cisco Live!

Let's go

The bridge to possible

ACI Troubleshooting: A deep dive into PBR

Roland Ducomble, Principal Engineer CX EMEA
CCIE 3745

BRKDN-3615

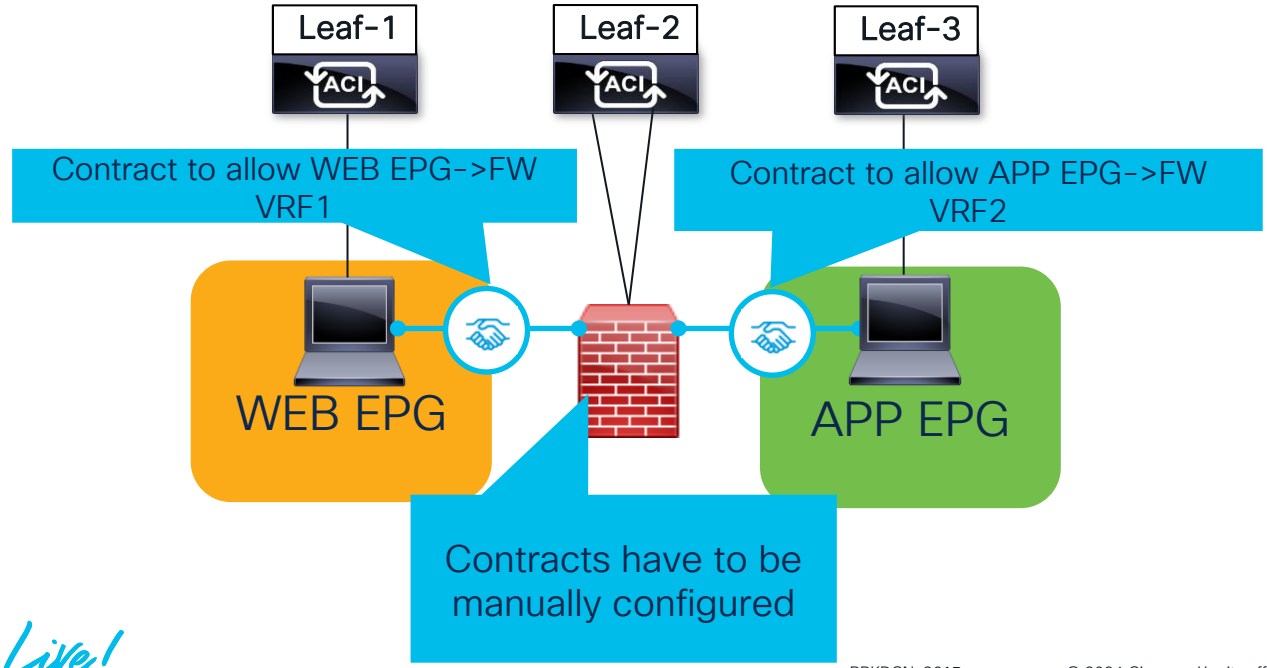
Agenda

- Introduction
- PBR packet walk with CLI
- Use case :
  - Multipod East-West PBR
  - Multisite East-West PBR
  - Multisite vzAny to vzAny PBR

Why and How PBR ?

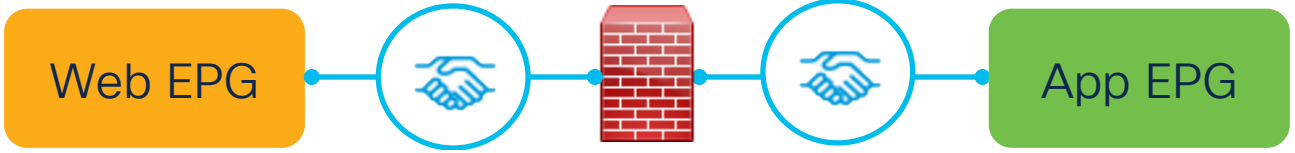
Service Insertion without PBR

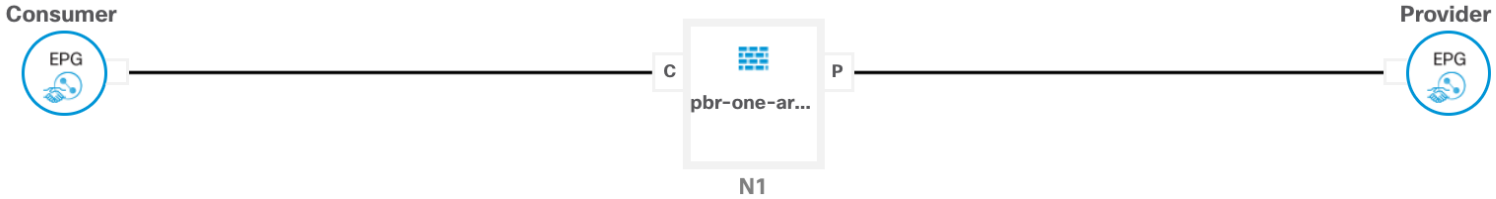
- Requires more manual contract
- More complex route configuration to send traffic to firewall usually with vrf stitching
- VRF stitching mandate two-arm firewall

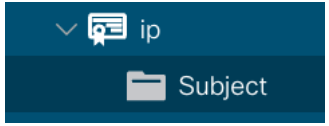


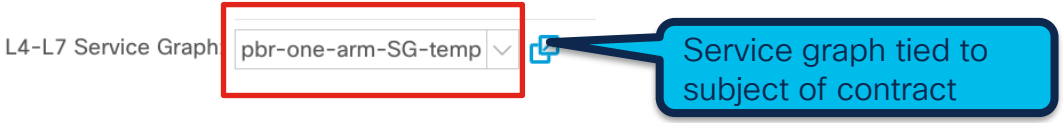
Why service graphs with PBR in ACI?

- Service graphs automate contract configuration
- Easy insertion within a VRF (no VRF stitching)
- One-arm or two-arm possible
- PBR gives us the ability to attach forwarding constructs to contracts
- Like regular contract they are implemented using zoning-rule on leaf.





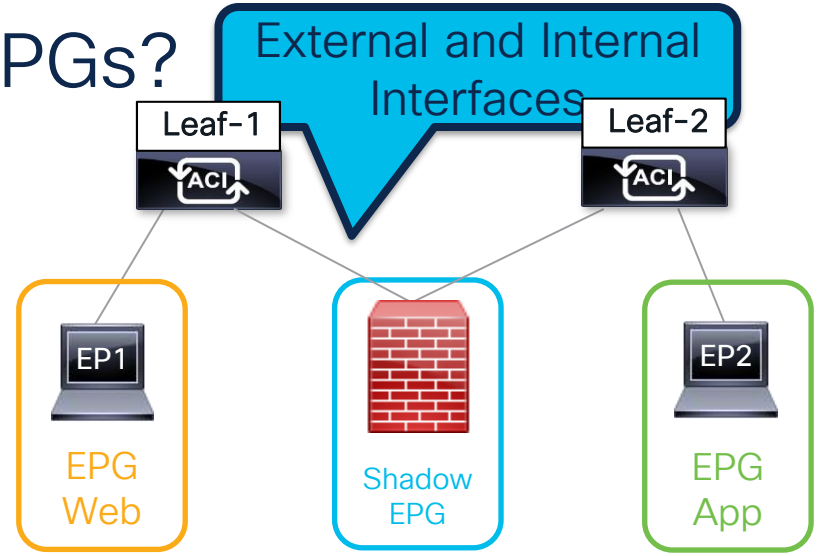


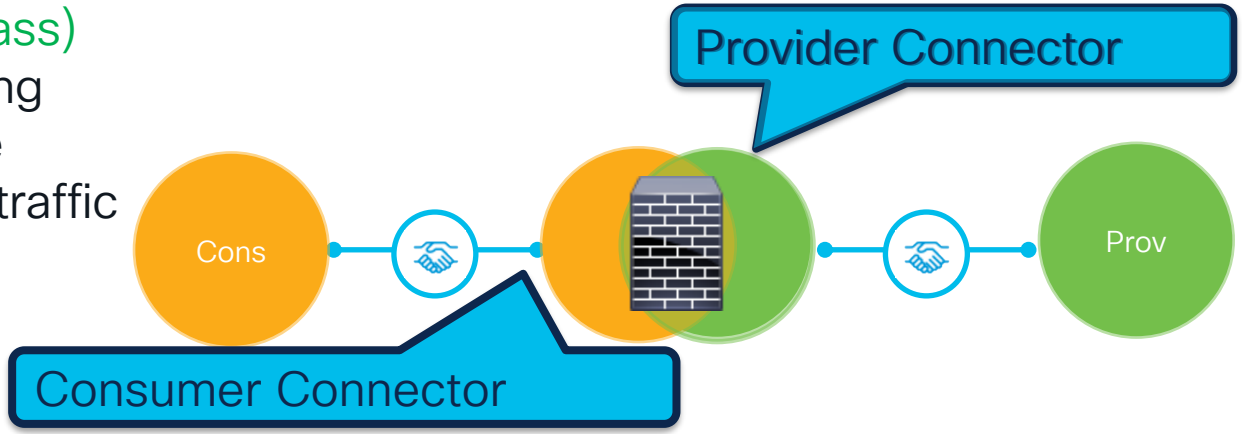


What are shadow/Service EPGs?

A *'two armed'* example

- Shadow/Service EPGs connect to the service Device
- External Interface is called the "Consumer Connector"
- Internal interface is the "Provider Connector"
- Each is represented by a VLAN and has its own PcTag (sclass)
- Kind of a way of stitching EPG VLAN and service node VLAN to "steer" traffic to service node

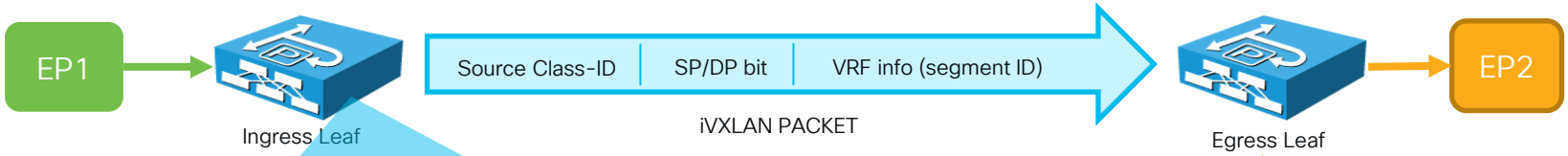




Contract enforcement – reminder

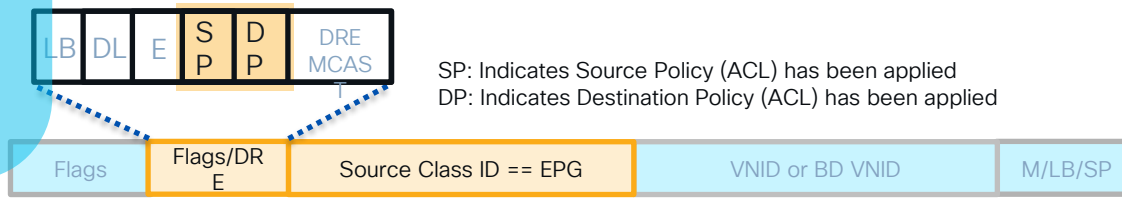
Sources and Destinations Must be Classified into EPG's or ESG's

Every EPG is mapped to its own `unique` classID (or pcTag or sclass)



- Ingress Leaf assigns the source ClassID (**sclass**) to the packet based on ingress vlan/interface (EPG) or Src IP policy prefix(L3out)
- If the destination ClassID can be derived from either known dest EP or (**dclass**) , Applies the policy (ACL – zoning-rule) and sets a `Policy Applied` bit in iVXLAN header if done (SP/DP bit to 1)
- if dest classID can't be derived, do not set SP/DP bit (SP/DP bit to 0)

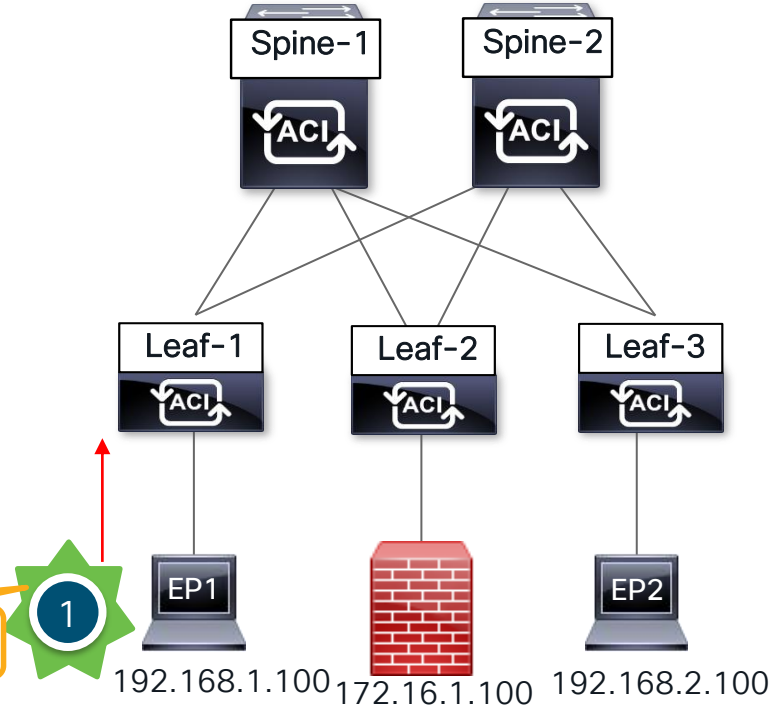
If policy has not been applied (sp=0/dp=0), egress leaf checks the EPG for the destination (local) and applies policy



SP: Indicates Source Policy (ACL) has been applied
DP: Indicates Destination Policy (ACL) has been applied

PBR packet walk

PBR Packet Walk – Step 1

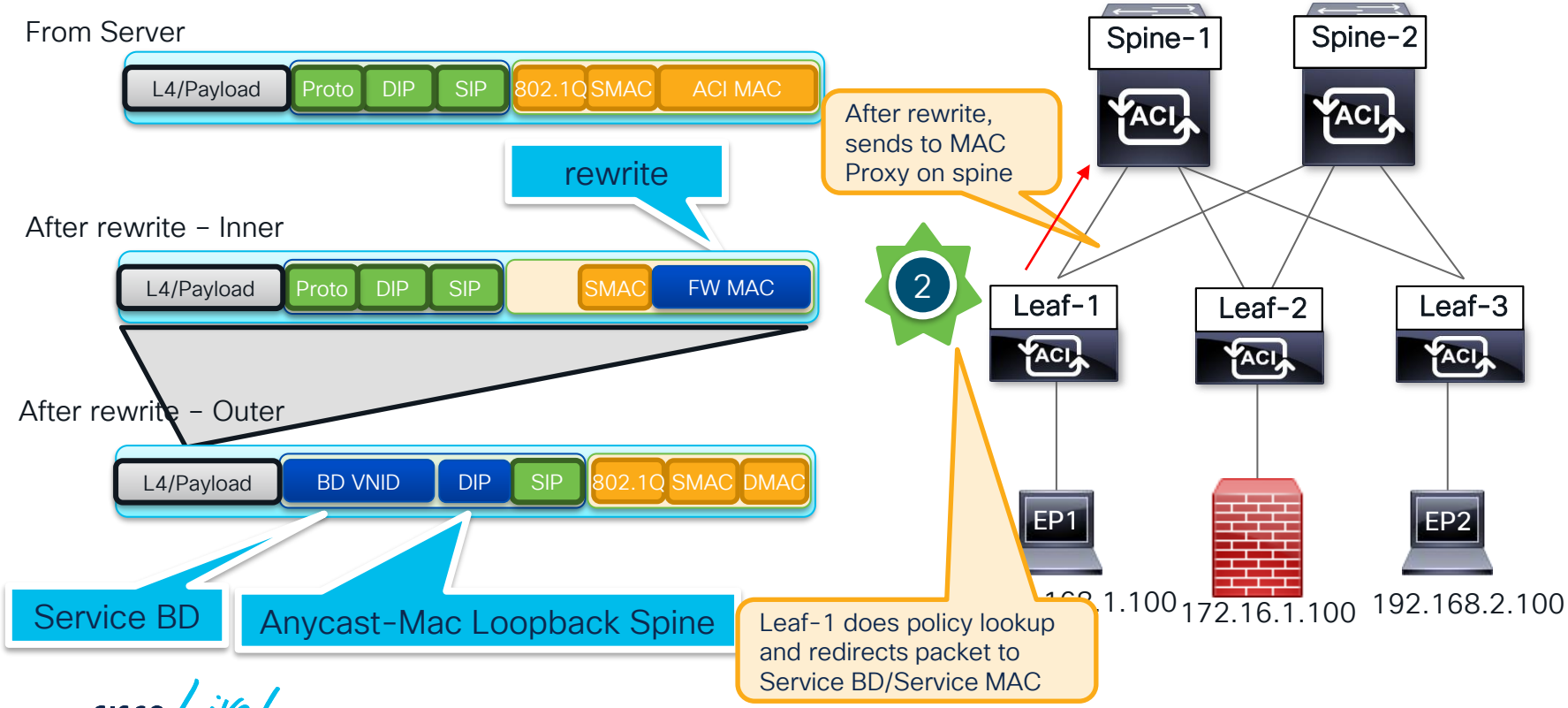


ELAM trigger Leaf-1 in vsh_lc

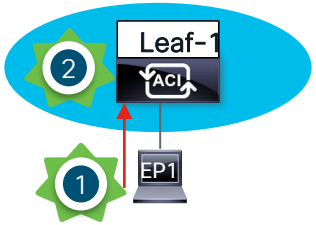
EP1 sends packet to EP2 via Leaf-1

```
debug platform internal roc elam asic 0
    trigger init in-select 6 out-select 1
      reset
      set outer ipv4 src_ip 192.168.1.100 dst_ip 192.168.2.100
      start
```

PBR Packet Walk – Step 2 (Packet Rewrite)



Command Line Verification



Dest EP is known by leaf-1

If destination EP is known : redirect happens on ingress leaf
---
If destination EP is unknown: redirect will happen on egress leaf

Confirm pcTags of traffic flow

```
a1-leaf1# show system internal epm endpoint ip
192.168.1.100 | egrep "VRF vnid|sclass"
BD vnid : 16089032 ::: VRF vnid : 2293762
Flags : 0x80004c04 ::: sclass : 32771
```

sclass for source EP

```
a1-leaf1# show system internal epm endpoint ip
192.168.2.100 | egrep "VRF vnid|sclass"
BD vnid : 15826920 ::: VRF vnid : 2293762
Flags : 0x80004c04 ::: sclass : 16386
```

dclass for dest EP

Verify zoning-rule has 'redir' action and matches desired traffic type (ex. ip traffic)

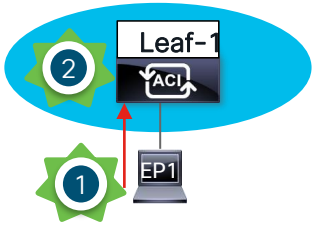
```
a1-leaf1# show zoning-rule scope 2293762 src-epg 32771 dst-epg 16386
+---------+--------+--------+----------+--------+---------+-----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  Dir   | Scope   |  Action         |
+---------+--------+--------+----------+--------+---------+-----------------+
|   4308  | 32771  | 16386  |    1     | bi-dir | 2293762 | redir(destgrp-3)|
```

Redirect happening on ingress leaf since destination is known

```
a1-leaf1# show zoning-filter filter 1
+----------+------+--------+-------------+-------------+-------------+-------------+
| FilterId | Name | EtherT |  SFromPort  |   SToPort   |  DFromPort  |   DToPort   |
+----------+------+--------+-------------+-------------+-------------+-------------+
|    1     | 1_0  |   ip   | unspecified | unspecified | unspecified | unspecified |
```

IP Filter

Command Line Verification



Check redirect policy to see how packets will be redirected

```
a1-leaf1# show service redir info group 3
=========================================================================
GrpID       Name            destination
3           destgrp-3       dest-[172.16.1.100]-[vxlan-2293762]
```

Service node redirect IP

Service node VRF VNID

```
a1-leaf1# show service redir info destination ip 172.16.1.100 vnid 2293762
=========================================================================================
Name                                    bdVnid          vMac                rf
====                                    ======          ====                ==
dest-[172.16.1.100]-[vxlan-2293762]     vxlan-16744311  00:50:56:A8:48:97   mg-cisco-live:v1
```

Parameters used to build redirected vxlan packet

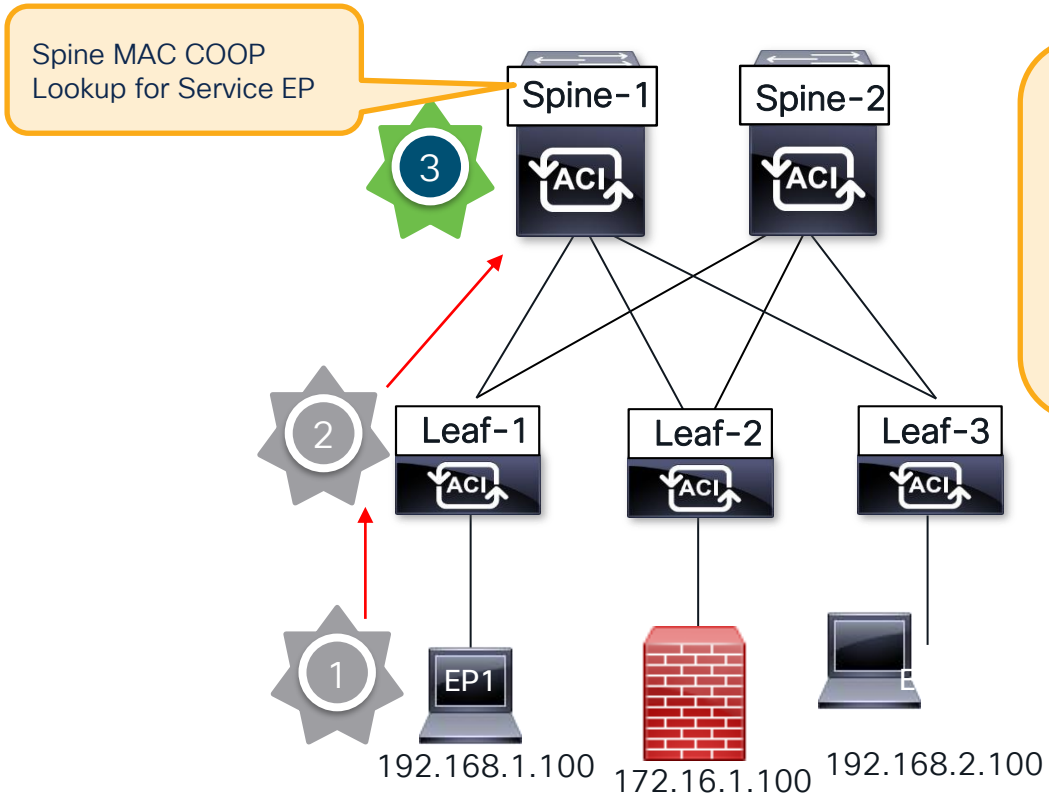
```
module-1(DBG-elam-insel6)# report detail | grep service_redir

sug_luc_latch_results_vec.luc3_0.service_redir: 0x1
```

0x1 – yes, redirected
0x0 – not redirected

Packet is being redirected

PBR Packet Walk – Step 3 (Spine COOP MAC Lookup)

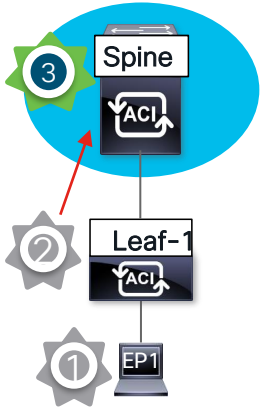


Note : Mac address misconfiguration in redirect policy will lead to packet drop on spine (COOP lookup miss)

Command Line Verification

Outer L3 Header	
L3 Type	IPv4
Destination IP	10.0.200.65 (MAC Spine-Proxy)
Source IP	10.0.216.64 (a1-leaf1)

ELAM ingress on Spine sees Destination TEP as MAC Proxy



Source EPG (sclass / src pcTag)	0x8003 / 32771 (mg-cisco-live:a1:e1)
VRF/BD VNID	0xFF7F77 / 16744311 (mg-cisco-live:service_bd)

Rewrite info! Service BD VNID

Verify Spine has installed MAC of service node in COOP

```
a1-spine1# show coop internal info repo ep key 16744311 00:50:56:A8:48:97 | egrep
"Tunnel|EP" | head -n 3
EP bd vnid : 16744311
EP mac :  00:50:56:A8:48:97
Tunnel nh : 10.0.216.68
```

Service BD VNID

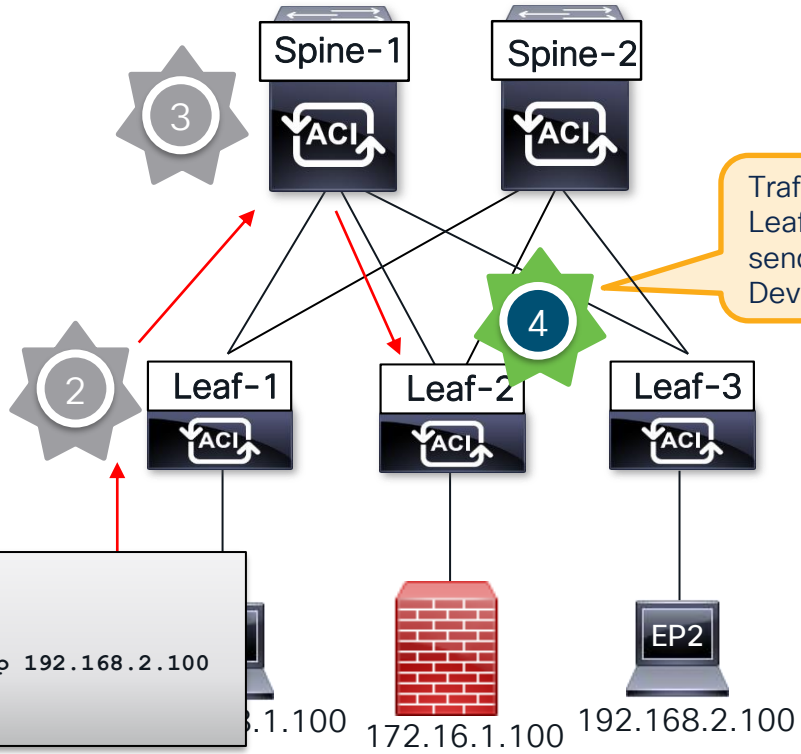
Service Device MAC

Map tunnel destination address to leaf

Tunnel points to Leaf 102 where service device is connected

```
apic1# moquery -c ipv4Addr -f 'ipv4.Addr.addr=="10.0.216.68"' | grep dn
dn                  : topology/pod-1/node-102/sys/ipv4/inst/dom-overlay-1/if-[lo0]/addr-[10.0.216.68/32]
```

PBR Packet Walk – Step 4 (Service Leaf)



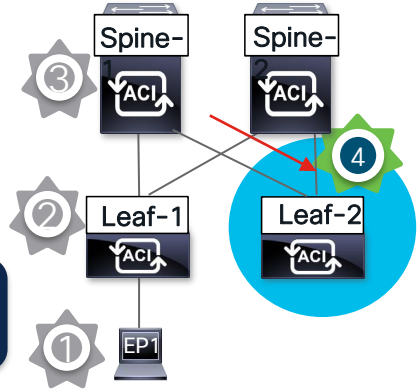
Traffic is sent to Service Leaf (Leaf-2) & Leaf-2 sends traffic to Service Device

ELAM trigger Leaf-2 in vsh_lc

```
debug platform internal roc elam asic 0
    trigger init in-select 14 out-select 1
      reset
      set inner ipv4 src_ip 192.168.1.100 dst_ip 192.168.2.100
      set inner l2 dst_mac <FW-MAC>
      start
```

Command Line Verification

Verify Service Device programming on Service Node (Leaf-2)




```
a1-leaf2# show system internal epm endpoint mac 0050.56a8.4897

MAC : 0050.56a8.4897 ::: Num IPs : 1
IP# 0 : 172.16.1.100 ::: IP# 0 flags : host-tracked
Vlan id : 57 ::: Vlan vnid : 10792 ::: VRF name : mg-cisco-live:v1
BD vnid : 16744311 ::: VRF vnid : 2293762
Phy If : 0x1a02a000 ::: Tunnel If : 0
Interface : Ethernet1/43
Flags : 0x80004c04 ::: sclass : 16389 ::: Ref count : 5
EP Create Timestamp : 03/28/2023 15:23:44.027077
EP Update Timestamp : 04/14/2023 13:52:41.683129
EP Flags : local|IP|MAC|sclass|timer|
```

Checking Service MAC Learning

Shadow EPG pcTAG

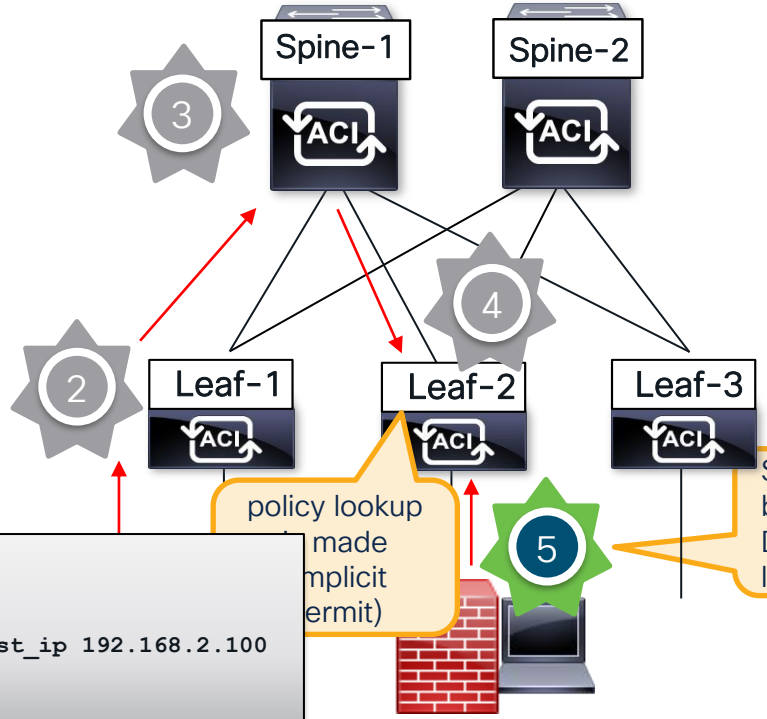
```
a1-leaf2# show vlan id 57 extended

 VLAN Name                                  Encap            Ports
 ---- ---------------------------------- ---------------- --------------------
 57   mg-cisco-live:pbr-one-arm-          vlan-1417        Eth1/43
      CL2023ctxv1:provider:
```

FD_VLAN

Access Encap VLAN

PBR Packet Walk – Step 5 (Return from FW)



ELAM trigger Leaf-2 in vsh_lc

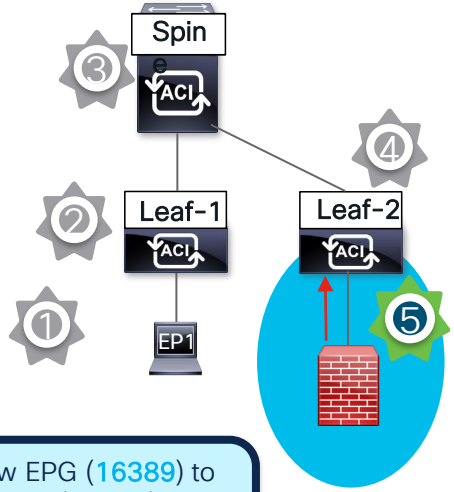
```
debug platform internal roc elam asic 0
    trigger init in-select 6 out-select 1
      reset
      set outer ipv4 src_ip 192.168.1.100 dst_ip 192.168.2.100
      set outer l2 src_mac <FW-MAC>
      start
```

policy lookup
made
mplicit
ermit)

Service Device sends traffic back to router MAC (19:FF). Destination IP is EP2 and policy lookup is made

5 Command Line Verification

Traffic is sent to 1-ARM FW device. After inspection, traffic comes back to Leaf-2 via Service EPG VLAN




```
a1-leaf2# show system internal epm endpoint mac 0050.56a8.4897
| egrep "VRF vnid|sclass"
BD vnid : 16744311 ::: VRF vnid : 2293762
Flags : 0x80004c04 ::: sclass : 16389
```

Source mac (FW)
pcTAG of service EPG

```
a1-leaf1# show system internal epm endpoint ip 192.168.2.100 |
egrep "VRF vnid|sclass "
BD vnid : 15826920 ::: VRF vnid : 2293762
Flags : 0x80000c80 ::: sclass : 16386
```

pcTAG of dest EPG
(provider of PBR flow)

Shadow EPG (16389) to provider (16386) is implicitly allowed (default filter) by service graph

```
a1-leaf2# show zoning-rule scope 2293762 src-epg 16389 dst-epg 16386
+---------+--------+--------+----------+---------+---------+---------+------+--------+----------------+
| Rule ID | SrcEPG | DstEPG | FilterID |   Dir   |  operSt |  Scope  | Name | Action |    Priority    |
+---------+--------+--------+----------+---------+---------+---------+------+--------+----------------+
|   4152  | 16389  | 16386  | default  | uni-dir | enabled | 2293762 |      | permit | src_dst_any(9) |
+---------+--------+--------+----------+---------+---------+---------+------+--------+----------------+
```

```
a1-leaf2# contract_parser.py --vrf mg-cisco-live:v1 --depg tn-mg-cisco-live/ap-a1/epg-e2

[9:4152] [vrf:mg-cisco-live:v1] permit any tn-mg-cisco-live/G-pbr-one-arm-CL2023ctxv1/C-provider(16389) tn-mg-
cisco-live/ap-a1/epg-e2(16386) [contract:uni/tn-mg-cisco-live/brc-ip] [hit=172]
```

Contract hit

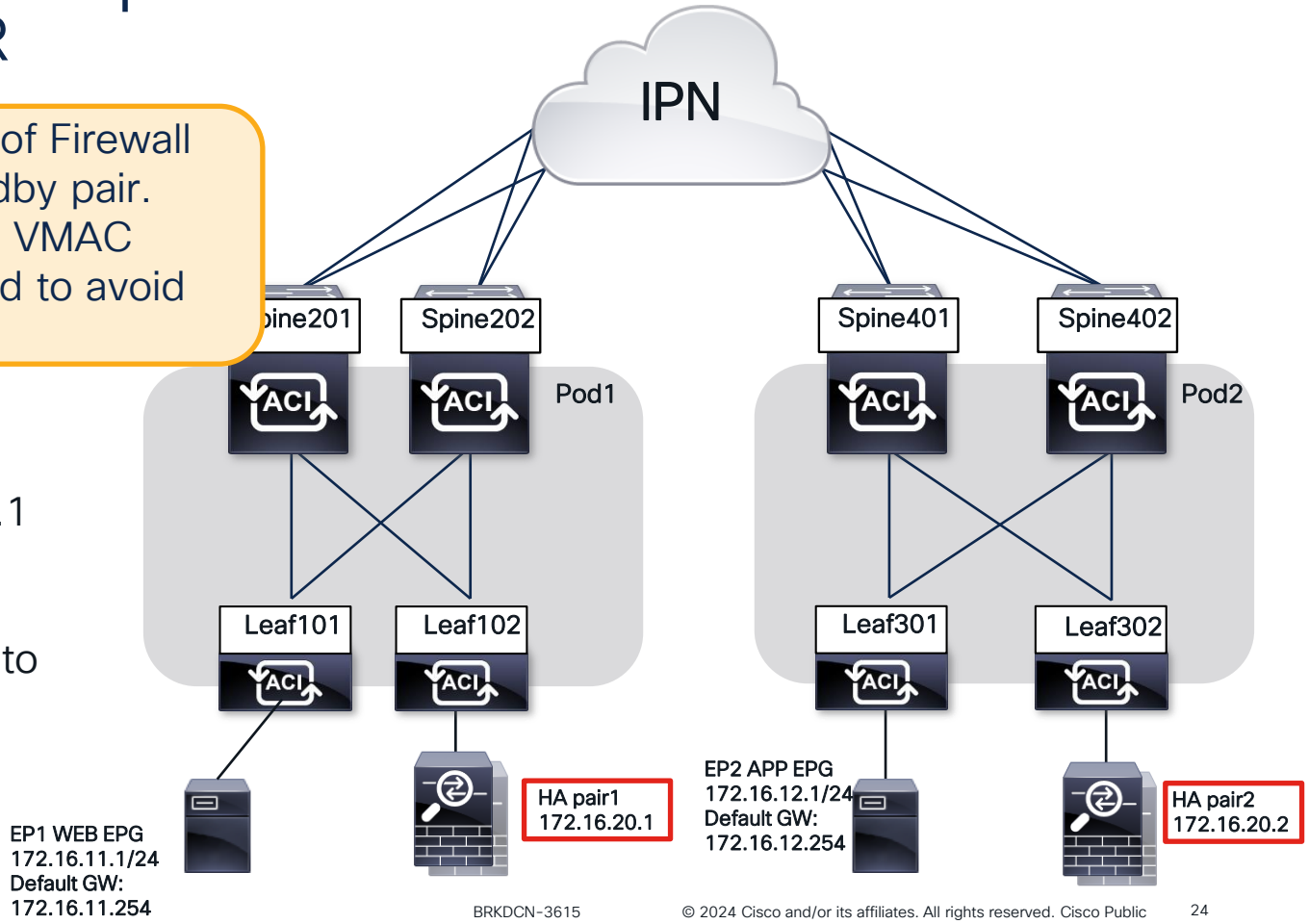
*NOTE FOR RETURN TRAFFIC: EP2->EP1 PBR flow is the same

Multipod PBR : East-West Symmetric PBR

Topology – Multipod East-West
Symmetric PBR

Each Pod have a cluster of Firewall independent active/standby pair.
Each pair have a VIP and VMAC
Symmetric PBR is needed to avoid asymmetric traffic

Routed flow between 172.16.11.1 to 172.16.12.1
Redirected to one the Firewall HA pair
FW are one-arm attached to ACI



Config Gotcha - Redirect policy

Only to be considered in North-South PBR scenario

Create L4-L7 Policy-Based Redirect

Name: REDIRECT-HA

Description: optional

Destination Type: L1 | L2 | L3

Rewrite source MAC: ☐

IP SLA Monitoring Policy: select an option

Enable Pod ID Aware Redirection: ☐

Hashing Algorithm: Destination IP | Source IP | Source IP, Destination IP and Protocol number

Enable Anycast: ☐

Resilient Hashing Enabled: ☐

L3 Destinations:

IP	Destination Name	MAC	Redirect Health Group	Additional IPv4/IPv6	Description	Oper Status
172.16.20.1		00:ea:bd:07:3d:...				Enabl...
172.16.20.2		50:2f:a8:cb:9b:...				Enabl...

Information on which we hash
Note – hash is symmetric :
Hash(A to B proto X) ==
Hash(B to A proto X)

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

PBR dest IP and MAC (MAC can be omitted in 5.2 with PBR tracking

Config Gotcha – L4/L7 devices for Symmetric PBR)

Cluster interface contains path to both HA Pair of firewall (one arm – one cluster interface

2 physical devices (one per pod)

L4-L7 Devices - FW-HA

Policy | Faults | History

General

Name: FW-HA
Alias:
Service Type: Firewall
Device Type: PHYSICAL
Physical Domain: phys
Promiscuous Mode:
Context Aware: Multiple | Single
Function Type: GoThrough | GoTo | L1 | L2

Devices

Name	Interfaces
HA-PAIR1	HA-PAIR1 (Pod-1/Node-102/eth1/19)
HA-PAIR2	HA-PAIR2 (Pod-2/Node-302/eth1/19)

Cluster Interfaces:

Name	Concrete Interfaces	Encap
LIF-FW-HA	HA-PAIR1/[HA-PAIR1], HA-PAIR2/[HA-PAIR2]	vlan-720

Packet – symmetric PBR

Packet walk Consumer WEB to Provider APP Epg

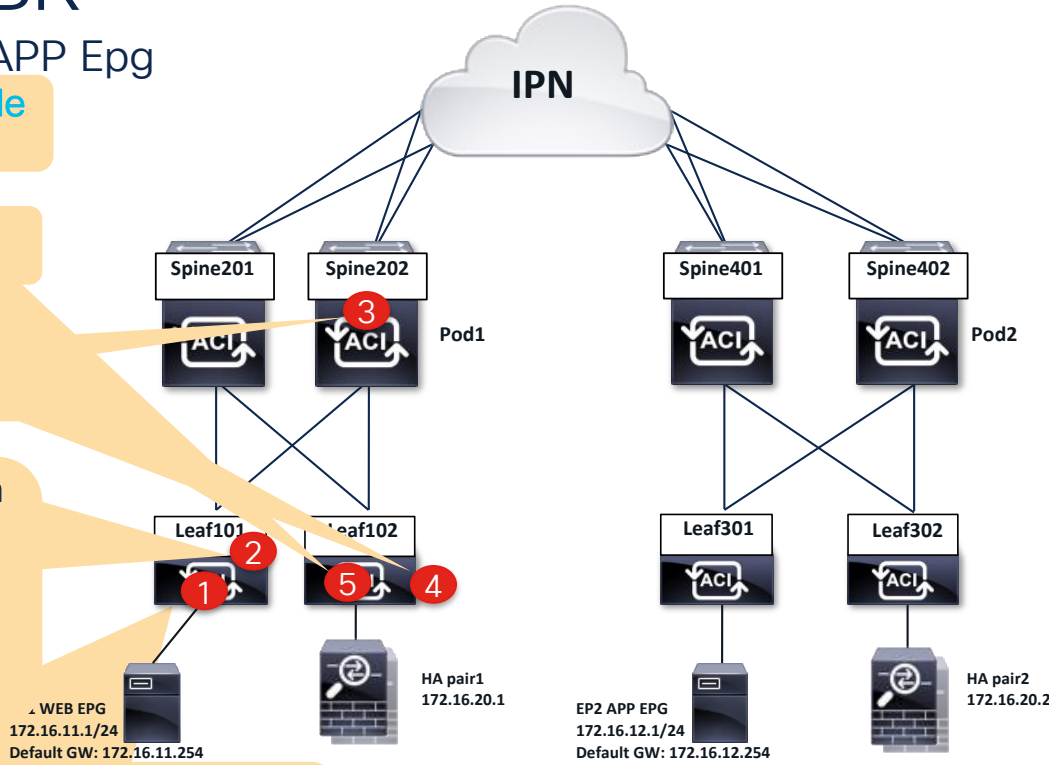
5 Return from firewall on service leaf hits permit rule to egress leaf

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

3 COOP lookup in BD VNID for Redirect mac and will send it toward service leaf

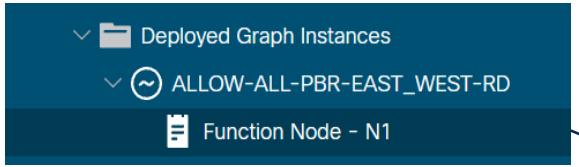
2 Leaf doing the redirect (101 or 301) is doing hash on the packet to choose one Fw cluster pair
DMAC is rewritten to Firewall Pair1 or Pair 2 (hash)
No Mac lookup happening on leaf.
Packet is encapsulated to Service BD VNID and send to vxlan tunnel to anycast-mac on spine.

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



Check 1 – Is the Graph deployed

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



Function Node - N1

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

Properties

Name: N1
Function Type: GoTo
Devices: FW-HA
Cluster Interfaces:

Name	Concrete Interfaces
LIF-FW-HA	HA-PAIR1/[HA-PAIR1], HA-PAIR2/[HA-PAIR2]

Function Connectors:

Name	Encap	Class ID	L3OutPBR Service pcTag
consumer	vlan-720	49157	any
provider	vlan-720	49157	any

If not deployed, usually it is Contract (no cons or prov)
If deployed but there are some fault (aka graph rendering failure), it is usually config related

Check 2 – Is the Service EPG deployed

```
Leaf102# show vlan encap-id 720

 VLAN Name                                Status    Ports
 ---- -------------------------------- --------- ----------
-
 15   RD-MPOD:FW-HActxRD:LIF-FW-HA:      active    Eth1/20
..

Leaf102# show system internal epm vlan 15 detail

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 the defined encap vlan-720.
- Service VLAN is deployed on the service leafs and is using the correct service EPG pcTag (sclass 49157).
- The VLAN is marked as a service EPG.
  - Service EPG do have ip dataplane learning disable

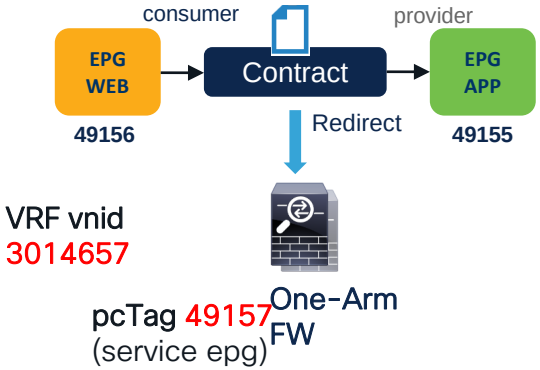
Check 3 – Zoning-rules

Summary | Dashboard | Policy | Operational | Stats | Health | Faults | History

Endpoints | Flows | Packets | Policy Tags | Resource IDs

Bridge Domains | VRFs | EPGs | ESGs | L3Outs | External Networks (Bridged)

Take note of all vnid and sclass involved



Expected zoning-rules:

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

```
Leaf101# show zoning-rule scope 3014657
+---------+--------+--------+----------+---------+---------+----------------+---------------+
| Rule ID | SrcEPG | DstEPG | FilterID |  operSt |  Scope  |      Action    | Priority      |
+---------+--------+--------+----------+---------+---------+----------------+---------------+
| 4 4128  | 49157  | 49156  | default  | enabled | 3014657 |     permit     | src_dst_any(9)|
| 3 4190  | 49155  | 49156  |   11     | enabled | 3014657 |redir(destgrp-1)| src_dst_any(9)|
| 2 4191  | 49157  | 49155  | default  | enabled | 3014657 |     permit     | src_dst_any(9)|
| 1 4189  | 49156  | 49155  |   11     | enabled | 3014657 |redir(destgrp-1)| src_dst_any(9)|
+---------+--------+--------+----------+---------+---------+----------------+---------------+
```

Check 4 - Redirect info

Redir group should have the VIP of each HA pair

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

```
Leaf101# show service redir info group 1
==========================================================================================
GrpID Name            destination                                   op

1     destgrp-1       dest-[172.16.20.2]-[vxlan-3014657]            er
                      dest-[172.16.20.1]-[vxlan-3014657]


Leaf101# show service redir info destination ip 172.16.20.2 vnid 30
==========================================================================================
Name                                      bdVnid          vMac                vrf
====                                      ======          ====                ====
dest-[172.16.20.2]-[vxlan-3014657]        vxlan-14843887  50:2F:A8:CB:9B:3C   RD-MPOD:RD

Leaf101# 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 5 – Check load balancing

for a given flow (in vsh_lc mode)

172.16.11.1 to 172.16.12.1 using TCP hash to HA pair with VIP 172.16.20.2

```
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 using ICMP hash to HA pair with VIP 172.16.20.1

```
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
```

Datapath Troubleshooting Tool:

ftriage from APIC CLI

Before service device

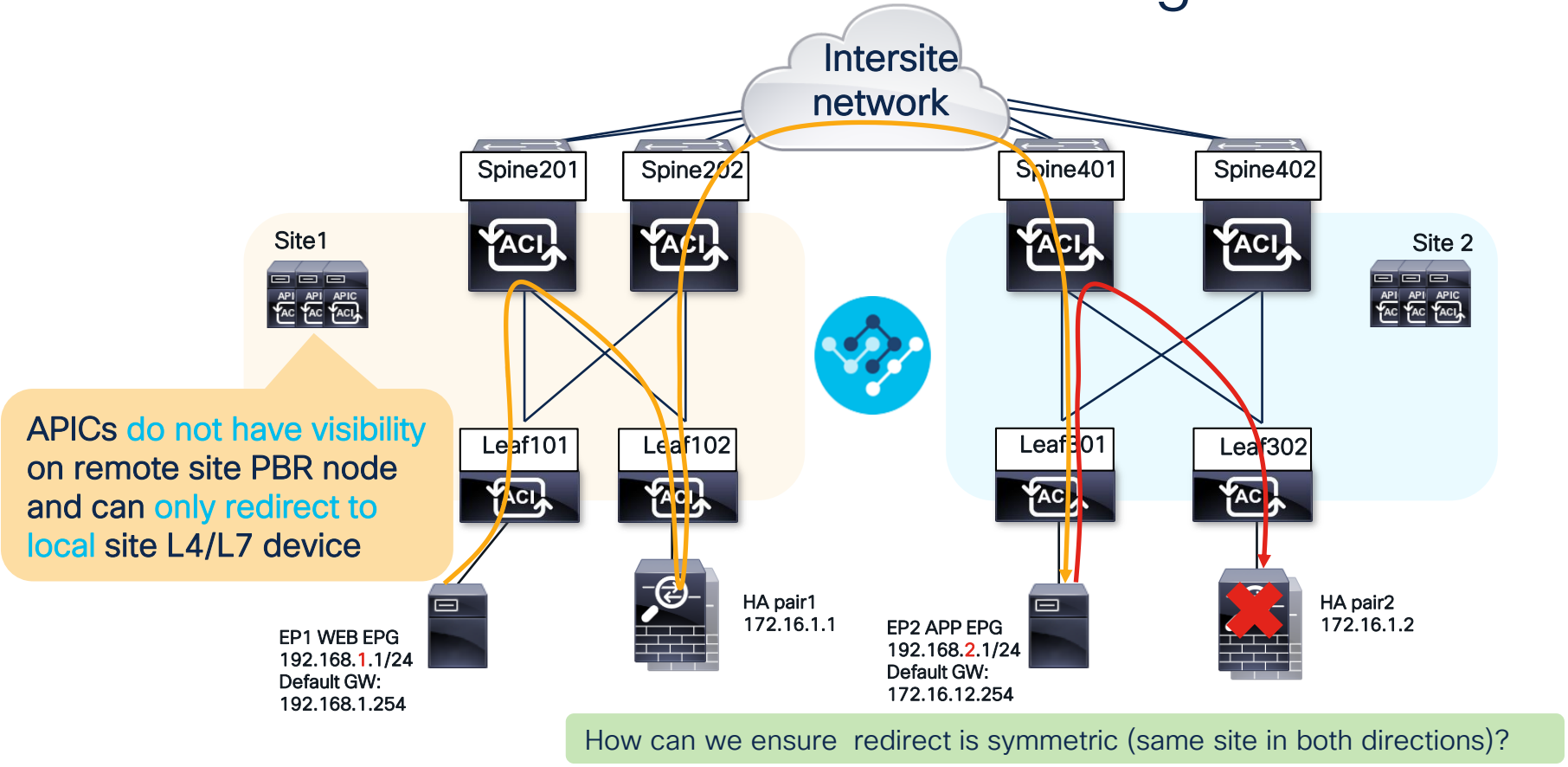
```
Apic1# ftriage route -ii LEAF:101 -sip 172.16.11.2 -dip 172.16.12.2
2023-01-27 08:28:41,179 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf101 Ingress: Eth1/11 Egress: Eth1/49  Vnid: 14909416
2023-10-27 08:29:27,042 INFO     ftriage:  unicast:1543 S1P1-Leaf101: traffic is redirected to vnid:14843887 mac:00:EA:BD:07:3D:7C via tenant:RD-
    MPOD graph:EAST_WEST contract: ALLOW-ALL-PBR
2023-01-27 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]
2023-01-27 08:31:28,056 INFO     ftriage:  unicast:2196 S1P1-Spine201: EP is known in COOP (DIPo = 10.0.0.67)
2023-01-27 08:31:41,494 INFO     ftriage:     main:958  Found peer-node S1P1-Leaf102 and IF: Eth1/49 in candidate list
2023-01-27 08:31:51,918 INFO     ftriage:       ep:128  S1P1-Leaf102: pbr traffic with dmac: 00:EA:BD:07:3D:7C
2023-01-27 08:32:06,748 INFO     ftriage:     main:1796 Packet is Exiting fabric with peer-device: POD1-router1 and peer-port: Ethernet1/19
2023-01-27 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]
2023-01-27 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
2023-01-27 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

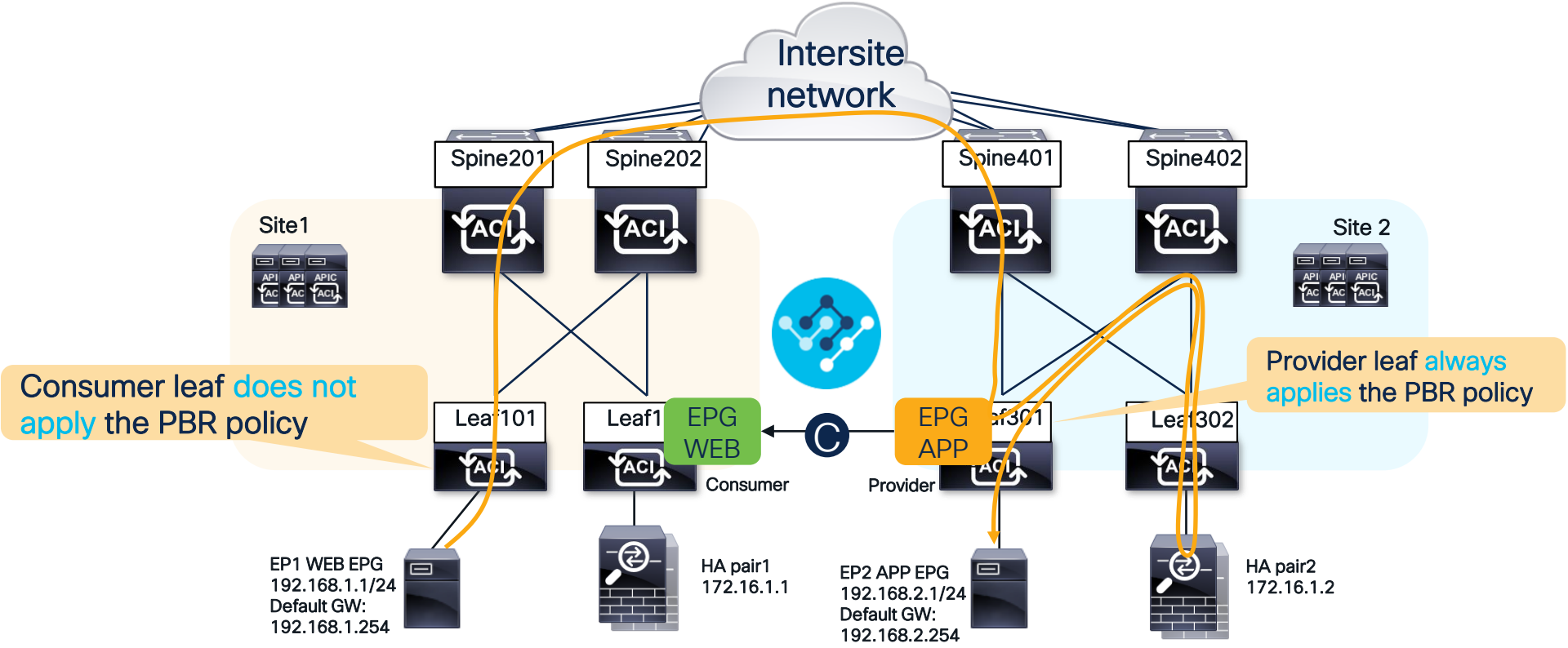
```
2023-01-27 08:32:21,224 INFO     ftriage:     main:1821 pbr return path, nxt_nifs {S1P1-Leaf102: ['Eth1/19']}, nxt_dbg_f_n ig, nxt_inst ig, eg_ifs
    Eth1/19, Vnid: 720
2023-01-27 08:32:33,581 INFO     ftriage:     main:1295 L3 packet Seen on  S1P1-Leaf102 Ingress: Eth1/19 Egress: Eth1/49  Vnid: 3014657
2023-01-27 08:33:14,060 INFO     ftriage:     main:958  Found peer-node S1P1-Spine201 and IF: Eth1/2 in candidate list
```

Multi-Site East-West PBR

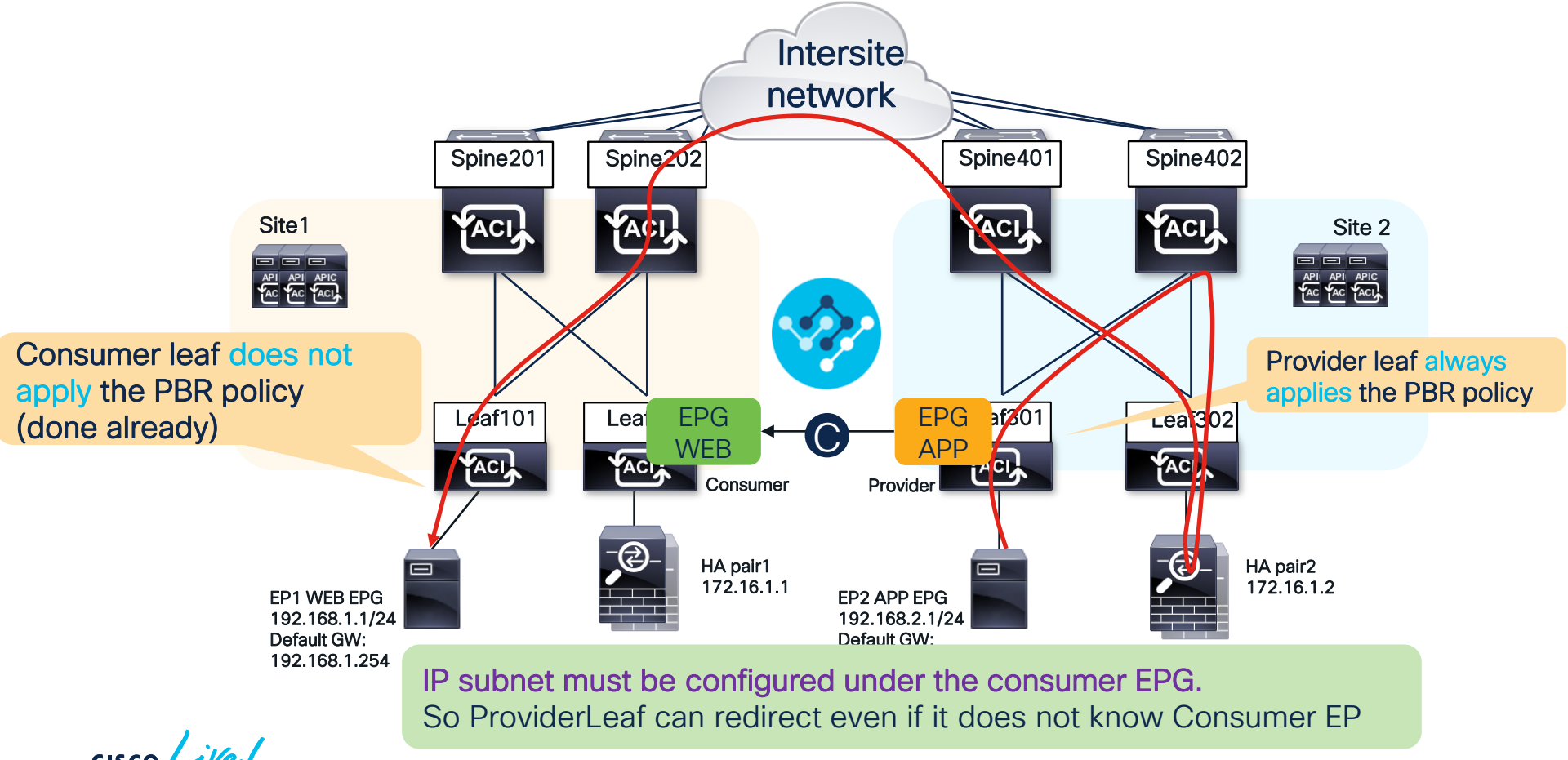
Multi-Site PBR – East-West Challenge



Multi-Site PBR – East-West Consumer to Provider



Multi-Site PBR – East -West Provider to Consumer



Config Gotcha Multi-Site PBR – East-West

RD

RD

TEMPLATES

BothSite
ACI

Site1
ACI

Site2
ACI

SITES

Site1 (ACI) 5.2(5c)

BothSite

Site1

Site2 (ACI) 5.2(7f)

BothSite

Site2

BothSite Version 21
Applied to 2 sites
Tenant: RD

FILTERS

Application Profile App

EPGs

APP

WEB

Contracts

EPG WEB

USED IN CURRENT TEMPLATE 1 | USED BY OTHER TEMPLATES 0

Common Properties

Display Name*
WEB

Description

Contracts

Name

To-FW
Type: consumer

Add Contract

Properties

On-Premises Properties

Bridge Domain *
BD-APP

Subnets

Gateway IP

192.168.1.1/24
Primary: No

EPG APP

USED IN CURRENT TEMPLATE 1 | USED BY OTHER TEMPLATES 0

Common Properties

Display Name*
APP

Description

Contracts

Name

To-FW
Type: provider

Add Contract

Properties

On-Premises Properties

Bridge Domain *
BD-WEB

Subnets

Gateway IP

Add Subnet

EPG WEB is the consumer of the contract and Subnet is under EPG
EPG APP is provider of the contract and subnet does not need to be under the EPG

Consumer to Provider
Ingress Consumer leaf zoning-rule – site1

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

```
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)|
+---------+------+------+----------+---------+---------+-------------------------------------------+


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                                  bdVnid          vMac                 vrf     operSt
====                                  ======          ====                 ====    =====
dest-[172.16.1                        an-16187319     00:EA:BD:07:3D:7C    RD:RD   enabled
```

Only local PBR is available

Multi-Site PBR – East-West

Limitation and requirement

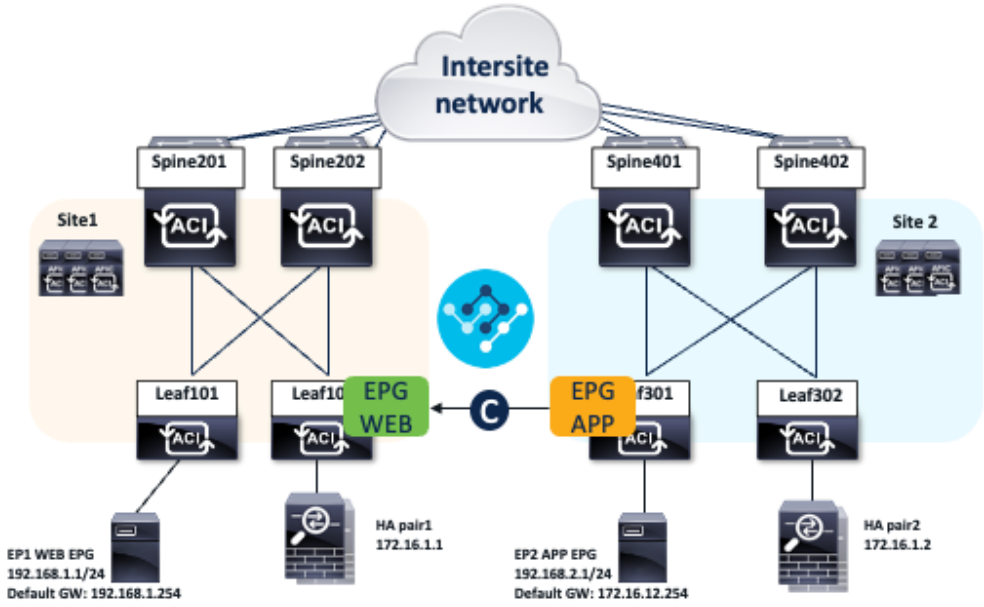
- Redirect only to site local PBR devices
- We need to ensure both direction flow through same firewall

Implementation :

- Policy redirect always applied in site where provider is → How to implement Any to Any PBR ?

Implementation :

- Consumer subnet must be configured under the consumer EPG → How to implement App centric ?



Multi-Site
vzAny to vzAny PBR

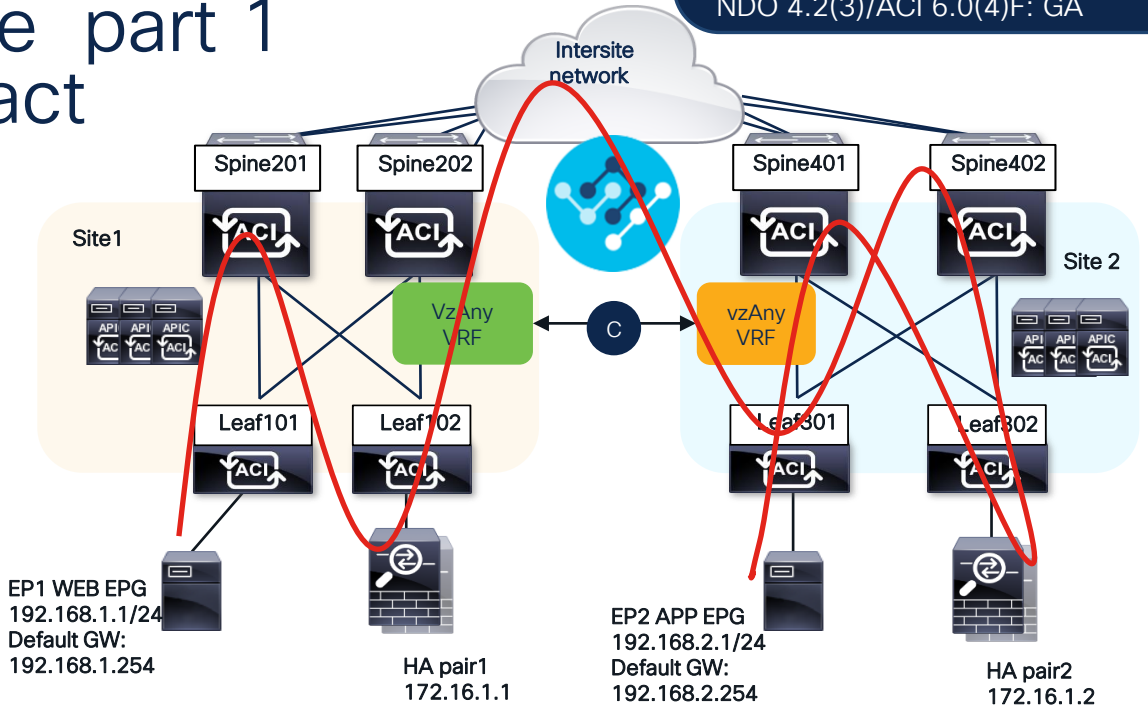
General Availability
Apic 6.0(4) NDO – 4.2(3)

Fixing the Challenge part 1 Directionless contract

NDO 4.2(1)/ACI 6.0(3)F: Beta
NDO 4.2(3)/ACI 6.0(4)F: GA

How to implement Any to Any PBR (no identified Provider and Consumer) and keep traffic symmetry across Firewall ?

Ensure all traffic goes to Firewall on both side !



East – West – Any to Any – Known EP

VNID/sclass translate

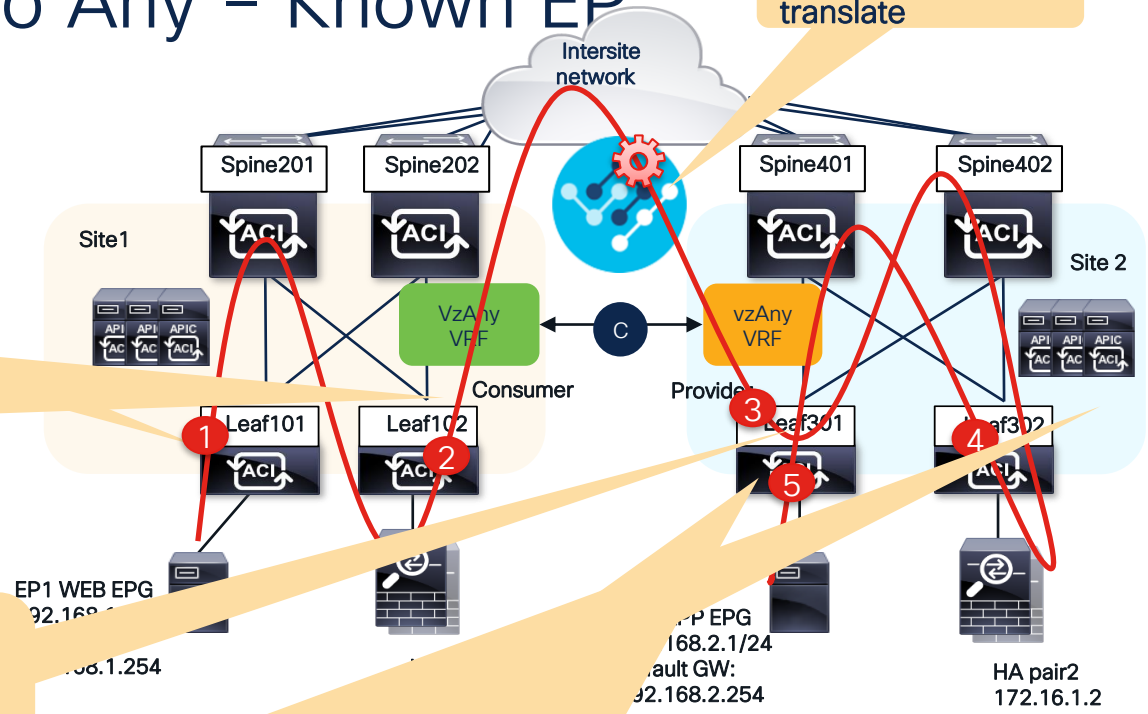
1 Traffic from Client to Server
Assuming Dest IP is known
Ingress server leaf apply Redirect to HA pair1 (rewrite Dmac to FW – BD vnid to anycast MAC spine)
Redirect Rule is any to any

2 Packet back from FW site 1.
Sclass = Serv EPG site 1 – dclass any
Hits a new rule "permit handoff" mark
Packet with Sp=0/Dp=1 (signal Src applies policy but dest should still do it)
Send to site 2

3 Destination leaf
Sclass Serv EPG → dclass EPG2, bypass regular rule to permit from serv EPG if coming from site 1 (Outer SIP)with Sp=0/Dp=1 and
Hits a redirect rule to HA pair 2

4 Back from FW HA pair 2
Leaf 302 applies permit from service EPG to App EPG if ep is known or proxy with no policy applied if unknown

5 Leaf 301 gets it from second time but Outer SIP is not Site1
Hence it hits "regular" rule from Service EPG to EPG2 permit (unless polict was applied by 302)



Zoning-Rule

1 Leaf101 ingress site redirect to FW site 1

RuleID	SrcEPG	DstEPG	FilterID	Dir	operSt	Scope	Action
4213	0	0	5	uni-dir	enabled	3112963	redir(destgrp-3)

2 Site 1 permit handoff to let know Site2 it also needs to redirect

Rule ID	SrcEPG	DstEPG	FilterID	Dir	operSt	Scope	Action
4198	5481	0	5	uni-dir	enabled	3112963	permit_handoff

5 Site 2 leaf 301 gets packet in step 3 and 5

Rule ID	SrcEPG	DstEPG	FilterID	Dir	operSt	Scope	Action
4182	10934	0	5	uni-dir	enabled	2883584	permit_handoff

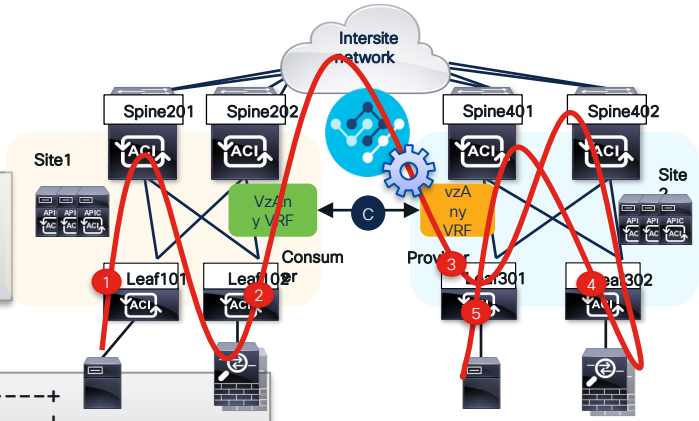
3

```
Site2-Leaf301# show service redir info aclrule
AclRuleVnid      DestGroup        OperSt
vxlan-2883584    destgrp-2        enabled
S2P1-Leaf102# show service redir info group 2
destgrp-2        dest-[172.16.1.2]-[vxlan-2883584]
```

Rule ignored in step 3
Used in Step 5
Outer Source IP being Site 1
TEP makes the difference

4 Site2 leaf 302 after FW permit handoff and go to final leaf 301

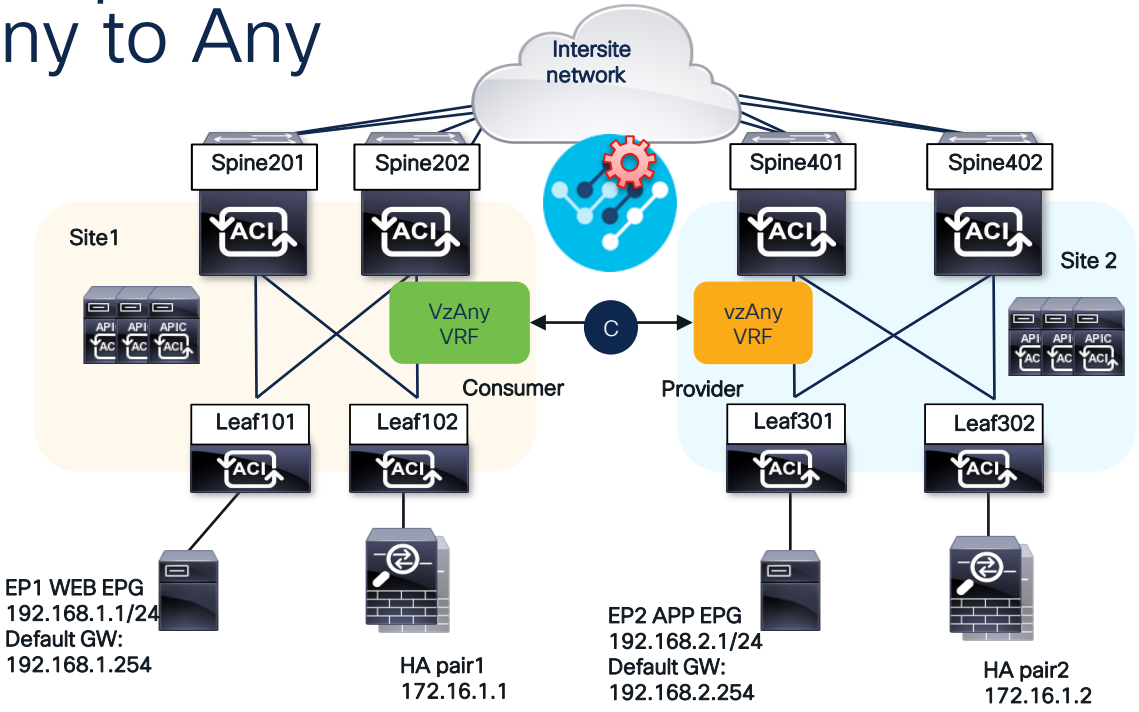
Rule ID	SrcEPG	DstEPG	FilterID	Dir	operSt	Scope	Action
4182	10934	0	5	uni-dir	enabled	2883584	permit_handoff



Fixing the Challenge part 2
App Centric PBR Any to Any

How to implement App Centric (aka no EPG subnet, but only BD subnet)

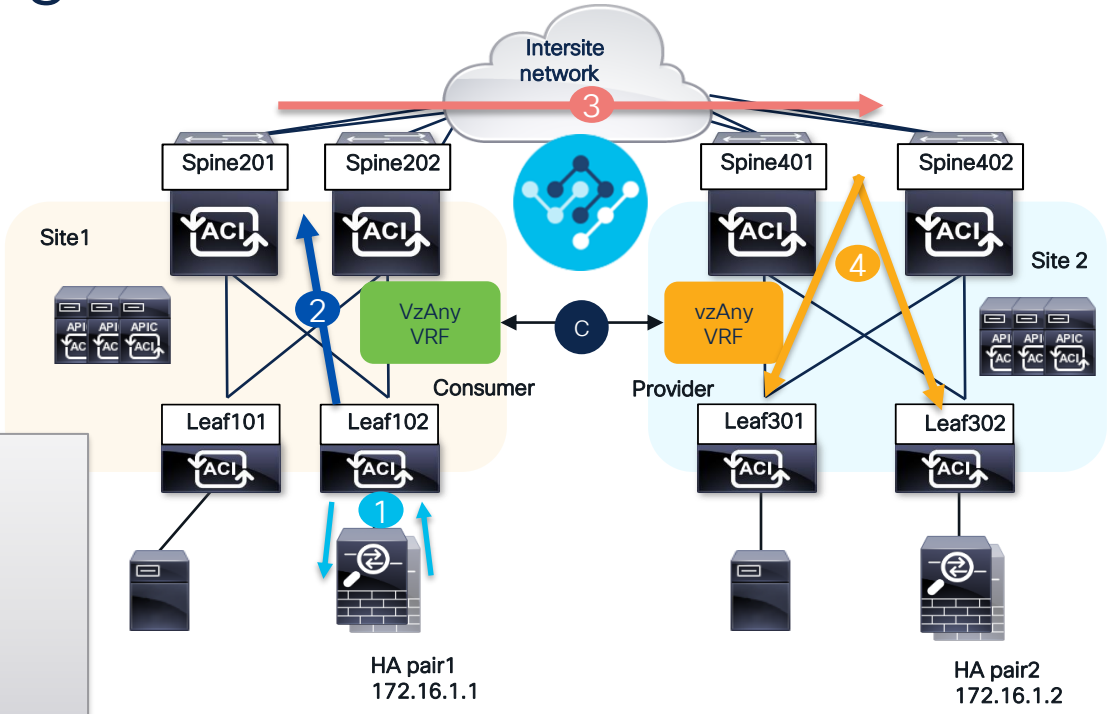
1. Ensure *cross site PBR node visibility* and availability
2. *Capability to trombone traffic during learning* (redirect from site 2 to site 1 FW) – **site aware zoning-rule**
3. Ensure ingress knows Dest EP ASAP (*forced control plane learning*) to limit tromboning



Multisite PBR tracking

1. Each Service leaf use local SLA monitoring to track PBR node
2. Service leaf informs spine about tracked service
3. Spine to spine update to inform other sites

```
Site2-Spine1# show system internal slamon tracked-
services  msite rx detail
SLAMON TRACKED SERVICE MSITE:
-------------------------------------------
IP address                                  : 172.16.1.1
VRF VNID                                    : 2883584
State                                       : Up
Site ID                                     : 1
BD-VNID                                     : 15171529
Monitoring TEP IP                           : 172.16.1.4
```



4. Spine to leaf to inform leaf in site 2 about PBR devices of site 1

Allows a site to redirect to PBR devices in remote site (site-aware zoning-rule)

`Show zoning-rule scope <VRF-VNID> site X`

Any to Any PBR – App Centric – Dataplane Tromboning

1

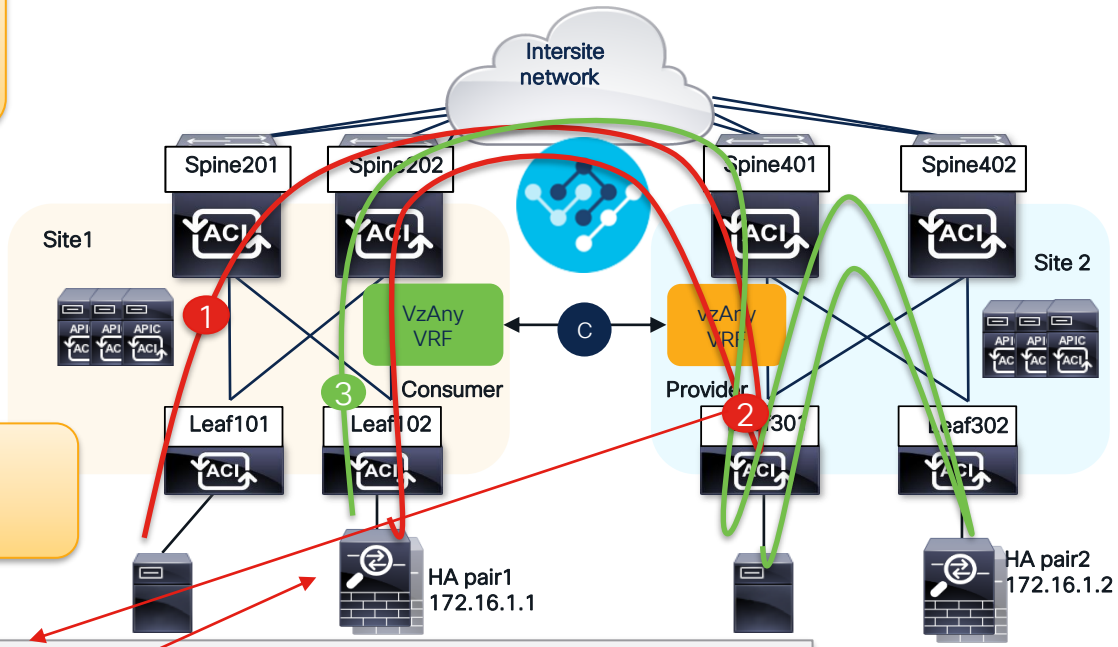
EP unknown in site 1 leaf
→ Forwarded to Site 2
→ Site 2 allow redirect to Site 1 Fw

3

From site 1 Firewall similar to previous example

2

Site 2 leaf – site aware zoning-rule
Redirecting to site 1 FW AND
Punt to CPU to force ingress learning




```
Site2-Leaf301# show zoning-rule scope 2883584 site-id 1
+---------+--------+--------+----------+---------+---------+---------+-------------------------------+
| Rule ID | SrcEPG | DstEPG | FilterID |   Dir   | operSt  |  Scope  |       Action                  |
+---------+--------+--------+----------+---------+---------+---------+-------------------------------+
|   4216  |   0    |   0    |    5     | uni-dir | enabled | 2883584 | punt_to_learn,redir(destgrp-4) |
Site2-LEaf301# show service redir info group 4
GrpID Name            destination
===== ====            ===========
4     destgrp-4       dest-[172.16.1.1]-[vxlan-2883584]
```

Any to Any PBR – App Centric – Force learning

1

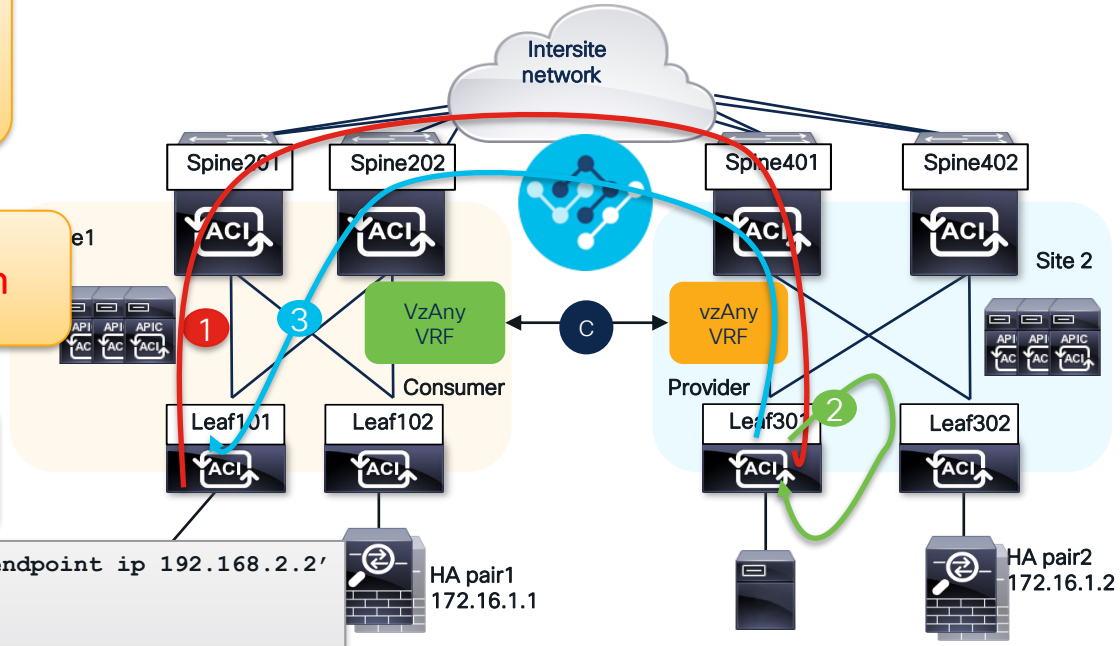
EP unknown in site 1 leaf
→ Forwarded to Site 2
→ Site 2 allow redirect to Site 1 Fw

2

Site 2 leaf – site aware zoning-rule
Redirecting to site 1 FW AND punt_to_learn
Punt to CPU to force ingress learning

3

Control plane packet from Site 2 to Site 1 to force learning in site 1 leaf (stop tromboning)




```
Site1-Leaf101# vsh_lc -c 'show system internal epm endpoint ip 192.168.2.2'
.
IP# 0 : 192.168.2.1
VRF name : DC:DC::: VRF vnid : 3112963
phy if : 0 ::: tunnel if : 0x1801000c ::: Interface : Tunnel12
Ref count : 3 ::: sclass : 32772
::: Learns Src: EPM
Times EP Skipped Sync : 1
EP Flags : IP|timer|control-ep|
Aging: Timer-type : control-ep ::: Timeout-left : 86284 ::
```

Control plane learned entry in site 1 leaf (24 hours timeout)

Summary

Summary of zoning-rule action used

Action Name	Purpose	Scenario used
permit	Regular permit and set policy applied bit	Usual contract – simple PBR from service EPG to destination
redirect	Redirect to a PBR destination group (hash if multiple)	Cons to Prov and Prov to Cons in most PBR scenario
redir_override	Allow to bypass redirect and just permit without setting policy applied bit	Used together with redirect in multisite PBR scenario. Used only if destination EP is not local
permit_handoff	Permit and set dest policy applied bit and not src policy applied bit	Used from service epg to destination in case of Any to Any PBR in multisite
punt_to_learn	Punt to cpu to force learning of the source across the site and limit tromboning	Used in multisite Any to Any PBR together with redirect
deny	Deny the packet	Not PBR related

More info ?

- ACI PBR white paper : https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739971.html
- PBR Any to Any in Multisite : https://www.cisco.com/c/en/us/td/docs/dcn/ndo/4x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-421/ndo-configuration-aci-use-case-vzany-pbr-42x.html

Thank you

cisco Live!

Let's go