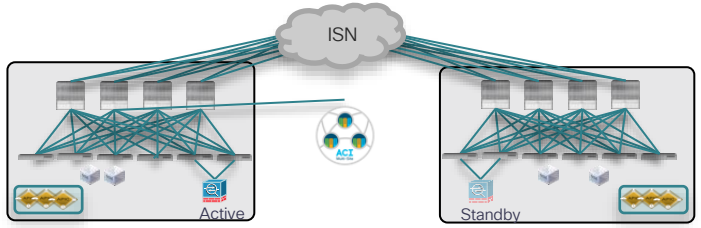
Challenges with Multisite PBR

- We need to ensure traffic symmetry across site
- **APIC cluster do not have visibility on remote site PBR node and can only redirect to local site L4/L7 device**
  - How can we ensure redirect is symmetric (same site in both direction)

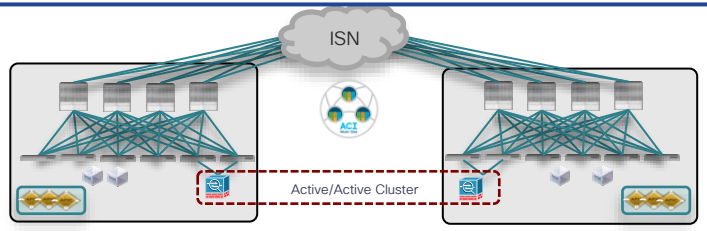
Rule 1 : Firewall likes symmetry

- Implementation is the following (post 4.x)
  - **East-West – Redirect ALWAYS apply in the site where Provider EP sits.**
    - **Extra requirement – Consumer EPG should have subnet under them**
  - **North-South – Redirect is always apply on Server leaf site** (non BL)

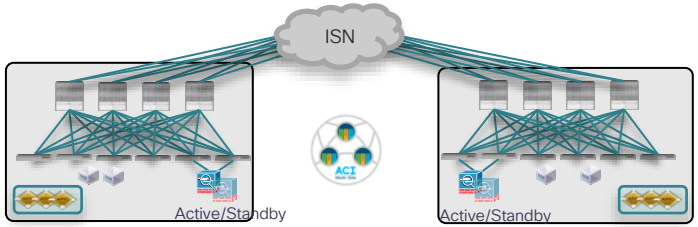
ACI Multi-Site and Network Services
Integration Models



- Active and Standby pair deployed across Pods
- Currently supported only
  - if the ACI is in L2 mode and FW is acting as default gateway for the endpoints
  - Of if FW is behind L3 out
- Aka not supported with PBR



- Active/Active FW cluster nodes stretched across Sites (single logical FW)
- Requires the ability of discovering the same MAC/IP info in separate sites at the same time
- Not supported

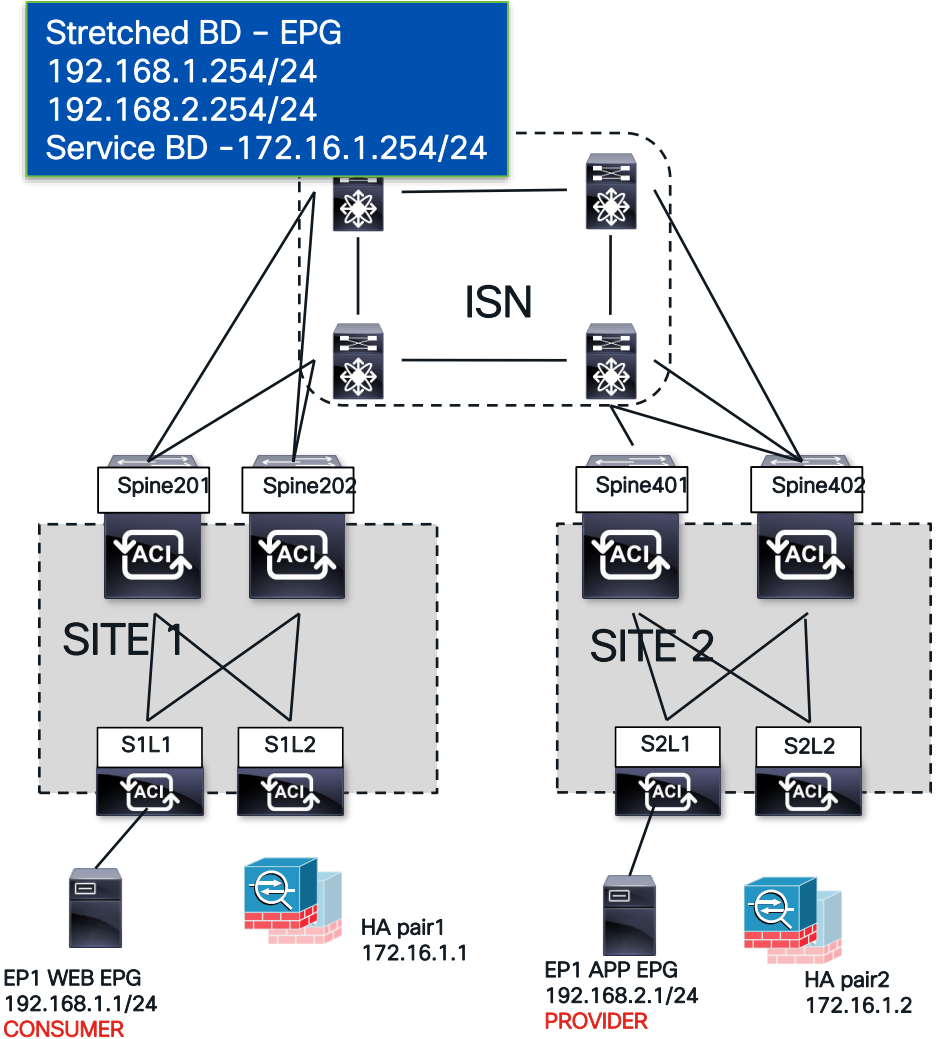


- Recommended deployment model for ACI Multi-Site
- supported from 3.2 release with the use of Service Graph with Policy Based Redirection (PBR).
- Recommended 4.x or later

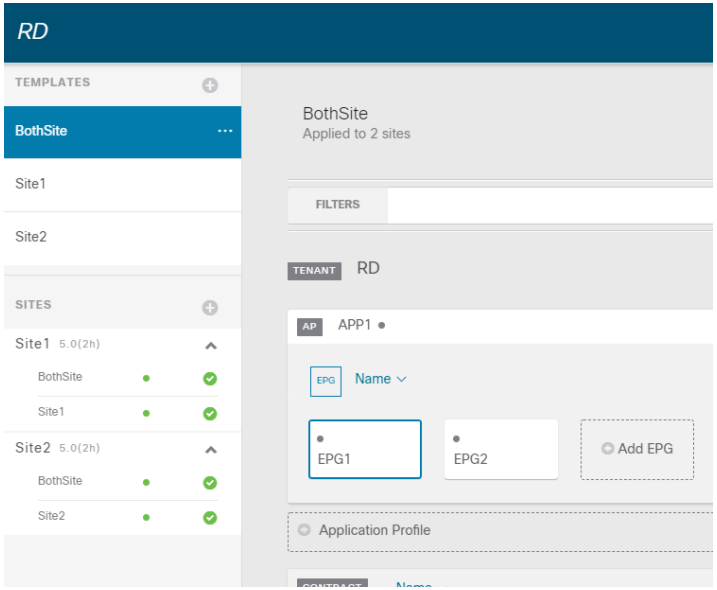
Multisite East-West PBR

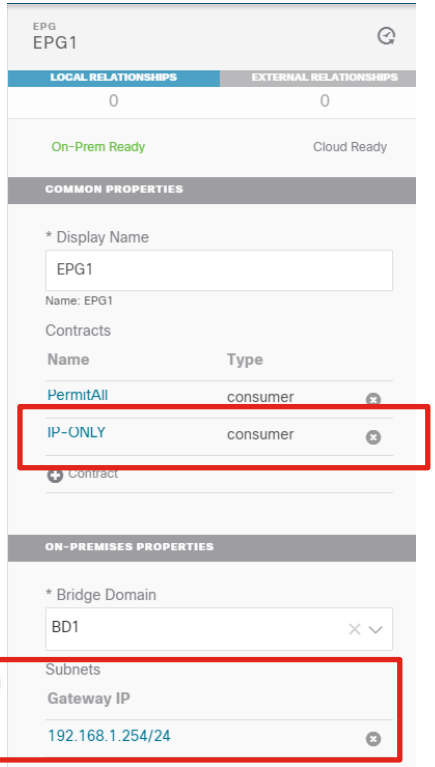
Multisite East-West PBR

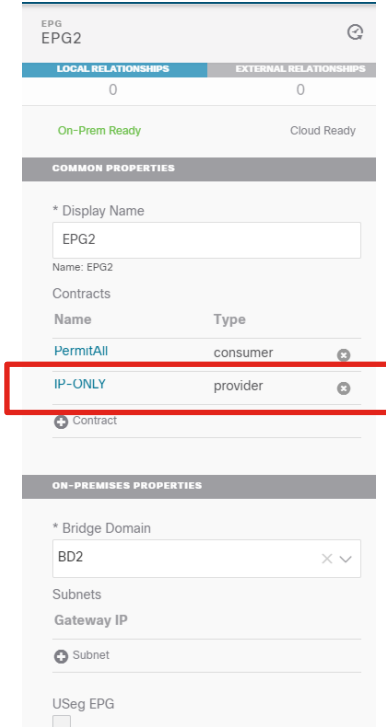
- Multisite PBR requirement
  - A site can only redirect to site local PBR Devices
- Rule 1 : we need to go through same Firewall pair in both direction
- Per Implementation :
  - Redirect happens in the site where provider endpoint is
- Extra requirement Consumer subnet must be configured under the consumer EPG



Multisite East-West PBR – Config gotcha







EPG1 is the consumer of the contract and Subnet is under EPG
EPG2 is provider of the contract and subnet do not need to be under the EPG

Multisite – East-West PBR Consumer to Provider

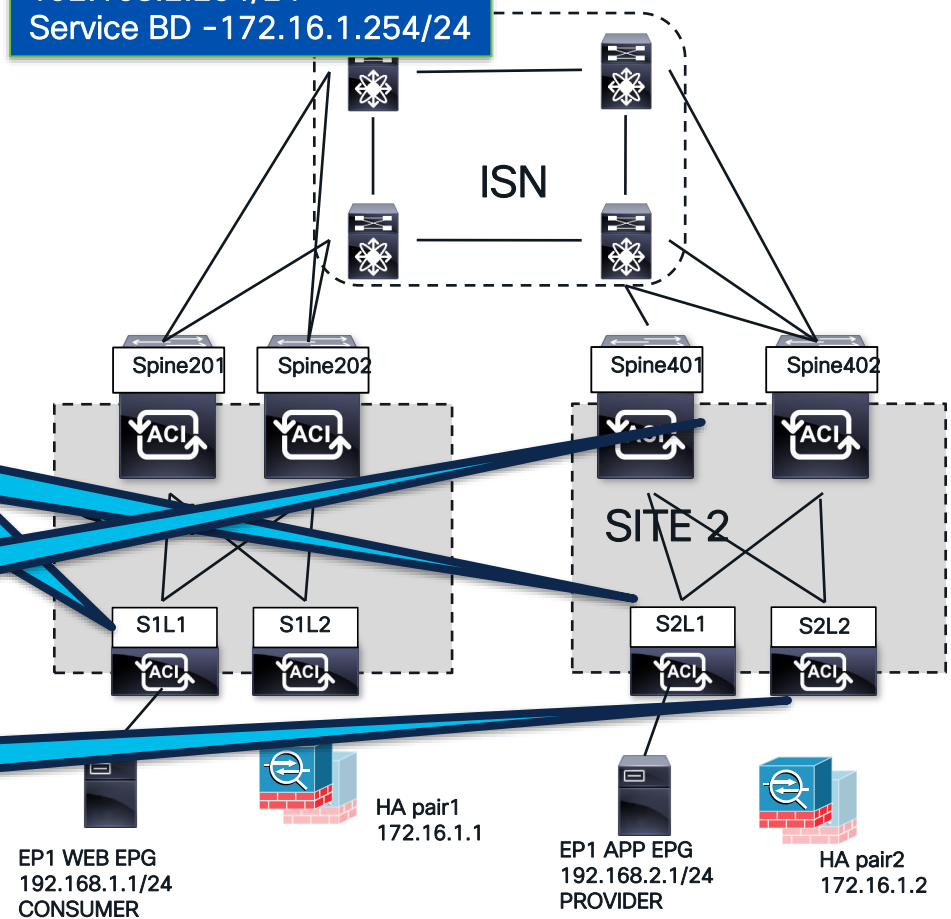
Stretched BD – EPG
192.168.1.254/24
192.168.2.254/24
Service BD -172.16.1.254/24

Ingress leaf (S1L1) will
- Apply redirect only if Provider EP is local (not the case here)
- Will use override rule (permit – no policy applied) in all other case → going to Provider leaf in site 2 per regular multisite dataplane

Egress provider leaf will apply the redirect rule and will redirect to firewall HA Pair2 in Site2

Spine site 2 : COOP lookup for HA pair2 VMAC in service BD → Service leaf site 2 (S2L1) to the firewall

After firewall , permit rule to reach provider leaf and provider Endpoint



Consumer to Provider
Ingress Consumer leaf zoning-rule – site 1

Unless the destination EP is local redir_override rule will be used(bypass PBR and do not mark policy)

```
S1P1-Leaf101# show zoning-rule scope 2719744 src-epg 32772 dst-epg 32771
+---------+------+------+----------+---------+---------+---------+-----------------------------------+------------+
| Rule ID |SrcEPG|DstEPG| FilterID |  operSt |  Scope  |       Action                        | Priority   |
+---------+------+------+----------+---------+---------+---------+-----------------------------------+------------+
|   4120  |32772 |32771 |    10    | enabled | 2719744 |redir(destgrp-1),redir_override|fully_qual(7)|
+---------+------+------+----------+---------+---------+---------+-----------------------------------+------------+


S1P1-Leaf101# show service redir info
==========================================================================================
List of Dest Groups
GrpID Name            destination                      HG-name          BAC  operSt
===== ====            ===========                      ===============  ===  =======
1     destgrp-1       dest-[172.16.1.1]-[vxlan-2719744] Not attached      N   enabled

 List of destinations
Name                                    dVnid          vMac               vrf      operSt
====                                    ==             ====               ====     =====
dest-[172.16.1.1]-[vxlan-2719744]       vx      7187319  00:EA:BD:07:3D:7C  RD:RD    enabled
```

Only local PBR is available

Multisite – East-West PBR Provider to Consumer

Traffic Symmetry (rule number 1)

We need to be able to redirect in site2 before sending it to site1 → Consumer pcTag must be available in provider site even if consumer endpoint is unknown

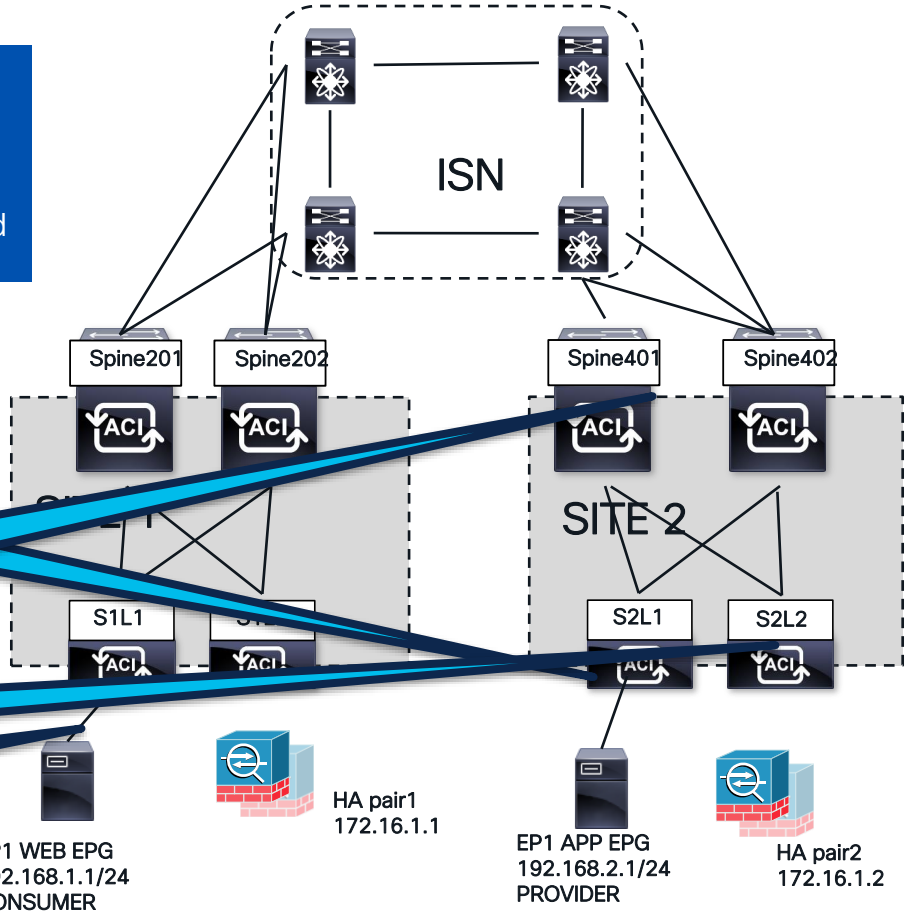
→ The need of configure Consumer Subnet under EPG instead of BD

Provider leaf derives dclass (pcTag of destination EPG aka consumer EPG) from EPG consumer subnet → Redirect zoning-rule always applied there and → spine 2 anycast-mac

Spine site 2 COOP lookup for VMAC HA pair 2 → service leaf site 2 and to firewall

After firewall permit rule to go to consumer (across ISN per regular multisite forwarding

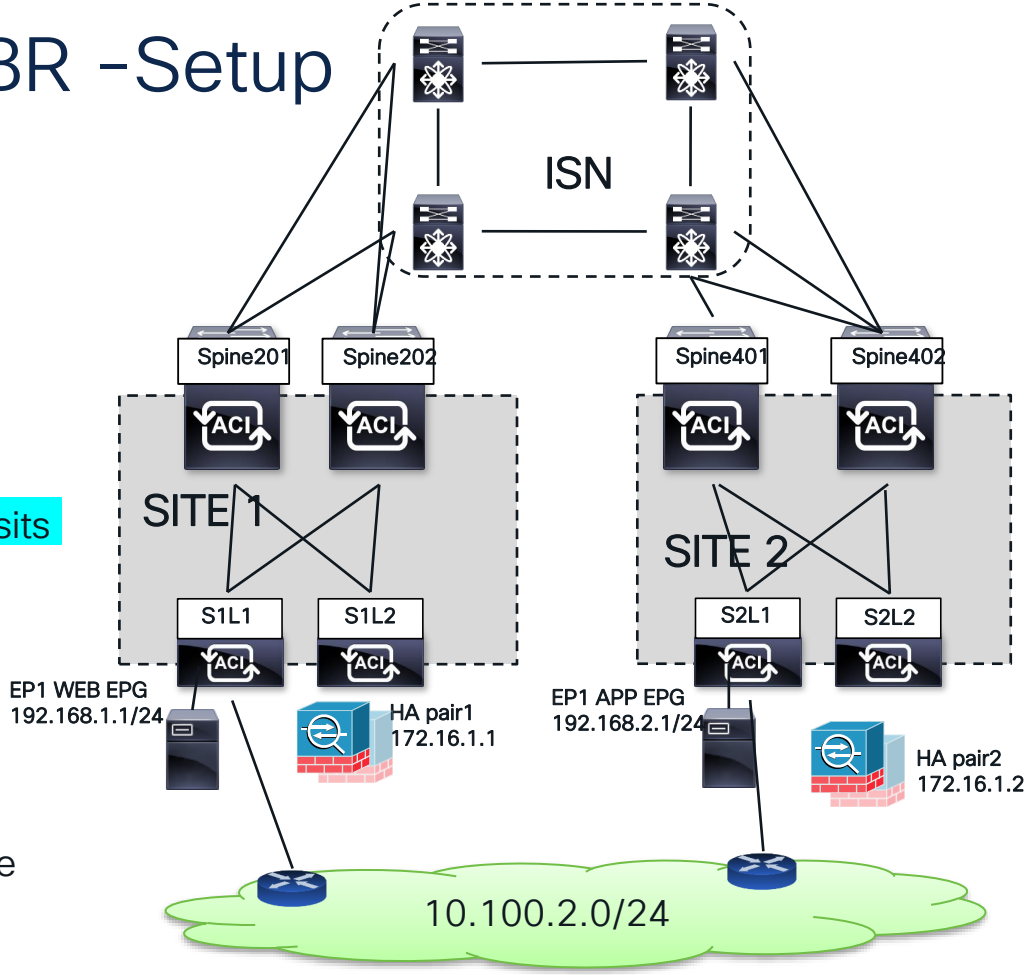
Packet reached consumer



Multisite North-South PBR

Multisite North-South PBR -Setup

- Multisite PBR requirement
  - A site can only redirect to site local PBR Devices
- **Rule 1 : we need to go through same Firewall pair in both direction**
- Per implementation
  - Redirect happens in the site where Server EP sits (not Border leaf)
- North-South specific :
  - Provider or consumer location do not matter
  - Only Server and Border leaf site matters
  - Only vrf enforcement mode ingress supported (default). Needed to ensure all Server leaf have rule to apply contract for external prefix

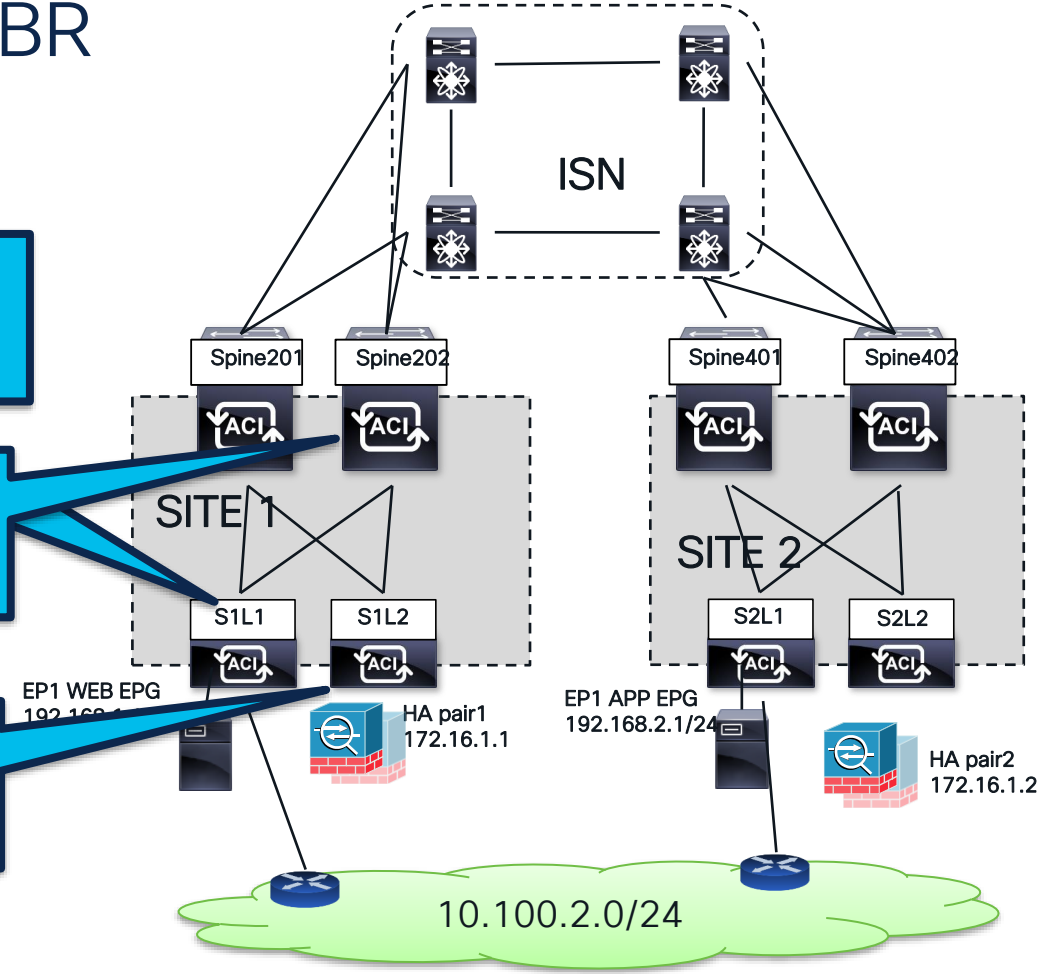


Multisite North-South PBR Endpoint to L3out

Ingress leaf (S1L1) will apply redirect always (dclass or dest pcTag from zoning prefix).
No override rule from EP to L3 out.
Redirect will be to local site HA pair (HA pair1)

Spine site 1 COOP lookup for VMAC lookup in service BD and sends it to S1L2 (service leaf in site1), that forward it to FW

Back from FW a permit rule allows to reach L3 out which may be local (likely) or remote site L3 depending on routing table



Zoning Rule EPG to L3 out – Ingress server leaf

On ingress server leaf dclass to reach L3 out will either by 15 (0.0.0.0/0 prefix) or external EPG pcTag (specific prefix here 16390).
As VRF enforcement is ingress, dclass is always knows

In all case the rule is always redirect with no option override → redirect always apply on this leaf

```
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |   Dir   |  operSt |  Scope  | Action             |    Priority    |
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+
|   4123  | 32772  |   15   | default  | uni-dir | enabled | 2621440 | redir(destgrp-4)   | src_dst_any(9) |
|   4114  | 32772  | 16390  | default  |  bi-dir | enabled | 2621440 | redir(destgrp-4)   | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+---------+--------------------+----------------+

S1P2-Leaf302# show service redir info group 4
4     destgrp-4       dest-[192.168.2.1]-[vxlan-2621440]

S1P2-Leaf302# show service redir info destination ip 192.168.2.1 vnid 2621440
dest-[192.168.2.1]-[vxlan-2621440]       vxlan-15892444  00:EA:BD:07:3D:7C    RD-PBR:RD
```

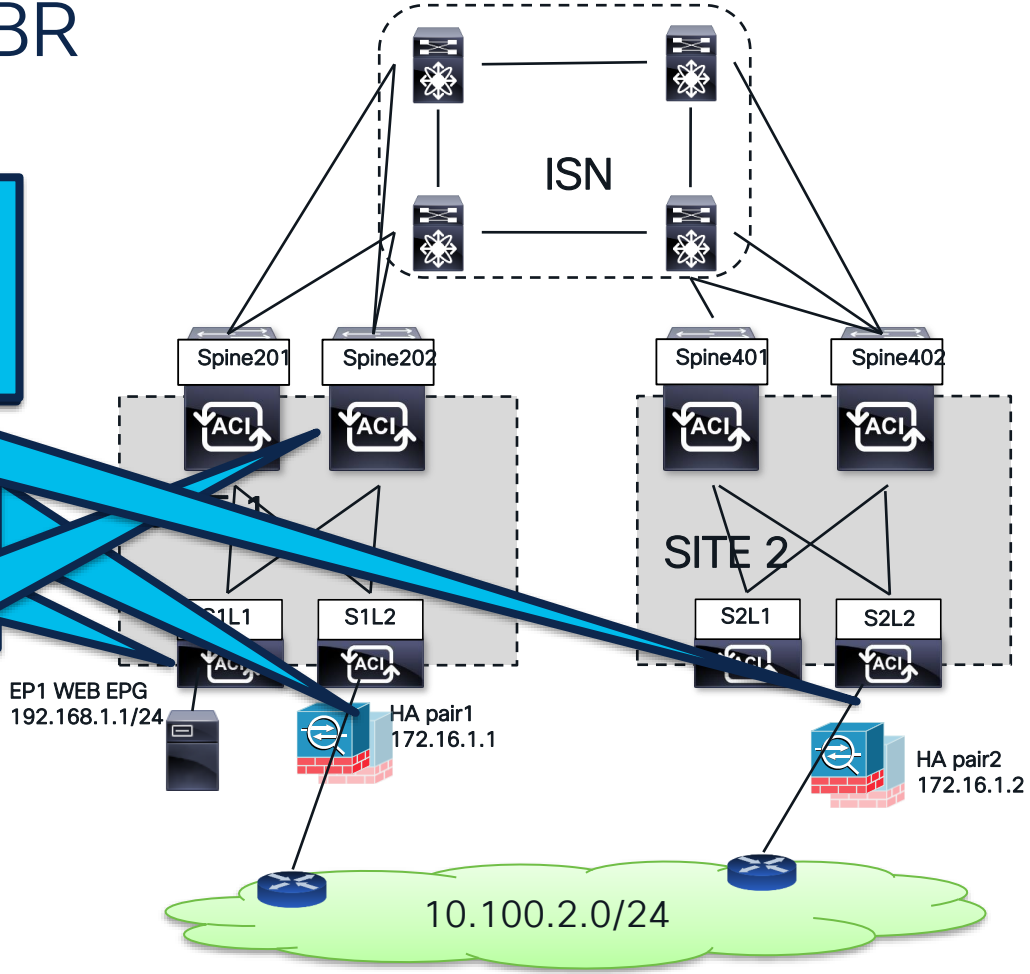
Here both rule are from EPG to L3 out
We will use one or the other depending on the zoning-rule subnet in the external EPG (0.0.0.0/0 or specific subnet)

Multisite North-South PBR L3out to EP

Ingress leaf (BL – S1L2 or S2L2) will hit REDIR+OVERRIDE rule and as destination EP is not local. It will not redirect but will follow regular forwarding to reach destination server leaf (S1L1)

Server leaf S1L1 will apply redirect to HA pair of server site (here HA pair1 in site1).
It will send it to spiny-mac proxy (site1)

Spine site 1 COOP lookup for VMAC lookup in service BD and sends it to S1L2 (service leaf in site1), that forward it to FW



Zoning Rule L3out EPG – Ingress Border leaf

On ingress server leaf sclass from L3out will either by 32770 (vrf pcTag if 0.0.0.0/0 prefix) or external EPG pcTag 16390 (specific prefix).

Rule are always redir +override, so BL will apply permit override unless the destination server End point is also local

```
Leaf# show zoning-rule scope 2621440 dstepg 32772
+---------+--------+--------+----------+---------+---------+---------------------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |              Action             |    Priority    |
+---------+--------+--------+----------+---------+---------+---------------------------------+----------------+
|   4187  | 16390  | 32772  | default  | enabled | 2621440 | redir(destgrp-2),redir_override | src_dst_any(9) |
|   4170  | 32770  | 32772  | default  | enabled | 2621440 | redir(destgrp-2),redir_override | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+---------------------------------+----------------+
```

Here both rule are from Ext Epg to EPG
We will use one or the other depending on the zoning-rule subnet in the external EPG (0.0.0.0/0 or specific subnet)

Zoning Rule L3 out EPG – Server leaf before redirect

On server leaf only redirect Action is present in rule, so we will always redirect here

```
ServerLeaf# show zoning-rule scope 2621440 src-epg 16390
+---------+--------+--------+----------+---------+---------+------------------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Action      |    Priority    |
+---------+--------+--------+----------+---------+---------+------------------+----------------+
|   4163  | 16390  | 32772  | default  | enabled | 2621440 | redir(destgrp-4) | src_dst_any(9) |
|   4127  | 32770  | 32772  | default  | enabled | 2621440 | redir(destgrp-4) | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+------------------+----------------+
```

Multisite PBR – One Slide Summary

Rule East-West

Consumer Subnet MUST BE under EPG

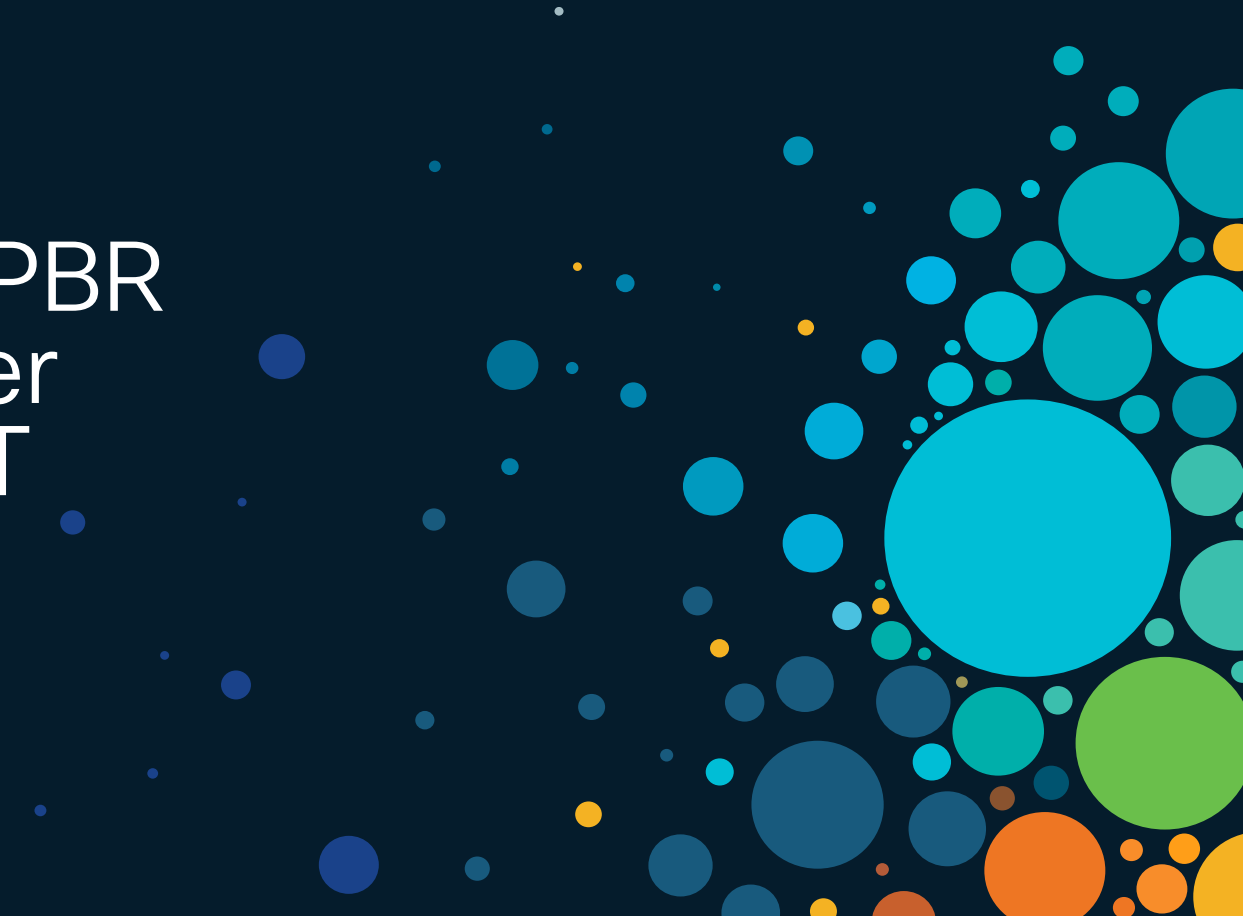
EPG – pcTag (sclass)	EPG pcTag (dclass)	Action	Remark
Consumer	Provider	REDIRECT + OVERRIDE	To ensure redirect is one on site where provider EP sits
Service EPG	Provider	Permit	
Provider	Consumer	REDIRECT	Redirect always done on provider ingress leaf
Service EPG	Consumer	Permit	

Rule North-South

Only vrf enforcement mode ingress supported (default)

EPG – pcTag (sclass)	EPG pcTag (dclass)	Action	Remark
Server EPG	External EPG	REDIRECT	Coming from EP we redirect directly on ingress server leaf
Service EPG	External EPG	Permit	
External EPG	Server EPG	REDIRECT + OVERRIDE	Coming from L3 out we do NOT redirect but we override to be apply redirect on site of incoming server EP
Service EPG	Server EPG	Permit	

Unidirection PBR Load Balancer with no SNAT



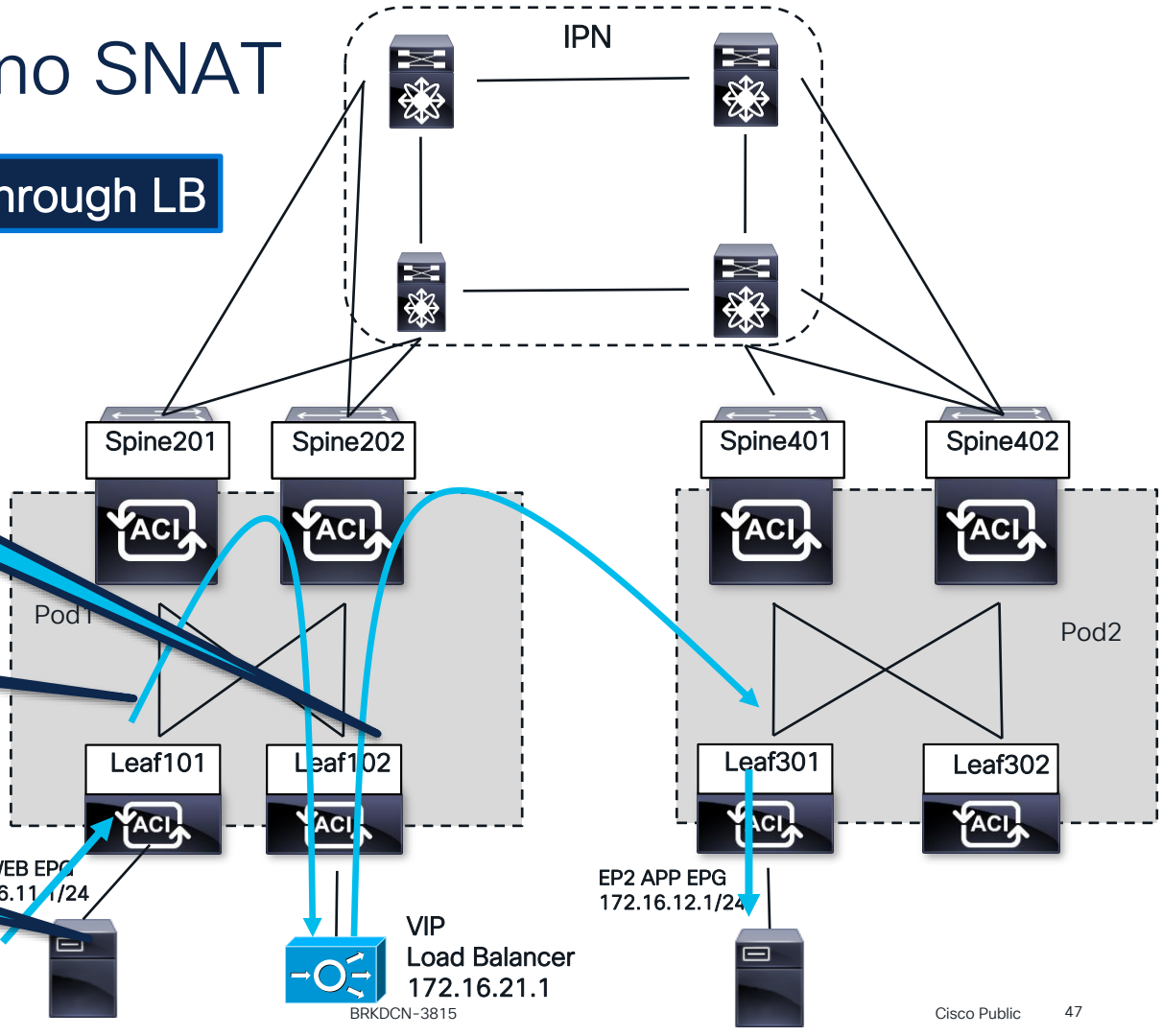
Setup – LB with no SNAT

Traffic from Client to Server through LB

LB does rewrite DIP but not source IP (NO SNAT)
Src IP: 172.16.11.1
Src MAC: VMAC LB
Dest IP: 172.16.12.1 - Real Server
Dest MAC: Leaf MAC

No PBR routing to VIP

Traffic from Client
Src IP: 172.16.11.1
Src MAC: MAC EP1
Dest IP: 172.16.21.1 - VIP
Dest MAC: Leaf MAC



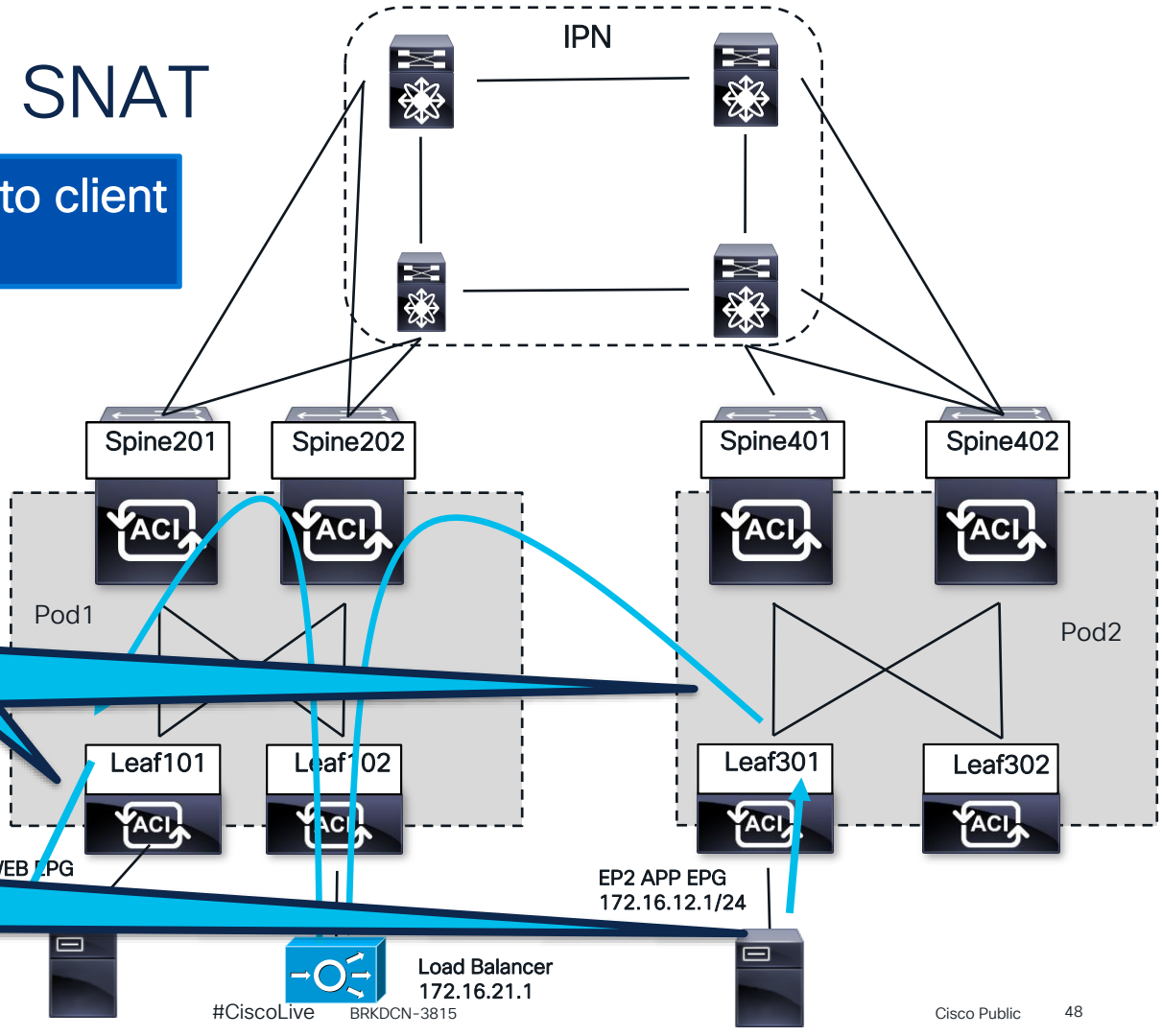
Setup – LB with no SNAT

Traffic back from Real server to client Through LB

Traffic from Service node (After PBR)
Src IP: 172.16.21.1 (VIP)
Src MAC: Leaf MAC
Dest IP: 172.16.11.1
Dest MAC: EP A

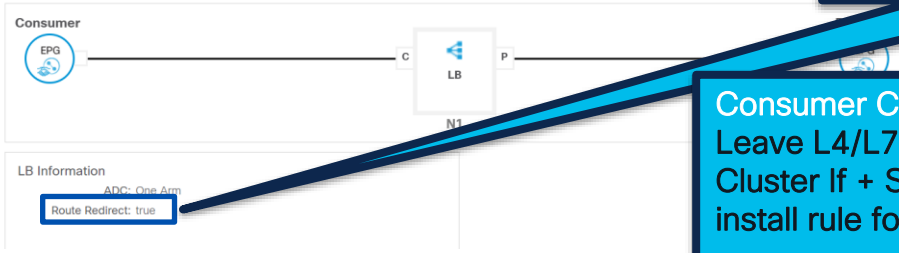
PBR for return traffic
Inner:
Src IP: 172.16.12.1
Dest IP: 172.16.11.1
Dest MAC: VMAC LB
Outer :
Leaf301 TEP to Anycast-mac on spine
VNID : Service BD VNID

Real server replies to Client IP not VIP (NO SNAT) → will bypass LB in return direction if PBR is not used
Src IP: 172.16.12.1
Src MAC: EP2 MAC
Dest IP: 172.16.11.1
Dest MAC: Leaf MAC



Config Gotcha – Unidirectional PBR

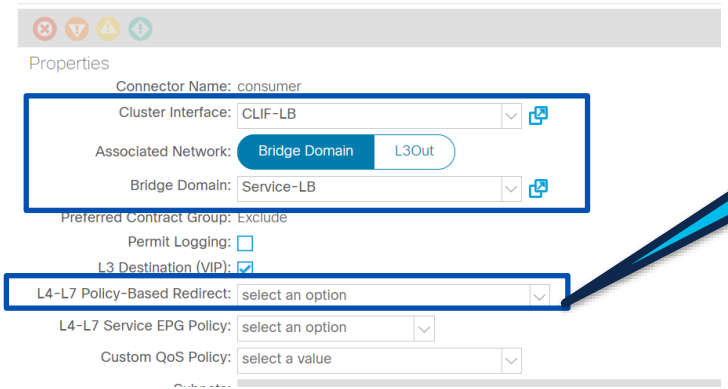
L4-L7 Service Graph Template - LB-NO-SNAT



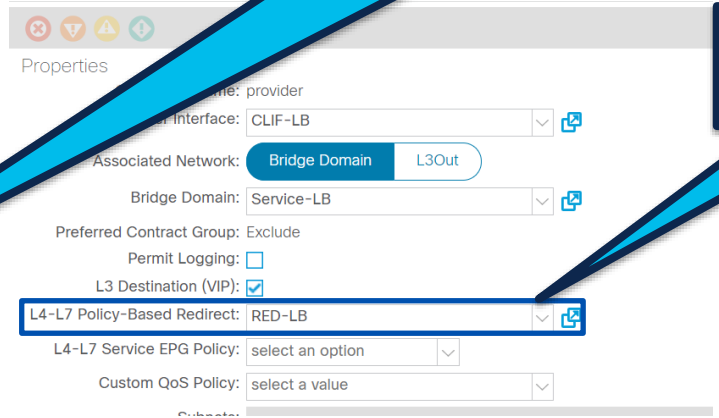
In service graph template
Keep route redirect : True
Even if only one leg needs redirect

Consumer Connector (no PBR):
Leave L4/L7 redirect empty
Cluster If + Service BD will instruct ACI to install rule for consumer to reach service EGP

Logical Interface Context - consumer



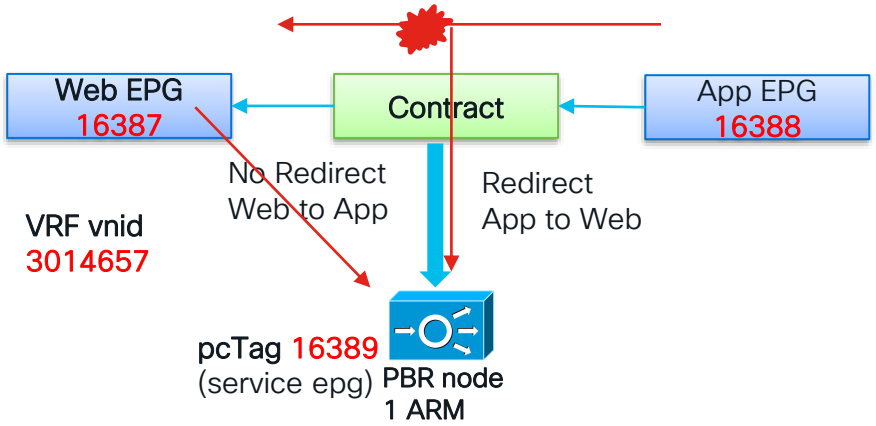
Logical Interface Context - provider



Provider Connector
PBR as usual

Zoning-rule

- Make note all all vnid and sclass involved



Note : There are no rule from Consumer to Provider. As a result of service graph with NO PBR leg we do install rule directly to service EPG Tough there are no direct contract between Web and service EPG

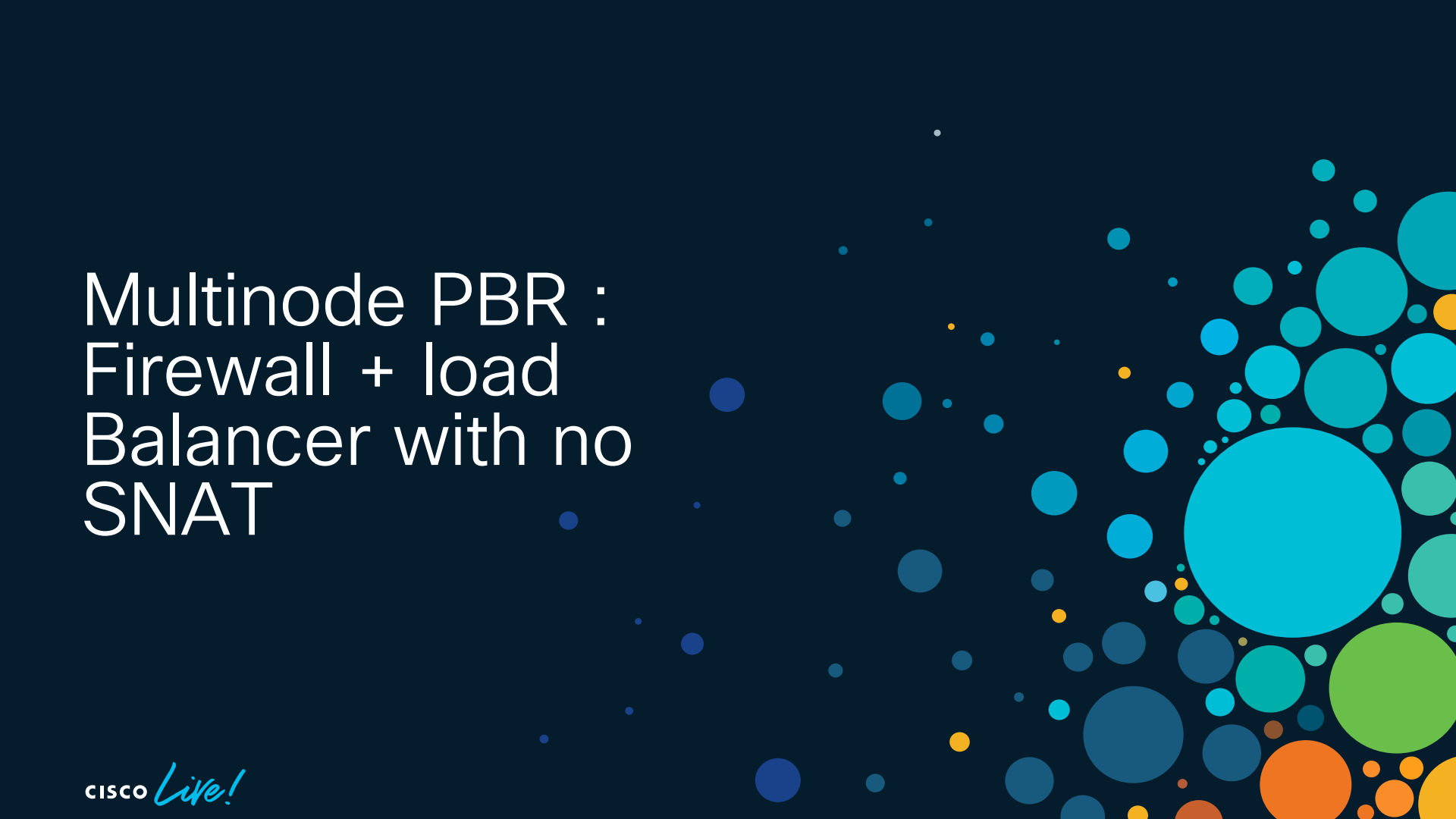
- Check expected zoning-rule

1. Cons to Prov (replaced by Consumer to Service EPG because DIP is VIP in service EPG) : 16387 to 16389 : PERMIT
2. Shadow to Prov : 16389 to 16388 : PERMIT
3. Prov to Cons : 16388 to 16387 : REDIRECT
4. Shadow to Cons : 16389 to 16387 : PERMIT

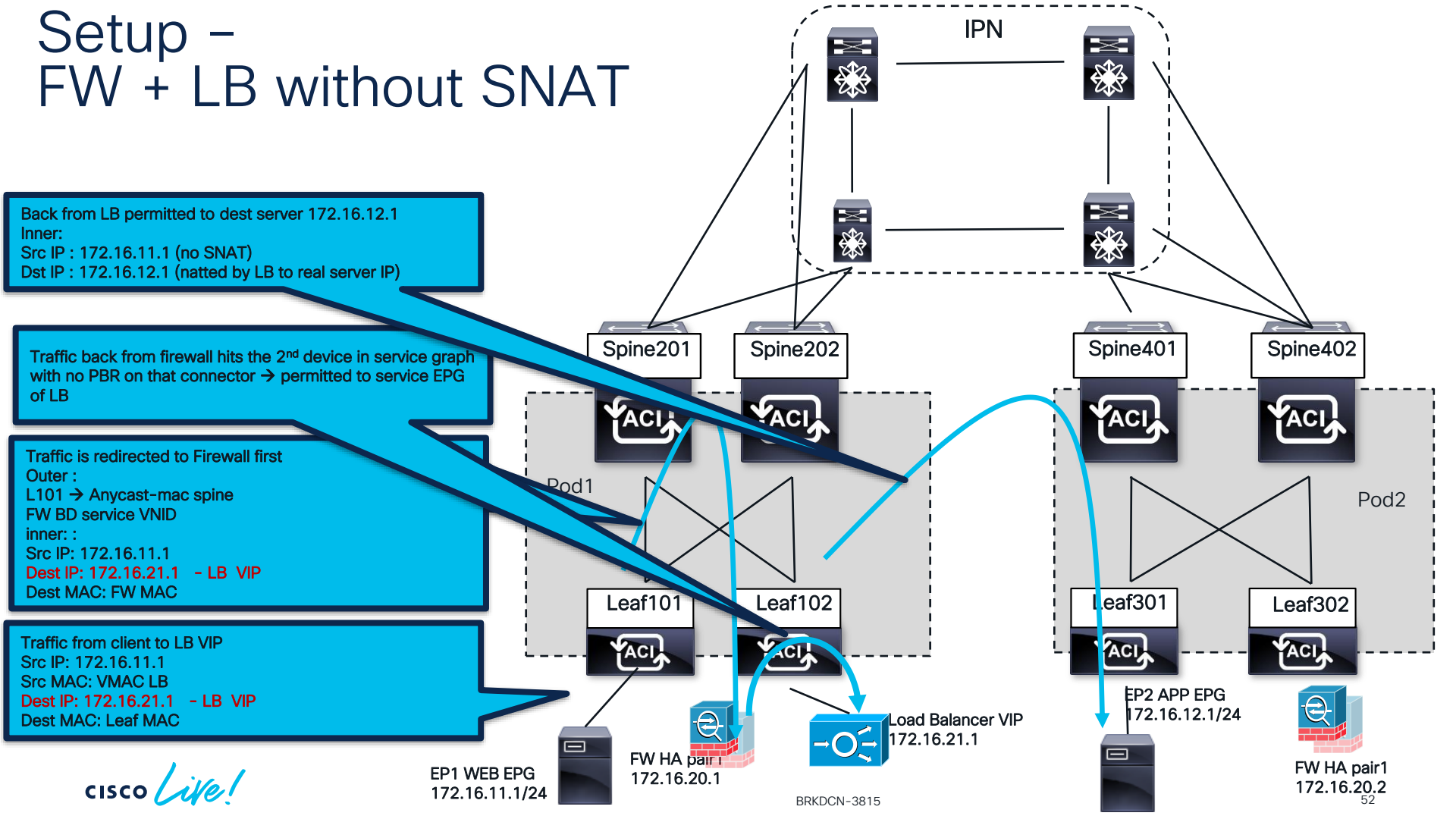
```
S1P1-Leaf101# show zoning-rule scope 3014657
+---------+--------+--------+----------+---------+---------+------------------+---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Action      |   Priority    |
+---------+--------+--------+----------+---------+---------+------------------+---------------+
| 4195    | 16388  | 16387  | default  | enabled | 3014657 | redir(destgrp-9) | src_dst_any(9)|
| 4177    | 16389  | 16387  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
| 4197    | 16387  | 16389  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
| 4196    | 16389  | 16388  | default  | enabled | 3014657 |      permit      | src_dst_any(9)|
```

3 4 1 2

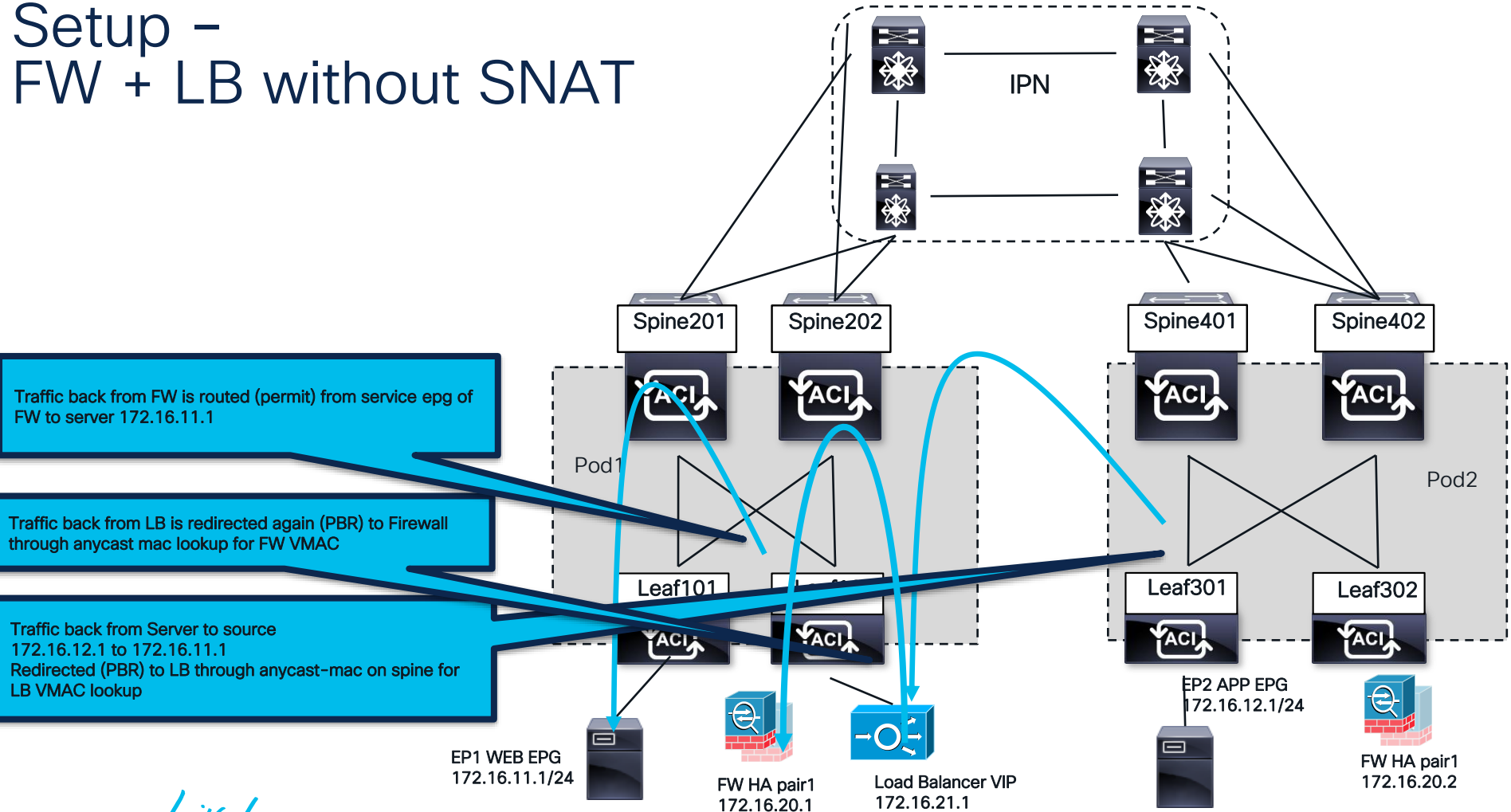
Multinode PBR : Firewall + load Balancer with no SNAT



Setup – FW + LB without SNAT

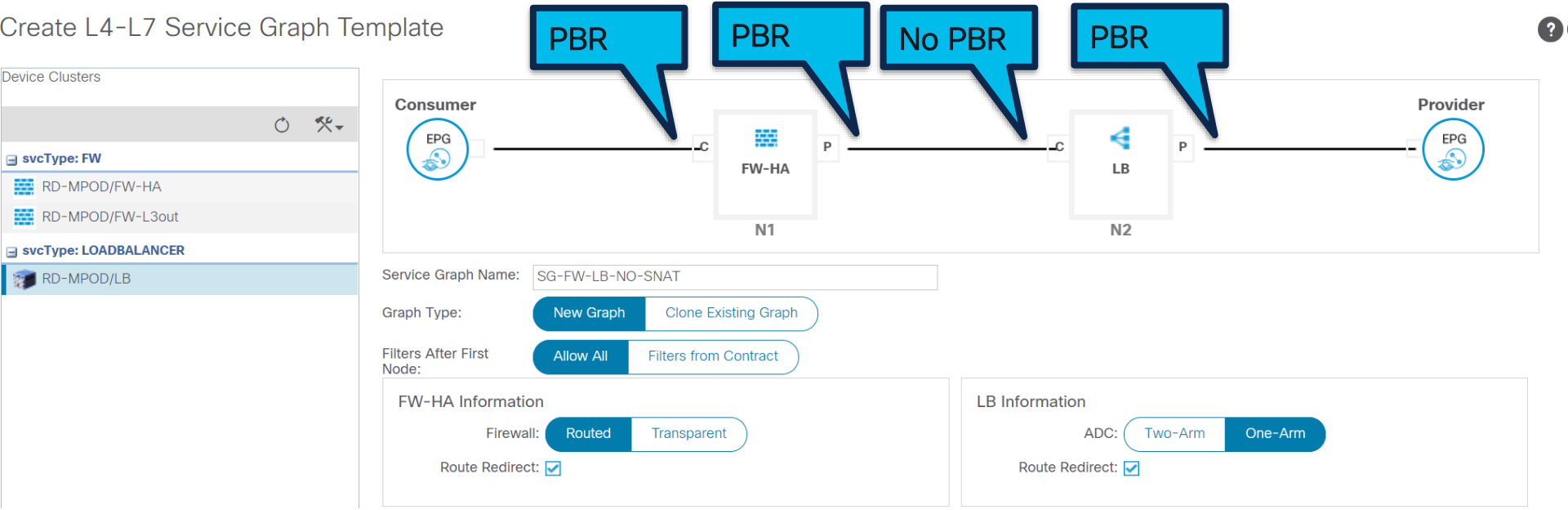


Setup – FW + LB without SNAT



Config Gotcha- Service Graph – L4/L7 device chaining

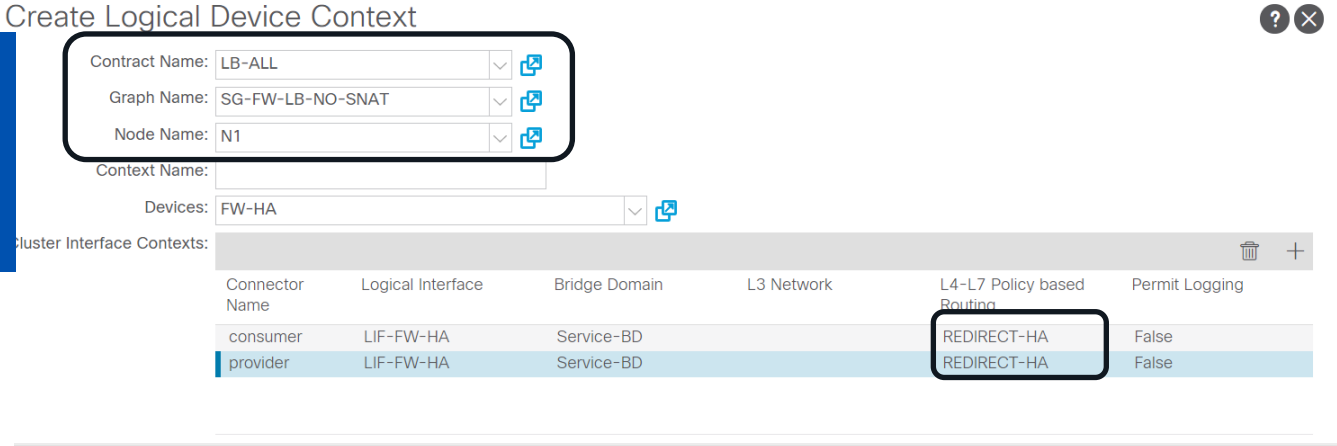
Out of the 4 connectors only Consumer connector of N2 (LB)
Will not use PBR



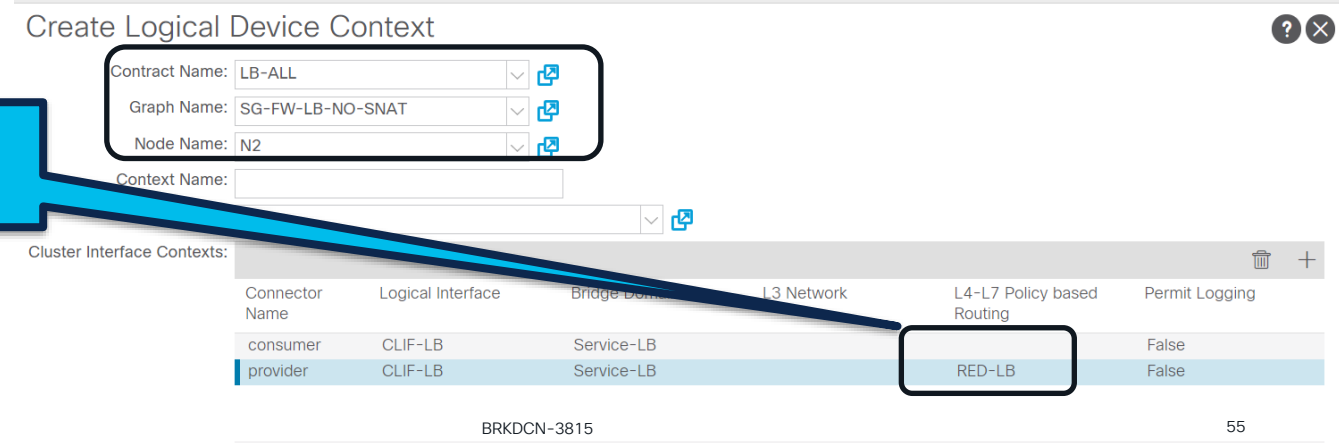
Config – Device Selection policies

One logical Device Context per node in the Service chain

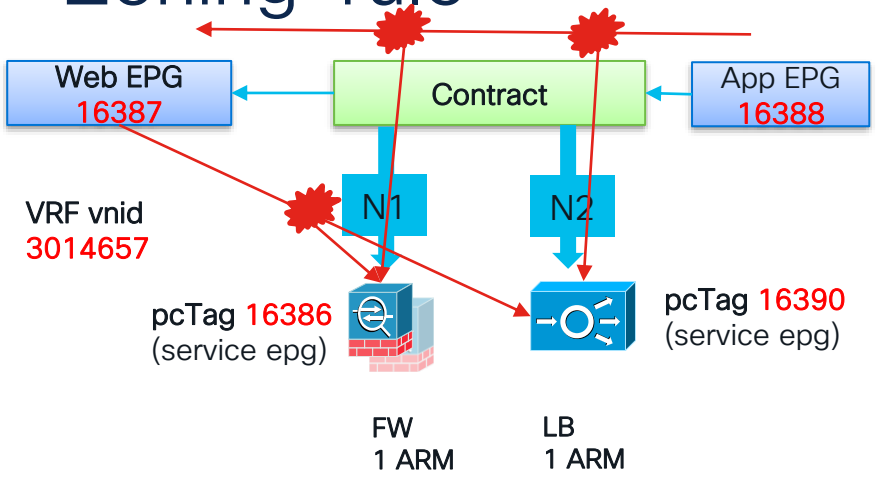
Both use same
Contract name
Graph Name
Different Node N1 or N2



There is no redirect on the Consumer leg of N2 (LB)



Zoning-rule

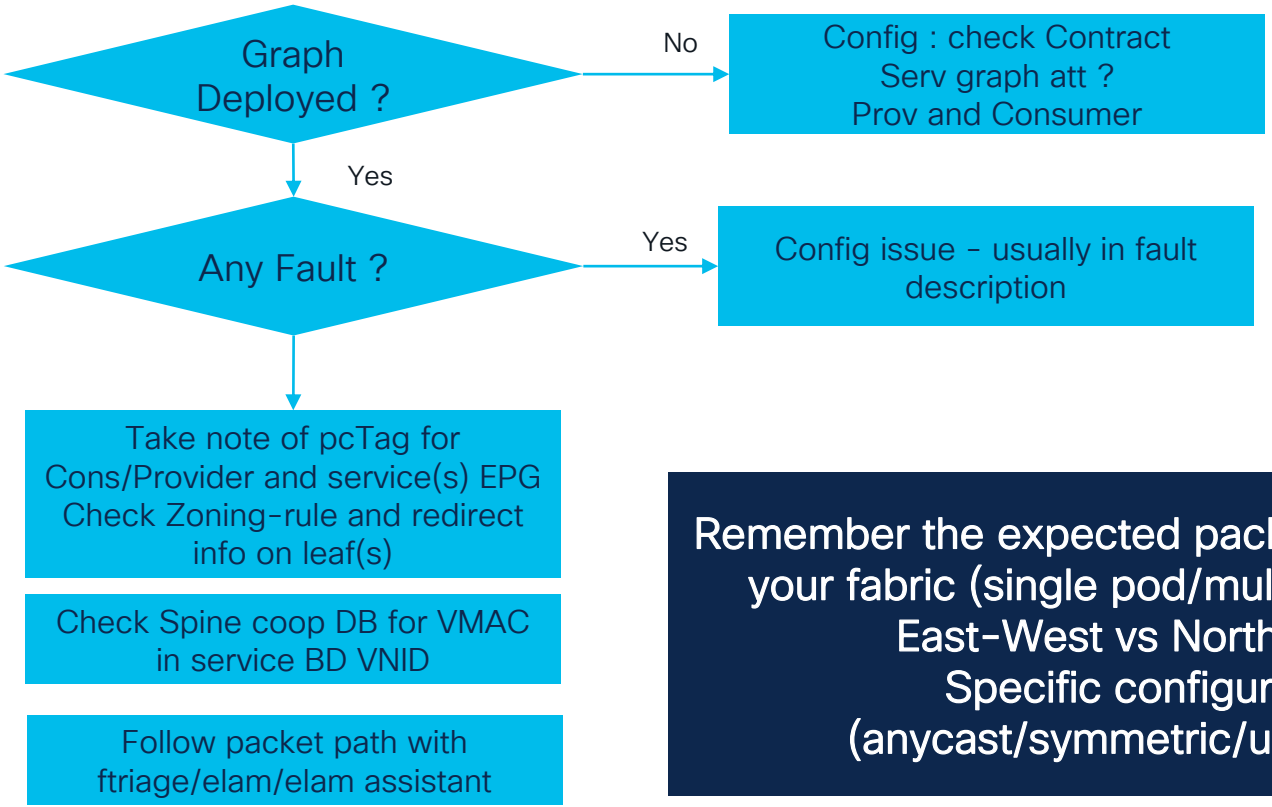


1. Cons to Prov (replaced by Consumer to Service EPG because DIP is VIP in service EPG) : 16387 to 16390 : REDIRECT to FW (group8)
2. Service EPG FW to VIP LB : 16386 to 16390 : PERMIT (no PBR on Cons Leg N2)
3. Service EPG LB to Prov : 16390 to 16388 : PERMIT
4. Prov to Cons : 16388 to 16387 : REDIRECT to LB (group 10) (as no SNAT on LB)
5. Service EPG LB to Cons : 16390 to 16387 : REDIRECT to FW (group 8)
6. Service EPG FW to Cons : 16286 to 16387 : PERMIT

```
S1P1-Leaf101# show zoning-rule scope 3014657
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Actio         | Priority      |
+---------+--------+--------+----------+---------+---------+--------------------+---------------+
| 1  4196 | 16387  | 16390  | default  | enabled | 3014657 | redir(destgrp-8)   | src_dst_any(9)|
| 2  4177 | 16386  | 16390  | default  | enabled | 3014657 |      permit        | src_dst_any(9)|
| 3  4197 | 16390  | 16388  | default  | enabled | 3014657 |      permit        | src_dst_any(9)|
| 4  4195 | 16388  | 16387  | default  | enabled | 3014657 | redir(destgrp-10)  | src_dst_any(9)|
| 5  4197 | 16390  | 16387  | default  | enabled | 3014657   redir(destgrp-8)  | src_dst_any(9)|
| 6  4177 | 16386  | 16387  | default  | enabled | 3014657 |      permit        | src_dst_any(9)|
```

Summary

Troubleshooting PBR checklist



Remember the expected packet path based on your fabric (single pod/multipod/multisite)
East-West vs North-South
Specific configuration (anycast/symmetric/unidir PBR/

Summary – PBR and firewall deployment options

Firewall integration model	Multipod East-West	Multipod North-South	Multisite East-West	Multisite North-South
Active and standby across Pod/Site	OK - Simple PBR	OK – simple PBR	NOK	NOK
Active/Active FW across POD	OK with anycast PBR	OK with anycast PBR	NOK	NOK
Active/Standby per pod site	OK symmetric PBR	OK either symmetric PBR or pod aware (+option Host based routing)	OK with PBR – Redirect on provider site	OK with PBR – Redirect on Server leaf site

Packet Format

From EP1 to ACI Leaf 101

1

Inner			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC ACI

From Leaf 101 to Fabric in case EP 2 is unknown in leaf 101

2

Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	xx	xx	VNID VRF	Sclass EPG1	DIP Anycast-v4	SIP Leaf101 PTEP

After Redirect (either by Leaf 101 or by leaf 301

3

Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC FW	VNID Service BD	Sclass EPG1	DIP Anycast-Mac	SIP Leaf 101

After FW MAC coop lookup in service BD

4

Inner				Outer			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC EP1	DMAC FW	VNID Service BD	Sclass EPG1	DIP - PTEP FW Leaf or Remote Pod	SIP Leaf101 PTEP

Coming back from FW on leaf 102 or 302

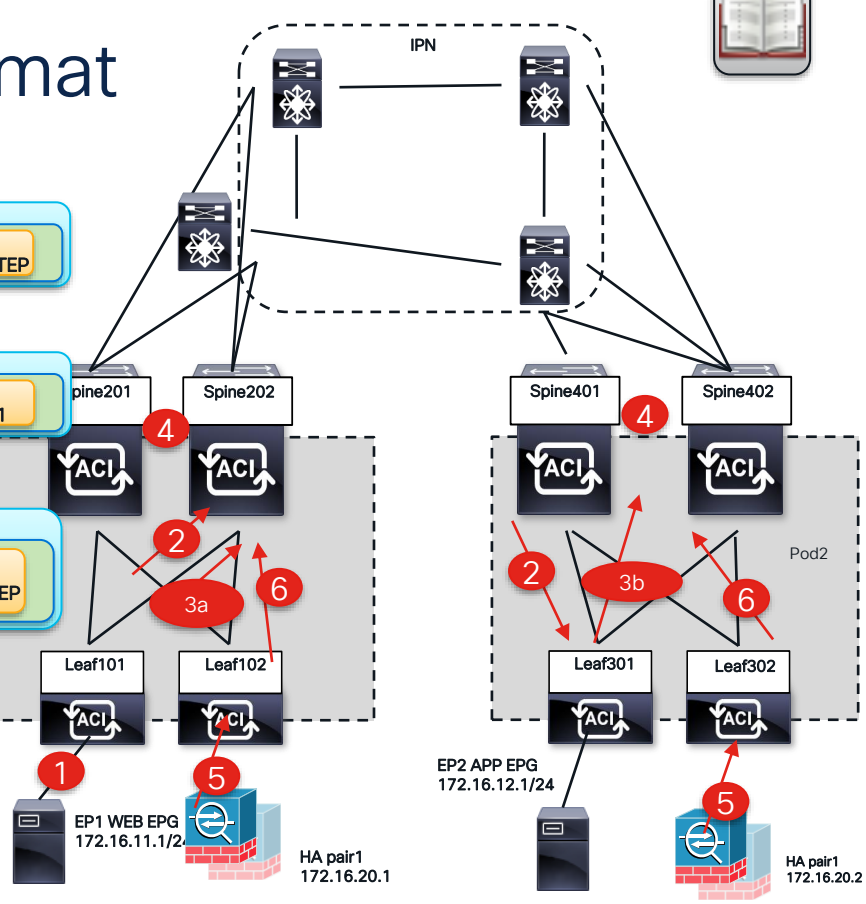
5

Inner			
DIP 172.16.12.1	SIP 172.16.11.1	SMAC FW	DMAC ACI

After FW, between Fw leaf and destination leaf 301

6

Inner							
DIP 172.16.12.1	SIP 172.16.11.1	xxx	xxx	VNID VRF	Sclass Shadow +No Learn	DIP - PTEP Leaf 301 or acast-v4	SIP - PTEP FW Leaf



Datapath Troubleshooting Tool Elam trigger to use

On leaf from front panel Port (ingress from server, or back from firewall)

```
debug platform internal roc elam asic 0
    trigger reset
    trigger init in-select 6 out-select 1
      set outer ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
      set outer l2 src_mac < MAC src EP >
```

On Spine or egress leaf Before Redirect (for example if EP is unknown in ingress leaf)

```
debug platform internal roc elam asic 0
    trigger reset
    trigger init in-select 14 out-select 1
      set inner ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
      set outer l4 tn-seg-id 0x2e001     (VRF VNID)
```

On Spine or egress leaf after redirect and before Firewall

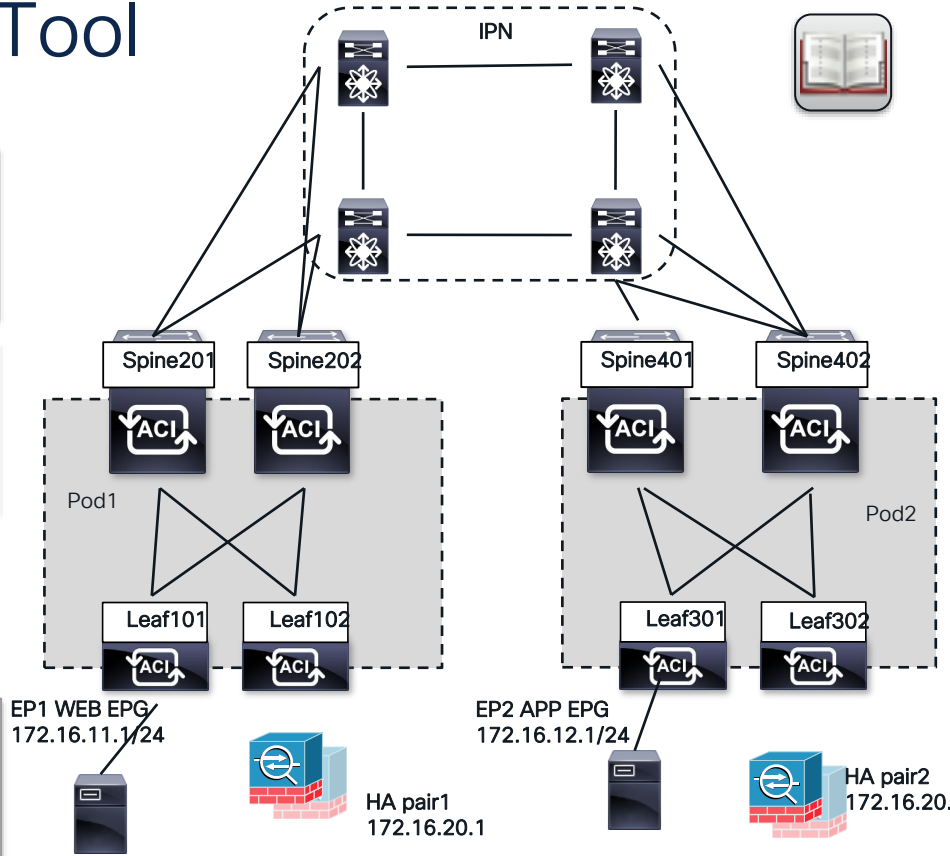
```
debug platform internal roc elam asic 0
    trigger reset
    trigger init in-select 14 out-select 1
      set inner ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
      set outer l4 tn-seg-id 0xe27fef     (service BD VNID)
```

On leaf from front panel Port (ingress from server, or back from firewall)

```
debug platform internal roc elam asic 0
    trigger reset
    trigger init in-select 6 out-select 1
      set outer ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
      set outer l2 src_mac < MAC FIREWALL >
```

On spine or egress leaf (ingress from server, or back from firewall)

```
debug platform internal roc elam asic 0
    trigger reset
    trigger init in-select 14 out-select 1
      set inner ipv4 src_ip 172.16.11.1 dst_ip 172.16.12.1
      set inner l2 src_mac < MAC FIREWALL >
```



Note : ASIC is roc in FX and later hardware and tah in EX

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

cisco Live!

Thank you

cisco Live!


cisco Live!
ALL IN