



ACI Multisite

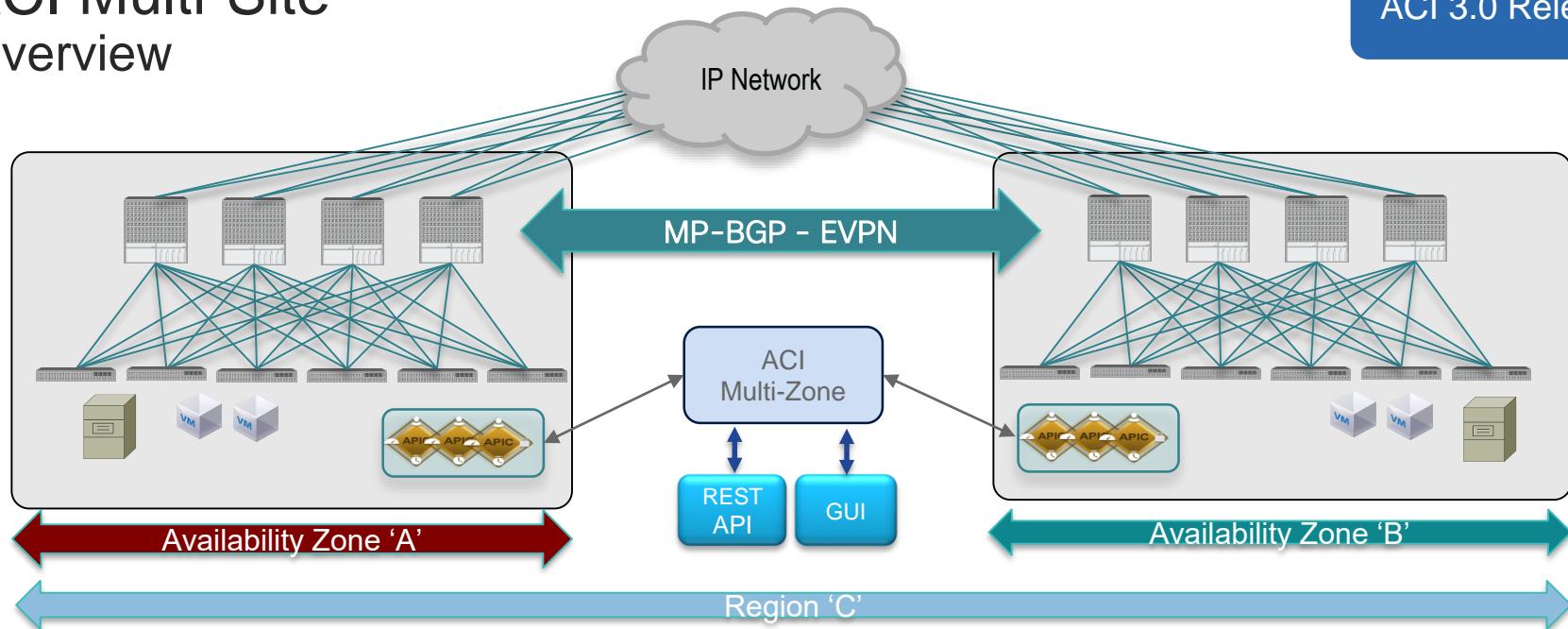
V1.1

Roland Ducomble - CCIE 3745
EMEAR ACI solution Tac Team - Technical Leader
24th May 2018

Overview

ACI Multi-Site Overview

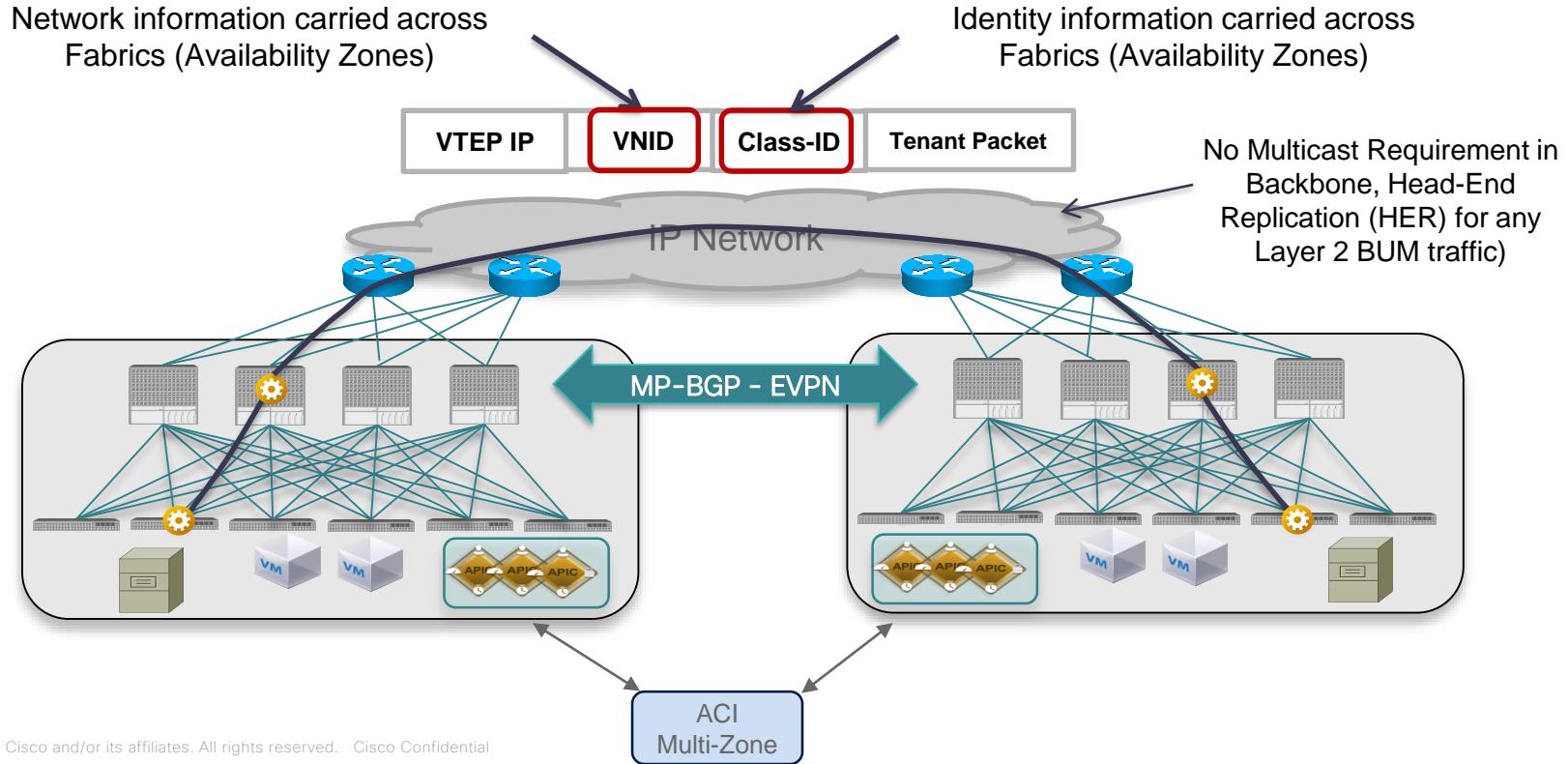
ACI 3.0 Release



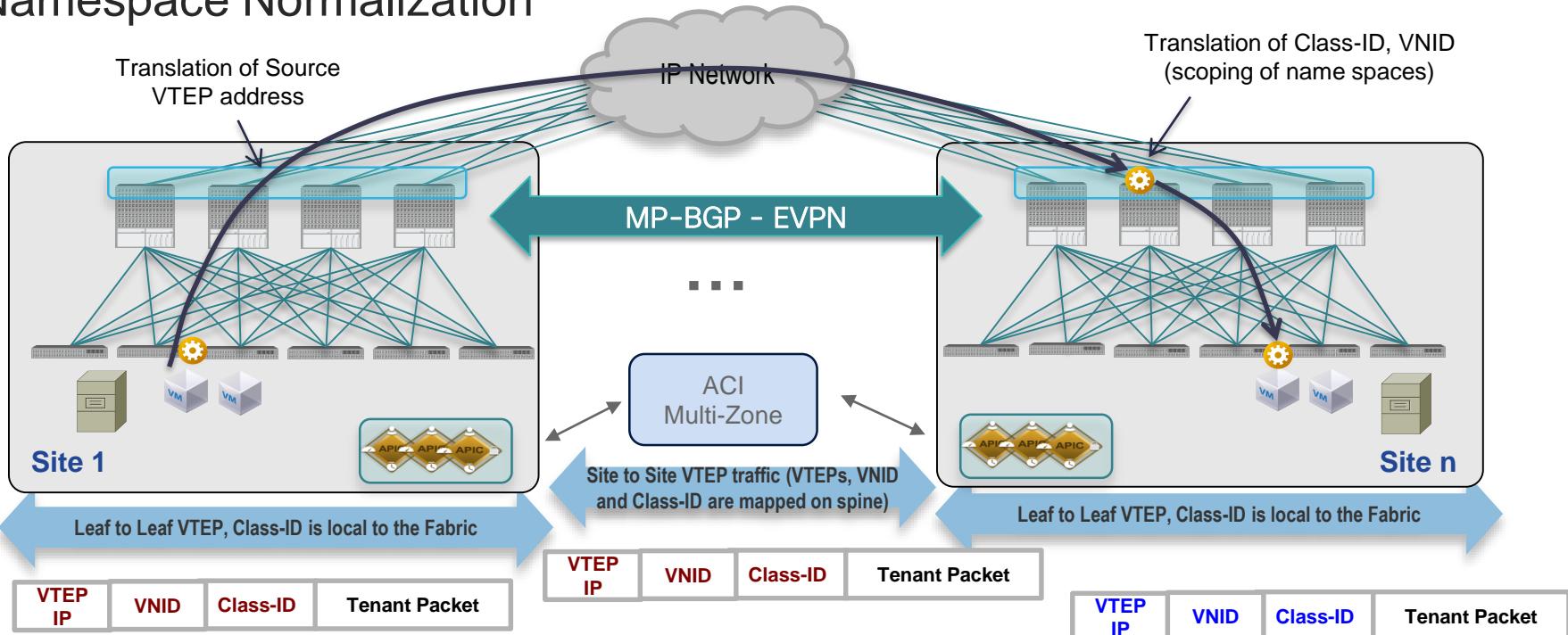
- Separate ACI Fabrics with independent APIC clusters
- ACI Multi-Zone pushes cross-fabric configuration to multiple APIC clusters providing scoping of all configuration changes
- MP-BGP EVPN control plane between sites
- Data Plane VXLAN encapsulation across sites
- End-to-end policy definition and enforcement

ACI Multi-Site

Network and Identity Extended between Fabrics



ACI Multi-Site Namespace Normalization

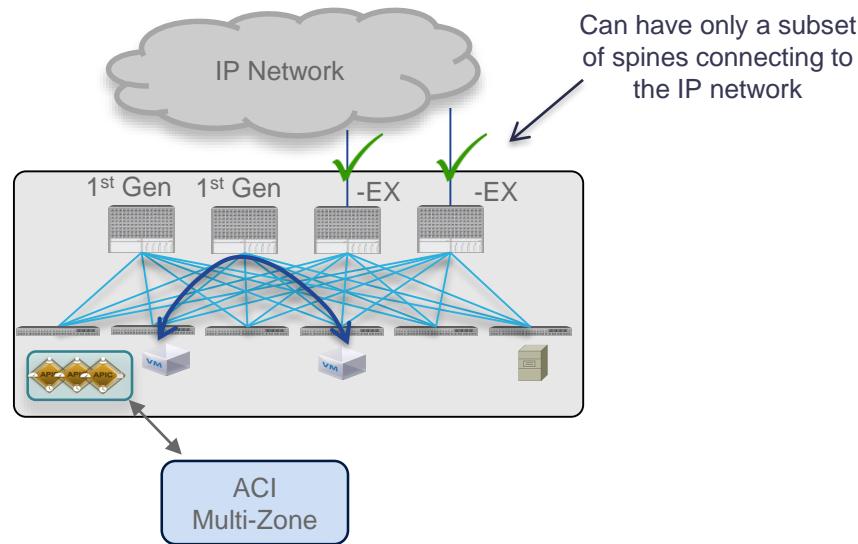


- Maintain separate name spaces with ID translation performed on the spine nodes
- Requires specific HW on the spine to support for this functionality

ACI Multi-Site

Hardware Requirements

- Support all ACI leaf switches (1st Generation, -EX and -FX)
- Only -EX spine nodes (or newer) to connect to the inter-site network
- New FX non modular spine (9364C, 64x40G/100G ports) will be supported for Multi-Site in Q1CY18 timeframe
- 1st generation spines (including 9336PQ) not supported
Can still leverage those for intra-site leaf to leaf communication



Multisite Day 0 Setup

Setup MultiSite Controller (MSC)

(MultiSite Policy Manager)

Follow the guide ☺

- [https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/aci_multi-site/sw/1x/installation/b Cisco ACI Multi-Site Installation Guide/b Cisco ACI Multi-Site Installation Guide chapter 010.html](https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/aci_multi-site/sw/1x/installation/b_Cisco_ACI_Multi-Site_Installation_Guide/b_Cisco_ACI_Multi-Site_Installation_Guide_chapter_010.html)

VM initial Setup

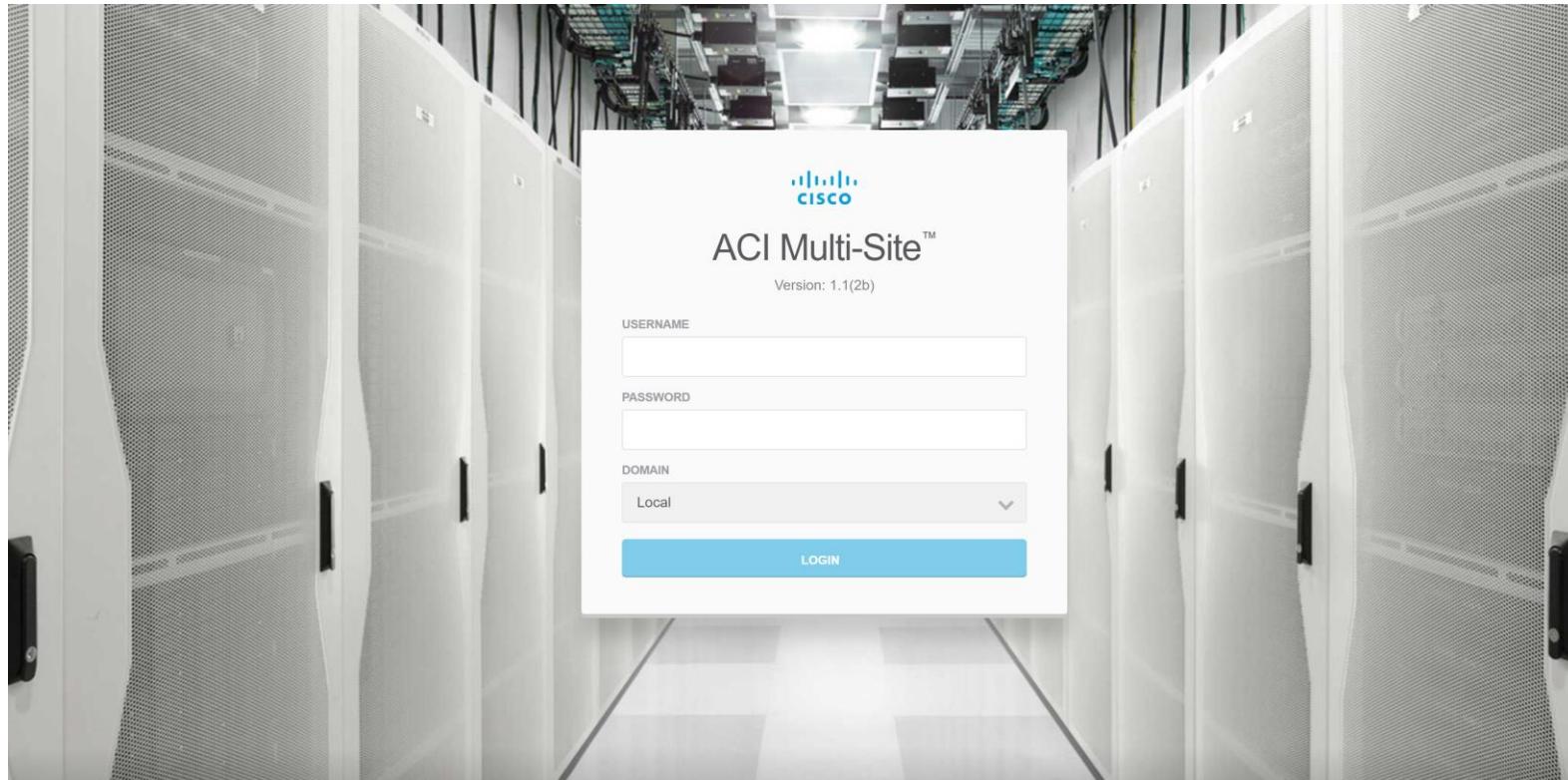
- Deploy 3 VM with the MSC OVA using Vcenter. Ideally each VM should be on different ESX host
- Each VM is a Centos7.3 VM
- When booted fill initial Setup
 - Set the hostname of each instances as “node1”, “node2”, “node3” (nothing else) !!
 - Ensure that the virtual machines are able to ping each other
- Then Follow the instruction in the config guide (lightly differ depending on MSC version)

Cluster Redundancy

- From : https://docs.docker.com/engine/swarm/admin_guide/#add-manager-nodes-for-fault-tolerance
- You should maintain an odd number of managers in the swarm to support manager node failures. Having an odd number of managers ensures that during a network partition, there is a higher chance that the quorum remains available to process requests if the network is partitioned into two sets
- With a cluster of 3, majority is 2 and we can lose one with no impact.
- Loss of 2 VM of the cluster would lead to cluster inactive
 - --> VM should ideally be on 3 different ESX host

Multisite Day 0 Setup Infra (overlay-1) config

Connect to MSC GUI



Step 1- In MSC - Add each site one by one

Cisco ACI Multi-Site

Controller Status 3/3 | ? | Welcome, Admin

Sites

Connection Settings

* NAME:

LABELS: Select or Create a Label.

* APIC CONTROLLER URL:

+ APIC CONTROLLER URL:

* USERNAME:

* PASSWORD:

SPECIFY LOGIN DOMAIN FOR SITE: OFF

* APIC SITE ID:

CONFIGURE INFRA | ADD SITE

SITE NAME/LABEL APIC CONTROLLER URLs ACTIONS

Cisco Confidential

SITE NAME/LABEL	APIC CONTROLLER URLs	ACTIONS

Sites



SITE NAME/LABEL

APIC CONTROLLER URLs

97 POD35

<https://10.48.18.241>

98 POD36

<https://10.48.18.251>

* NAME

POD35

LABELS

Select or Create a Label.

* APIC CONTROLLER URL

<https://10.48.18.241>

APIC CONTROLLER URL

* USERNAME

admin

* PASSWORD

SPECIFY LOGIN DOMAIN FOR SITE

 OFF

* APIC SITE ID

1

97 POD35

<https://10.48.18.241>

98 POD36

<https://10.48.18.251>

* NAME

POD36

LABELS

Select or Create a Label.

* APIC CONTROLLER URL

<https://10.48.18.251>

APIC CONTROLLER URL

* USERNAME

admin

* PASSWORD

SPECIFY LOGIN DOMAIN FOR SITE

 OFF

* APIC SITE ID

2

Note. Each Site
Must have a unique SITE id

Step 2 – in each APIC cluster

- On each site in the APIC you need to create the following :
 - iBGP AS and Route Reflector
 - Spine access policies for spine uplink to IPN with an AEP and phys domain
 - **External Dataplane TEP per Site** (unique Dataplane loopback per site)
- **No Need to configure L3 out or ospf !!**

Configure Multipod External Dataplane TEP per site in Infra/Policies/Protocol/FabricExtConnProfiles

Intrasite/Intersite Profile - Fabric Ext Connection Policy pod35

Properties

Fabric ID: 1

Name:

Community:
Ex: extended:as2-nn4:5:16

Peering Type: Full Mesh Route Reflector

Password:

Confirm Password:

Pod Connection Profile

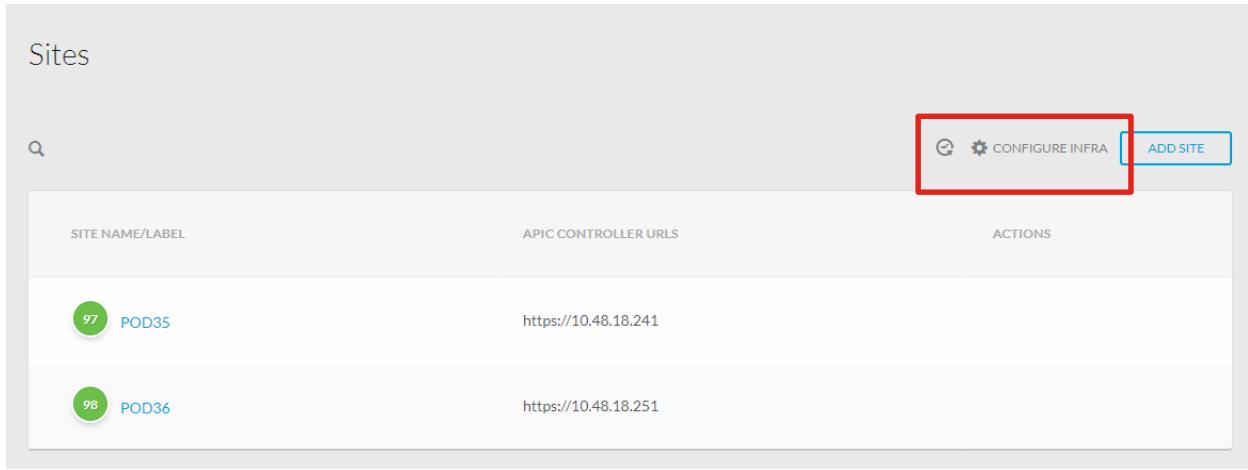
Pod ID	MultiPod Dataplane TEP
1	10.10.35.101/32

Specify
Community for
BGP and Mpod
DP TEP

Step 3 – in MSC configure infra policies per Site

- (optional) modify BGP and OSPF policies
- Per site :
 - Specify Site policies (ospf area 0, multicast TEP, ..)
 - Pod policies (Dataplane TEP set in APIC)
 - Spine policies (vlan 4 subinterface addresses and bgp RID (Control plane TEP))

Sites



The screenshot shows the 'Sites' page of the Cisco ACI Controller. At the top, there is a search bar with a magnifying glass icon. To its right are two buttons: 'CONFIGURE INFRA' (highlighted with a red box) and 'ADD SITE'. The main area displays a table with two rows. The columns are labeled 'SITE NAME/LABEL', 'APIC CONTROLLER URLs', and 'ACTIONS'. The first row contains a green circular badge with '97', the label 'POD35', and the URL 'https://10.48.18.241'. The second row contains a green circular badge with '98', the label 'POD36', and the URL 'https://10.48.18.251'.

SITE NAME/LABEL	APIC CONTROLLER URLs	ACTIONS
97 POD35	https://10.48.18.241	
98 POD36	https://10.48.18.251	

Fabric infra General settings

Fabric Connectivity Infra

SETTINGS

General Settings

SITES

- POD35
ENABLED
- POD36
DISABLED

Control Plane BGP

BGP PEERING TYPE

full-mesh

KEEPALIVE INTERVAL (SECONDS)

60

HOLD INTERVAL (SECONDS)

180

STALE INTERVAL (SECONDS)

300

GRACEFUL HELPER

ON

MAXIMUM AS LIMIT

5

BGP TTL BETWEEN PEERS

16

OSPF

OSPF POLICIES

NAME	NETWORK TYPE
msc-ospf-policy-default	point-to-point

+ ADD POLICY

Allow to modify
Bgp and ospf policies

Once done,
Select one of the site
On the left

Site 1 (Pod35) infra settings on MSC

The screenshot shows the Cisco Management Service Center (MSC) interface for managing fabric connectivity. The left sidebar has sections for SETTINGS, General Settings, SITES, and two entries: POD35 (ENABLED, highlighted in blue) and POD36 (DISABLED). The main pane displays Site 1 (POD35) with a count of 97. It shows a hierarchy: SITE > POD pod-1 > Spine pod35-spine1. A red box highlights the entire configuration pane, and another red box highlights the Spine node entry.

You can click on The Site , the Pod or the Spine,
It will allow different settings on the config pane (right pane)

Site settings

- Specify Dataplane Multicast TEP (one lo per site) used for HREP
- BGP AS (matching AS from the site configured in apic)
- OSPF area type and area id
- Ext Routed Domain

97 POD35 SETTINGS 

0 | 1 | 5 | 0

SITE IS ACI MULTI-SITE ENABLED **ON**

APIC SITE ID

DATA PLANE MULTICAST TEP

BGP AUTONOMOUS SYSTEM NUMBER

BGP PASSWORD

OSPF AREA ID

OSPF AREA TYPE ▼

EXTERNAL ROUTED DOMAIN ▼

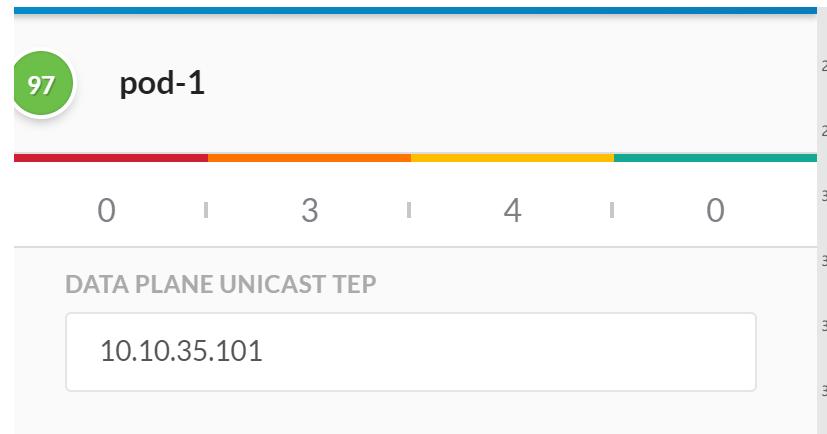
IP SUBNETS TO IMPORT

SUBNET 

 ADD SUBNET

Pod infra settings

- Specify the Dataplane TEP set in APIC in step 2



Spine Infra settings

- For each interface from spine to IPN, set IP address and mask
- Enable BGP peering
- Configure Control plane TEP (Router ID)

The screenshot shows a network configuration interface for a pod named "pod35-spine1".

PORTS

ID	IP ADDRESS/SUBNET	MTU	Actions
2/5	10.10.35.1/30	inherit	<input checked="" type="button"/>
2/6	10.10.35.5/30	inherit	<input checked="" type="button"/>

BGP PEERING

ON

CONTROL PLANE TEP

10.10.35.111

SPINE IS ROUTE REFLECTOR

OFF

Step 4 – Configure ISN (InterSite Network)with OSPF

```
vrf context IPN
router ospf 1
  vrf IPN
    router-id 10.10.35.100
```

```
interface Ethernet1/49.4
  encapsulation dot1q 4
  vrf member IPN
  ip address 10.10.35.2/30
  ip ospf network point-to-point
  ip ospf mtu-ignore
  ip router ospf 1 area 0.0.0.0
  no shutdown
```

```
interface Ethernet1/50.4
  encapsulation dot1q 4
  vrf member IPN
  ip address 10.10.35.6/30
  ip ospf network point-to-point
  ip ospf mtu-ignore
  ip router ospf 1 area 0.0.0.0
  no shutdown
```

```
interface Ethernet1/51.4
  encapsulation dot1q 4
  vrf member IPN
  ip address 10.10.35.10/30
  ip ospf network point-to-point
  ip ospf mtu-ignore
  ip router ospf 1 area 0.0.0.0
  no shutdown
```

```
interface Ethernet1/52.4
  encapsulation dot1q 4
  vrf member IPN
  ip address 10.10.35.14/30
  ip ospf network point-to-point
  ip ospf mtu-ignore
  ip router ospf 1 area 0.0.0.0
  no shutdown
```

```
interface loopback1
  vrf member IPN
  ip address 10.10.35.100/32
```

Step 4 – Verify config on each apic cluster

- Verify L3 out , ospf, bgp was set on each apic cluster
- Verify OSPF is up on spine and gets routes from IPN
- Verify BGP session is up to remote site

Resulting Intersite Profile when site is full configured in MSC

ALL TENANTS | Add Tenant | Tenant Search: Enter name, alias, descr | common | Infra | mgmt

Tenant infra

- > External Routed Networks
- > Dot1Q Tunnels
- > Contracts
- < Policies
 - < Protocol
 - > Route Maps
 - > BFD
 - > BGP
 - > OSPF
 - > EIGRP
 - > IGMP Snoop
 - > IGMP Interface
 - > Custom QOS
 - > End Point Retention
 - > DHCP
 - > ND Interface
 - > ND RA Prefix
 - > Route Tag
 - > L4-L7 Policy Based Redirect
 - > L4-L7 Redirect Health Groups
 - > Data Plane Policing
 - < Fabric Ext Connection Policies
 - Fabric Ext Connection Policy pod35
 - > HSRP
 - > First Hop Security
 - > IP SLA Monitoring Policies

Intrasite/Intersite Profile - Fabric Ext Connection Policy pod35

Properties

Fabric ID: 1
Name: pod35
Community: extended:as2-nn4:5:16
Ex: extended:as2-nn4:5:16

Pod Peering Profile

Peering Type: Full Mesh Route Reflector
Password:
Confirm Password:

Pod Connection Profile

Pod ID	MultiPod Dataplane TEP	Intersite Dataplane TEP
1	10.10.35.101/32	10.10.35.101/32

Fabric External Routing Profile

Name	Subnet
msc-routing-profile	10.10.35.0/24

L3 out automatically create

The screenshot shows the Cisco Application Policy Infrastructure Controller (APIC) web interface. The top navigation bar includes tabs for APIC, System, Tenants (selected), Fabric, Virtual Networking, L4-L7 Services, Admin, Operations, and Apps. Below the navigation is a search bar for 'Tenant Search' with fields for 'Enter name, alias, desc' and filters for 'common', 'infra', and 'mgmt'. The main left sidebar is titled 'Tenant infra' and contains a tree view of networking components. A red box highlights the 'intersite' node under 'Logical Node Profiles'. To the right, a detailed configuration panel is open for the 'intersite' profile, titled 'L3 Outside - intersite'. The properties include:

- Name: intersite
- Alias: (empty)
- Description: optional
- Tags: (empty)
- Global Alias: (empty)
- Provider Label: (empty)
- Target DSOP: Unspecified
- Route Control Enforcement: Import (unchecked), Export (checked)
- VRF: overlay-1
- Resolved VRF: infra/overlay-1
- External Routed Domain: ipn
- Route Profile for Interleaf: select a value
- Route Control For Dampening: (dropdown menu)
- Address Family Type: (dropdown menu)
- Enable BGP/EIGRP/OSPF: BGP (checked), OSPF (checked), EIGRP (unchecked)
- OSPF Area ID: 0
- OSPF Area Control: (checkboxes for OSPF and EIGRP)
- Send redistributed LSAs into NSSA area (checked)
- Originate summary LSA (checked)

L3 out logical node and I3 if in vlan 4

Logical Node Profile - node-201-profile

Name: node-201-profile
Description: optional
Alias:
Target DSCP: Unspecified
Nodes:

Node ID	Port ID	Static Routes	Loopback Address
topology/pod-1/node-201	10.10.35.111		

Logical Interface Profile - port-2-5

Routed Sub-Interfaces:

Path	IP Address	Secondary IP Address	MAC Address	MTU (bytes)	Encap
Pod-1/Node-201/eth2/5	10.10.35.1/30		00:22:BD:F8:19:FF	inherit	vlan-4

Spine BGP node role – Check

- Verify node-role for BGP is msite-speaker both in Object Model and in bgp

```
pod35-spine1# moquery -d sys/bgp/inst | egrep "dn|.*Role"
dn          : sys/bgp/inst
spineRole   : msite-speaker

pod35-spine1# show bgp internal node-role
Node role :                         : MSITE_SPEAKER
pod35-spine1#
```

Overlay-1 check – Spine interface

```
loopback13, Interface status: protocol-up/link-up/admin-up, iod: 119, mode: etep, dci-ucast, vrf_vnid: 16777199
  IP address: 10.10.35.101, IP subnet: 10.10.35.101/32
  IP primary address route-preference: 1, tag: 0
```

Dataplane unicast ETEP (anycast per site)

```
loopback14, Interface status: protocol-up/link-up/admin-up, iod: 120, mode: mcast-hrep, vrf_vnid: 16777199
  IP address: 10.10.35.121, IP subnet: 10.10.35.121/32
  IP primary address route-preference: 1, tag: 0
```

Dataplane multicast TEP

```
loopback15, Interface status: protocol-up/link-up/admin-up, iod: 122, mode: mscp-etepl, vrf_vnid: 16777199
  IP address: 10.10.35.111, IP subnet: 10.10.35.111/32
  IP primary address route-preference: 1, tag: 0
```

CP ETEP (Control plane - BGP router ID)

```
Ethernet2/5.37, Interface status: protocol-up/link-up/admin-up, iod: 121, mode: external, vrf_vnid: 16777199
  IP address: 10.10.35.1, IP subnet: 10.10.35.0/30
  IP primary address route-preference: 1, tag: 0
```

Spine loopback as concrete MO

```
pod35-spine1# cat /mit/sys/ipv4/inst/dom-overlay-1/if-\[lo13\]/summary | egrep "dn|mode"
dn      : sys/ipv4/inst/dom-overlay-1/if-[lo13]
mode    : dci-ucast,etep
modeExtn  :
pod35-spine1# cat /mit/sys/ipv4/inst/dom-overlay-1/if-\[lo14\]/summary | egrep "dn|mode"
dn      : sys/ipv4/inst/dom-overlay-1/if-[lo14]
mode    : dci-mcast-hrep
modeExtn  :
pod35-spine1# cat /mit/sys/ipv4/inst/dom-overlay-1/if-\[lo15\]/summary | egrep "dn|mode"
dn      : sys/ipv4/inst/dom-overlay-1/if-[lo15]
mode    :
modeExtn  : mscp-etep
```

Verify BGP L2vpn evpn and OSPF is up on each spine

```
pod35-spine1#  
pod35-spine1# show ip ospf neighbors vrf overlay-1  
OSPF Process ID default VRF overlay-1  
Total number of neighbors: 2  
Neighbor ID      Pri State          Up Time   Address           Interface  
10.10.35.100     1 FULL/ -        02:06:51  10.10.35.2    Eth2/5.37  
10.10.35.100     1 FULL/ -        02:06:27  10.10.35.6    Eth2/6.38  
pod35-spine1#
```

```
pod35-spine1# show bgp l2vpn evpn summary vrf overlay-1  
BGP summary information for VRF overlay-1, address family L2VPN EVPN  
BGP router identifier 10.10.35.111, local AS number 135  
BGP table version is 26, L2VPN EVPN config peers 1, capable peers 1  
13 network entries and 9 paths using 1864 bytes of memory  
BGP attribute entries [4/576], BGP AS path entries [1/6]  
BGP community entries [0/0], BGP clusterlist entries [0/0]
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
10.10.35.112	4	136	126	124	26	0	0	01:54:15	3

MTU in ISN and Mp-BGP

- MTU of MP-BGP EVPN control-plane communication between spine nodes in different sites: By default, the spine nodes generate 9000-byte packets for exchanging endpoint routing information. If that default value is not modified, the ISN must support an MTU size of at least 9100 bytes. The default value can be tuned by modifying the corresponding system settings in each APIC domain

The screenshot shows the Cisco Application Policy Infrastructure Controller (APIC) web interface. The top navigation bar includes links for Cisco APIC, System (which is highlighted in blue), Tenants, Fabric, Virtual Networking, L4-L7 Services, Admin, and Operations. Below the navigation bar is a secondary menu with links for QuickStart, Dashboard, Controllers, System Settings (which is also highlighted in blue), Faults, Config Zones, Events, Audit Log, and Active Sessions.

The main content area is titled "System Settings". On the left, there is a sidebar with a tree view containing the following items:

- Quota
- APIC Connectivity Preferences
- BD Enforced Exception List
- Control Plane MTU** (This item is currently selected, indicated by a blue background and a blue border around the icon)
- Endpoint Controls
- Fabric Wide Setting
- System Global GIPo
- BGP Route Reflector
- COOP Group
- Load Balancer
- Precision Time Protocol

To the right of the sidebar, a large panel displays the "CP MTU Policy - Control Plane MTU Policy". At the top of this panel, there are three small circular icons. Below them, the title "Properties" is followed by a configuration field labeled "MTU Size for Fabric Ports (bytes):" with the value "9000" and a dropdown arrow.

Multisite Day 1 Tenant Creation

Create multisite Tenant in MSC

The screenshot shows the ACI Multi-Site interface. On the left sidebar, the 'Tenants' option is selected and highlighted with a red box. In the main content area, the 'Tenants' page is displayed with two existing tenants: 'common' and 'MetalShop'. A red box highlights the 'ADD TENANT' button at the top right of the tenant list. A blue arrow points from a yellow callout box containing instructions to the 'DISPLAY NAME' field in the 'Tenant details' modal. Another blue arrow points from the 'DISPLAY NAME' field to the 'Associated Sites' section in the same modal, where three sites are selected: 'SITE', 'POD35', and 'POD36'. The 'Tenant details' modal has a blue header bar.

ACI Multi-Site

Controller Status 3/3 | Welcome, Admin

Tenants

SEARCH

ADD TENANT

NAME	DESCRIPTION	ASSIGNED TO SITES	ASSIGNED TO USERS	ASSIGNED TO SCHEMAS	ACTIONS
common	Common tenant for use ...	2	1	0	
MetalShop	We make Metal	2	1	1	

Tenant details

General Settings

* DISPLAY NAME
DC

Internal Name: DC

DESCRIPTION
Prod Datacenter Tenant

Associated Sites

SITE

POD35

POD36

SECURITY DOMAINS

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Add a tenant
Give tenant name
And associate it to one or more site

Tenant is created on each APIC cluster of site associated

See the warning that this tenant was created from Msite controller

The screenshot shows the Cisco Application Policy Infrastructure Controller (APIC) interface. The top navigation bar includes tabs for APIC, System, Tenants, Fabric, Virtual Networking, L4-L7 Services, Admin, Operations, and Apps. The 'Tenants' tab is selected, highlighted in blue. Below the tabs is a search bar labeled 'Tenant Search: Enter name, alias, descr' with a placeholder 'common'. A red box highlights a message in the center of the screen: 'This has been created from Multi-Site. It is recommended to only make changes from Multi-Site. Please review the documentation before making any changes here.' An arrow points from the yellow callout box above to this message. On the left, a sidebar for 'Tenant DC' shows a tree structure with 'Tenant DC' expanded, containing 'Application Profiles', 'Networking', 'Contracts', 'Policies', and 'Services'. The main content area is titled 'Tenant - DC' and displays a 'Health' section with a chart. The chart has 'Score' on the vertical axis and 'Time' on the horizontal axis. A legend at the top of the chart area indicates 'Zoom' options: '1H' (selected), '1D', and 'All'. The chart area contains the text 'No stats data to display...'. Above the chart, there are three small circular icons.

Multisite Day 1 Schema Management



What are Schema ?

Create New schema
In Schemas- Add Schema

- Schema list displays tenants associated to a schema as well as templates in it.
- User can edit, delete and create a new schema from this page.
- Schema can be seen as a use-case skeleton which could spread over one or more tenant and one or more site
- Schema contains Templates

The screenshot shows the Cisco ACI Multi-Site dashboard. On the left, a sidebar menu has 'Schemas' selected and highlighted with a red box. At the top right, there are status indicators for 'Controller Status' (3/3) and a welcome message for 'Admin'. Below the header, the main content area is titled 'Schemas'. It features a search bar and a table with columns: NAME, TEMPLATES, TENANTS, and ACTIONS. One row is visible in the table:

NAME	TEMPLATES	TENANTS	ACTIONS
Untitled Schema	Extreme Metal Shop	MetalShop	Edit

At the bottom right of the content area, there is a blue button labeled 'ADD SCHEMA' with a red box around it.



New schema Startup Screen - Templates

The screenshot shows the 'Untitled Schema' interface. On the left, there's a sidebar with 'TEMPLATES' and 'SITES'. Under 'TEMPLATES', 'Template 1' is selected. The main area displays 'Template 1' with a note: 'To build your schema please click here to select a tenant'. A red box highlights this note. At the bottom of the template area, there are 'Application profile' and 'Add EPG' buttons. In the top right corner, there are 'SAVE', 'DEPLOY TO SITES', and close ('X') buttons.

- You can name the Schema
- Schema starts with a default Template 1 that you can rename as well
- You can add extra templates in the scheme
- When building a templates, you must first assign it to a tenant

The screenshot shows the 'DC-Schema' interface. On the left, there's a sidebar with 'TEMPLATES'. Under 'TEMPLATES', 'L3-VRF-Stretched' is selected. The main area displays 'L3-VRF-Stretched' with the note 'Applied to 2 sites'. A red box highlights the 'L3-VRF-Stretched' button. In the top right corner, there are 'SAVE', 'DEPLOY TO SITES', and close ('X') buttons.



Select a tenant for the Template

To build your schema please click here to select a tenant

common
Common tenant for use with all other tenants

DC
Prod Datacenter Tenant

MetalShop
We make Metal

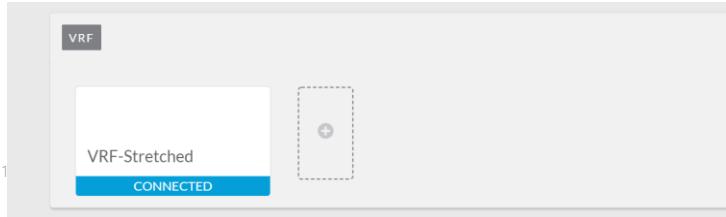
Once the tenant is Selected you can add App Prof/Epg VRF, BD and Contract

MSC GUI logic

The screenshot shows the Cisco MDS GUI interface. On the left, there's a sidebar with 'TEMPLATES' and 'SITES'. Under 'TEMPLATES', 'L3-VRF-Stretched' is selected and highlighted in blue. In the main panel, under 'TENANT DC', there's a section for 'AP' with 'App1' listed. A red box highlights the 'App1' entry. To the right, there's a 'APPLICATION NETWORK PROFILE App1' section with a 'NAME' field containing 'App1' and a 'DESCRIPTION' field containing 'My first ANP'. A yellow callout box says: 'Whenever you click somewhere in middle panel you can modify Name and property in right pane'.

We created ANP

No we create a VRF in the Tenant (name VRF-stretched)





Schema builder: build templates and objects

Templates allows to create objects:

- App Profile
- EPG
- VRF
- Contract
- BD
- Filter
- External EPG

The screenshot shows the 'Web-App-Use-Case1' template builder interface. On the left, a sidebar lists 'Templates' (selected), 'Web-App' (highlighted in blue), and 'Sites'. The main area is titled 'Web-App' with a 'PUSH' button. It shows a tenant named 'IPA-Test' and an Application Network Profile (ANP) named 'Web-App-ANP-Case1'. This profile contains two EPGs: 'Web-EPG-Case1' (Consumed) and 'App-EPG-Case1' (Provided). Below the ANP is a 'CONTRACT' section containing a single entry 'C1-Case1'. Further down are sections for 'VRF' (containing 'VRF-Stone-IPA-Case1') and 'BRIDGE DOMAIN' (containing 'BD1' and 'BD 2'). On the right, a sidebar displays 'CONTRACT' details for 'C1-Case1', showing 2 local relationships and 0 external relationships. It also includes sections for 'DISPLAY NAME' (C1-Case1), 'SCOPE' (vrf), and 'FILTER CHAIN'.



Schema builder: associate sites

Click the + sign next to Sites, there you can modify which templates applies to which sites.

The screenshot shows the Schema builder interface with a modal dialog titled "Add Sites". The left sidebar lists various templates: L3-VRF-Stretched, Pod35-only, SITES, POD35, L3-VRF-Stretched (with a warning icon), POD36, and L3-VRF-Stretched (with a green checkmark). A red box highlights the "+ sign" next to "SITES". The main area shows an "L3-VRF-Stretched" template applied to 2 sites, with tabs for "TENANT" (set to DC) and "AP" (set to App1). A red box highlights the "Add EPG" button. Below it is an "Application Profile" section with a red box highlighting the "+ Application Profile" button. The "CONTRACT" section is partially visible at the bottom. The "Add Sites" dialog has two columns: "NAME" and "ASSIGN TO TEMPLATE". In the "NAME" column, "POD35" and "POD36" are listed with checkboxes. In the "ASSIGN TO TEMPLATE" column, "L3-VRF-Stretched" and "Pod35-only" are selected, indicated by a blue border and a small circular icon with a dot. A red box highlights this selection. At the bottom right of the dialog is a "SAVE" button.

NAME	ASSIGN TO TEMPLATE
<input checked="" type="checkbox"/> POD35	L3-VRF-Stretched Pod35-only
<input checked="" type="checkbox"/> POD36	None available <small>Select or find an item here</small>

SAVE



Schema builder: associate sites

- Templates containing policies have to be associated to all sites where they have to be stretched
- User can assign specific template to selected sites

Add Sites X

NAME	ASSIGN TO TEMPLATE
<input checked="" type="checkbox"/> London	Stone-IPA-Stretched Web-App-London ▼
<input checked="" type="checkbox"/> New York	Stone-IPA-Stretched Web-App-NewYork ▼
<input checked="" type="checkbox"/> San Francisco	Stone-IPA-Stretched ▼
<input checked="" type="checkbox"/> Seattle	Stone-IPA-Stretched Web-App-Seattle ▼
<input checked="" type="checkbox"/> Toyko	Web-App-Toyko Stone-IPA-Stretched ▼

CONFIRM



Schema builder: add sites

- Click on “+” next to sites to select which template should be deployed on which site

Add Sites X

NAME	ASSIGN TO TEMPLATE
<input checked="" type="checkbox"/> London	Web-App ▼
<input checked="" type="checkbox"/> New York	Web-App ▼
<input type="checkbox"/> San Francisco	
<input type="checkbox"/> Seattle	
<input type="checkbox"/> Tokyo	

CONFIRM CANCEL

Example EPG screen in Template

- 1 Here Template Site1-only is selected
- 2 EPG : Web-EPG1 is selected in the template
- 3 We can modify EPG to BD assoc, contract cons or prov.
- 4 For any modif you need to Save

The screenshot shows the Cisco ACI EPG configuration interface for the 'DC-Stretched VRF' tenant. The left sidebar lists sites: Site1-only (selected), Site2-only, POD35, Global-All-Site, POD35, POD36, Global-All-Site, and Site2-only. The main panel shows the 'Site1-only' template applied to one site. In the 'AP' section, 'Web-EPG1' is selected (indicated by a red box and number 2). On the right, a detailed view of the 'Web-EPG1' configuration is shown, including its display name, subnets, gateway IP, and bridge domain BD1. A modified contract 'C1' is listed under 'CONTRACTS' with the status 'CONSUMED'. A red box and number 4 point to the 'SAVE' button at the top right of the configuration panel.

1 Site1-only

2 Web-EPG1

3 Web-EPG1

4 SAVE

DC-Stretched VRF

Site1-only
Applied to 1 sites

TENANT DC

AP Ap1

Site2-only

SITES POD35

Global-All-Site

POD35

POD36

Global-All-Site

Site2-only

Application Profile

CONTRACT

VRF

BRIDGE DOMAIN

EPG Web-EPG1

DISPLAY NAME Web-EPG1

Name: Web-EPG1

SUBNETS

GATEWAY IP

+ SUBNET

USEG EPG

USEG ATTR

N/A

INTRA EPG ISOLATION

Enforced

Unenforced

BRIDGE DOMAIN

BD1

CONTRACTS

C1 consumer

C2 consumer

CONTRACT

Changes Saved by not deployed yet

We see POD35/Site1-only
Template/site assoc which is not in sync
Click Deploy to site in Top right of UI to deploy

Site1-only	
Site2-only	
SITES	
POD35	
Global-All-Site	
POD35	
Site1-only	
POD36	
Global-All-Site	
POD36	
Site2-only	

Modif in Template/Site association

- 1 Here we selected Site/Template combo (Pod35 – Site1only)
- 2 Same EPG
- 3 But we can edit/add more property (local to site) such as Domain or static path

The screenshot shows the Cisco ACI DC Stretched VRF interface. On the left, the navigation pane lists templates: Global-All-Site, Site1-only, Site2-only, POD35, and POD36. The main area displays the 'DC' tenant configuration. A red circle labeled '1' highlights the 'Site1-only' entry under 'POD35'. Another red circle labeled '2' highlights the 'Web-EPG1' and 'App-EPG1' entries under the 'AP' section. A third red circle labeled '3' highlights the detailed configuration window for 'Web-EPG1', which includes fields for DISPLAY NAME, SUBNETS, USESEG EPG, STATIC PORTS, STATIC LEAF, INTRA EPG ISOLATION, BRIDGE DOMAIN, DOMAINS, and CONTRACTS.

1

2

3

POD35

Site1-only

Site2-only

POD36

Global-All-Site

DC

AP Ap1

Web-EPG1 App-EPG1

C1 CONSUMED

BD1 BD3 CONNECTED

all

EXTERNAL-EPG

Web-EPG1

DISPLAY NAME: Web-EPG1
Name: Web-EPG1

SUBNETS

GATEWAY IP

SUBNET

USESEG EPG

STATIC PORTS

PATH	TYPE	VLAN
eth1/25 (node-101)	port	101

STATIC PORT

USESEG ATTR
N/A

STATIC LEAF

NODE	VLAN	ACTION
Site2-only	101	STATIC LEAF

INTRA EPG ISOLATION

Enforced
Unenforced

BRIDGE DOMAIN
BD1

DOMAINS

PROFILE	TYPE
VMMPOD35	vmm
RD-Phys	physical

DOMAIN

CONTRACTS



Deploy schema

- We can deploy a template to associated sites.
Push button provides this functionality.
- Note : Push button was replaced by “deploy to site” in latest MSC software
- Push saves entire schema and then deploys to associated sites.

The screenshot shows the Cisco MDS Manager interface for a project named "Web-App-UseCase-1.1". On the left, there's a sidebar with sections for "Templates" and "Sites". Under "Templates", "Web-App-NewYork" is selected. Under "Sites", "London" and "New York" are listed. In the main content area, a "TENANT" tab is selected, showing an "IPA-Test" tenant. Inside the tenant, there's an "ANP" section for "Web-App-ANP" containing "Web-EPG1" and "App-EPG1" with a "+ Add EPG" button. Below it is an "Application Network Profile" section. Further down are sections for "CONTRACT" (with a box labeled "C1") and "VRF". A success message box at the top right says "Successfully deployed Web-App-NewYork". Buttons for "SAVE", "PUSH", "IMPORT", and "EXPORT" are visible at the top right.

Schema <> templates

- You can have multiple template per schema
- Anything in a Schema is pushed to the same group of site
- So if you want to stretch VRF but not BD you need different template
 - Template 1 : VRF pushed to both site
 - Template 2 : BD – subnet pushed to site 1 only
 - Template 2 : BD – subnet Pushed to site 2 only

Multisite Controller Troubleshooting log

MSC audit log

MAIN MENU

cisco ACI Multi-Site

Controller Status 3/3 | Welcome, Admin

Audit Logs (312 Logs)

DATE ACTION TYPE DETAILS USER

DATE	ACTION	TYPE	DETAILS	USER
Apr 11, 2018 10:50 AM	Created	Backup	Backup BackupRD_20180411085033 was created	admin (Admin User) Local
Apr 11, 2018 10:48 AM	Logged In	Authentication	User admin has successfully logged in	admin (Admin User) Local
Apr 11, 2018 10:08 AM	Deployed	Schema Site	Template Pod35-only was deployed on POD35	admin (Admin User) Local
Apr 11, 2018 10:08 AM	Deployed	Schema Site	Template Pod35-only was deployed on POD36	admin (Admin User) Local
Apr 11, 2018 10:08 AM	Updated	Schema	Schema DC-Schema was updated	admin (Admin User) Local
Apr 11, 2018 10:06 AM	Updated	Schema	Schema DC-Schema was updated	admin (Admin User) Local
Apr 11, 2018 10:06 AM	Updated	Template	Template Pod35-only was updated	admin (Admin User) Local
Apr 11, 2018 10:06 AM	Created	EPG	EPG test was created	admin (Admin User) Local

Can be downloaded
in Json or csv format

Generate Troubleshooting Report

Controller Status **3/3**

Welcome, Admin ^

Troubleshooting Report

Reset Password

Log Out

Troubleshooting Report

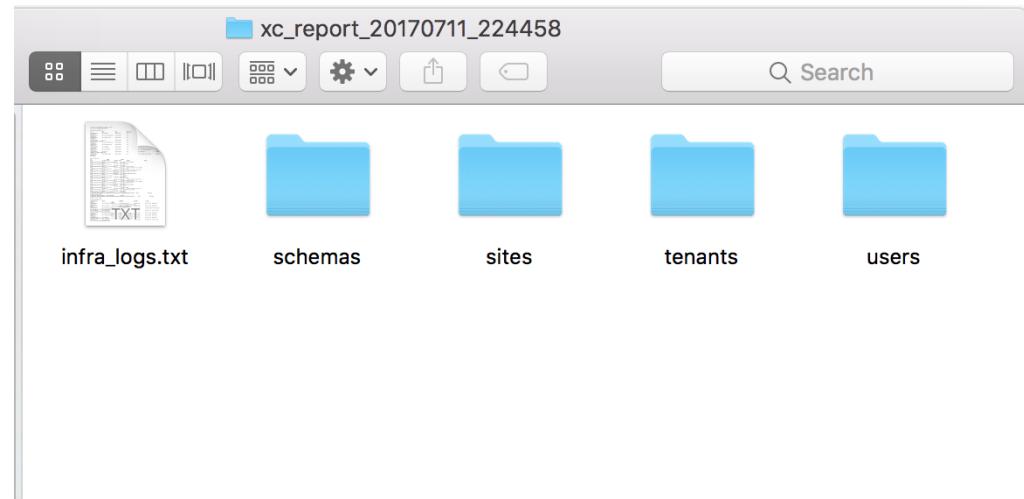
Select Logs

- Sites
- Tenants
- Schemas
- Users
- Infra Logs

DOWNLOAD

What you get for the troubleshooting report

- schemas All schemas in json format
- sites All sites definition in json format
- tenants All tenants definition in json format
- users All user definition in json format
- infra_logs.txt All logs of containers



Login to the VM to troubleshoot

- Useful commands for docker
 - Docker command cheatsheet
https://www.docker.com/sites/default/files/Docker_CheatSheet_08.09.2016_0.pdf
 - Inspect all docker services
 - Inspect logs for specific container
-
- SSH To MSC using root login (password is the same admin got MSC gui)

Inspect health of docker services

- docker service ls

ID	NAME	MODE	REPLICAS	IMAGE	PORTS
09tdb043c9f0	msc_siteservice	replicated	1/1	msc-siteservice:latest	*:9010->9010/tcp
e3um1no9gws2	msc_kong	replicated	1/1	msc-kong:1.1	*:8000->8000/tcp,*:8001->8001/tcp
jgp6p39r3v8v	msc_kongdb	replicated	1/1	postgres:9.4	*:5432->5432/tcp
mvhcf46tv572	msc_executionengine	replicated	1/1	msc-executionengine:latest	*:9030->9030/tcp
oxggeba7tv8y	msc_ui	replicated	1/1	msc-ui:latest	*:80->80/tcp,*:443->443/tcp
qk2ta0vxop9e	msc_userservice	replicated	1/1	msc-userservice:latest	*:9040->9040/tcp
qkh1346ttidp	msc_syncengine	replicated	1/1	msc-syncengine:latest	*:9060->9060/tcp
qw088qvr7tzk	msc_mongodb	replicated	1/1	mongo:3.4	*:27017->27017/tcp
w2nd7j92c1p6	msc_platformservice	replicated	1/1	msc-platformservice:latest	*:9050->9050/tcp
yvuanhjo7l5q	msc_schemaservice	replicated	1/1	msc-schemaservice:latest	*:9020->9020/tcp

The output is the expected health status. All services have at least 1 container replicated.
If any of them is down, the system might not work correctly.

Lab : getting execution engine log

```
[root@node1 containers]# docker ps | egrep executionengine
f15fc0522d15      msc-executionengine:1.1.2b    "bin/executionengine"    3 weeks ago      Up 3 weeks (healthy)   9030/tcp
msc_executionengine.1.r5u9ful5oa5ed460r8csu7n7s
[root@node1 containers]# ls -al | egrep f15fc0522d15
drwx-----. 4 root root 4096 Apr 19 07:05 f15fc0522d157317cdfab7acfb8035e3291b2ae8f0b2340231976a1b23d7770
[root@node1 containers]# pwd
/var/lib/docker/containers
[root@node1 containers]# cd f15fc0522d157317cdfab7acfb8035e3291b2ae8f0b2340231976a1b23d7770
[root@node1 f15fc0522d157317cdfab7acfb8035e3291b2ae8f0b2340231976a1b23d7770]# ls -al
total 6112
drwx-----. 4 root root 4096 Apr 19 07:05 .
drwx-----. 22 root root 4096 Mar 30 02:07 ..
drwx-----. 2 root root 6 Mar 30 02:07 checkpoints
-rw-r--r--. 1 root root 3849 Mar 30 02:07 config.v2.json
-r-----. 1 root root 1219224 Apr 23 08:01 f15fc0522d157317cdfab7acfb8035e3291b2ae8f0b2340231976a1b23d7770-json.log
-rw-r-----. 1 root root 5000090 Apr 19 07:05 f15fc0522d157317cdfab7acfb8035e3291b2ae8f0b2340231976a1b23d7770-json.log.1
-rw-r--r--. 1 root root 1181 Mar 30 02:07 hostconfig.json
-rw-r--r--. 1 root root 13 Mar 30 02:07 hostname
-rw-r--r--. 1 root root 173 Mar 30 02:07 hosts
-rw-r--r--. 1 root root 38 Mar 30 02:07 resolv.conf
-rw-r--r--. 1 root root 71 Mar 30 02:07 resolv.conf.hash
drwxrwxrwt. 2 root root 40 Mar 30 02:07 shm
```

Json exec engine log sample

```
{"log":"2018-04-19 14:05:05,933 [\u001b[36mdebug\u001b[0m] execution.service.actor.ApicActor - Site actor initialized\n5abe073a0e00002d007becf8\\n","stream":"stdout","time":"2018-04-19T14:05:05.93344042\n5Z"}\n{"log":"2018-04-19 14:05:05,954 [\u001b[36mdebug\u001b[0m] e.service.orchestration.OrchestratorService - Deploying schema: DC-Stretched VRF -\ntemplate: Site1-only (templateID: 5acdced420f00007601b0769e\n-Site1-only) -- tenant: DC to site -- POD36 (siteID: 5abe073a0e00002d007becf8)\\n","stream":"stdout","time":"2018-04-19T14:05:05.95556287Z"}\n{"log":"2018-04-19 14:05:05,954 [\u001b[36mdebug\u001b[0m] e.service.orchestration.OrchestratorService - Plan for schema: DC-Stretched VRF -\ntemplate: Site1-only (templateID: 5acdced420f00007601b0769e-\nSite1-only) -- tenant: DC to site -- POD36 (siteID: 5abe073a0e00002d007becf8)\\n","stream":"stdout","time":"2018-04-19T14:05:05.95558013Z"}\n{"log":"2018-04-19 14:05:05,955 [\u001b[36mdebug\u001b[0m] e.service.orchestration.OrchestratorService - Payload for APIC :\n\\n","stream":"stdout","time":"2018-04-19T14:05:05.955584084Z"}\n{"log": "\\n","stream":"stdout","time":"2018-04-19T14:05:05.955586892Z"}\n{"log": "\u003cpolUni\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955588954Z"}\n{"log": " \u003cfvTenant name=\\\"DC\\\"\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955591133Z"}\n{"log": " \u003cvzBrCP name=\\\"C1\\\" scope=\\\"context\\\"\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955593561Z"}\n{"log": " \u003cvzSubj name=\\\"msc-subject\\\"\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.95559607Z"}\n{"log": " \u003cvzRsSubjFiltAtt tnVzFilterName=\\\"all\\\" directives=\\\"none\\\"\\u003e\n\\u003c/vzRsSubjFiltAtt\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955598438Z"}\n{"log": " \u003c/vzSubj\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955601272Z"}\n{"log": " \u003c/vzBrCP\\u003e\\n","stream":"stdout","time":"2018-04-19T14:05:05.955603389Z"}
```

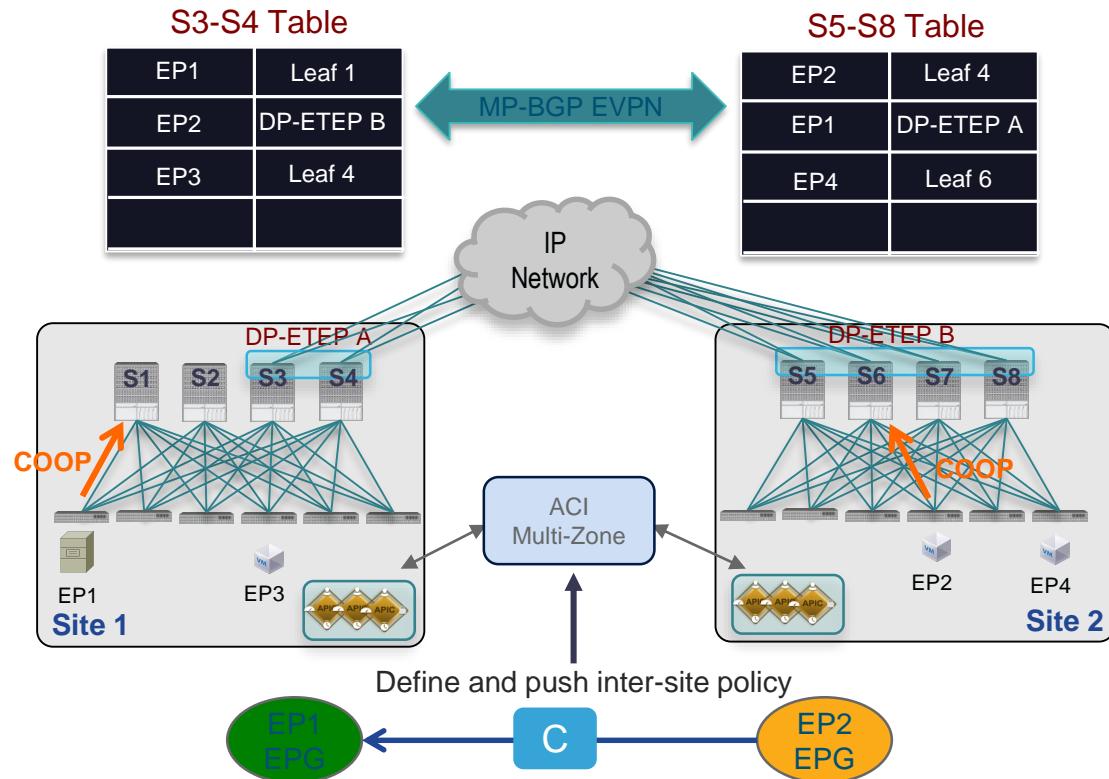
Reading the execution log

- Three kinds of log information
 - 1. Websocket refresh information (printed out every 5 mins)
 - 2. The schema to push, and the plan being generated (most important information)
 - 3. Websocket monitoring vnid for cross vnid programming (important)
- Signs of errors
 - Log lines starting with red “error”
 - Stacktrace for exceptions

Control Plane and Data plane
High level

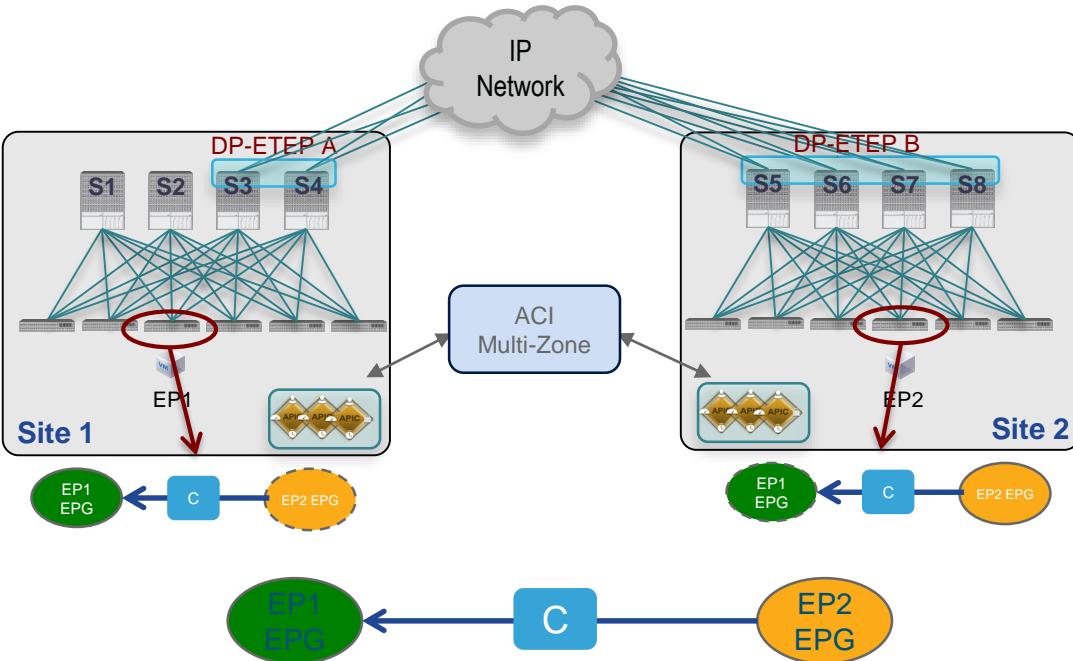
ACI Multi-Site Inter-Site MP-BGP EVPN Control Plane

- MP-BGP EVPN used to communicate Endpoint (EP) information across Sites
 - MP-iBGP or MP-EBGP peering supported across sites
 - Remote host route entries (**EVPN Type-2**) are associated to the remote site Anycast DP-ETEP address
- Automatic filtering of endpoint information across Sites
 - Host routes are exchanged only if there is a cross-site contract requiring communication between endpoints

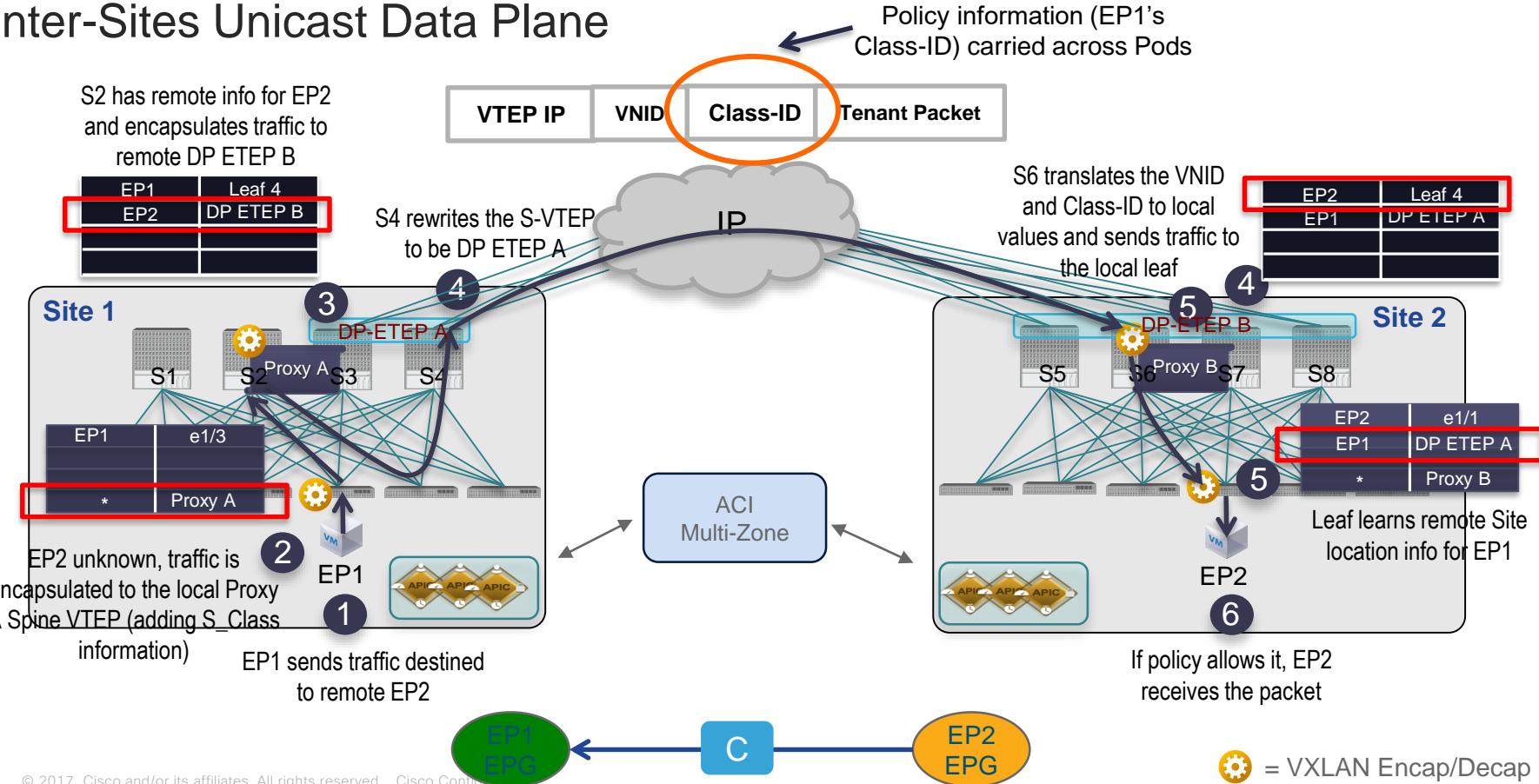


ACI Multi-Site Inter-Site Policies

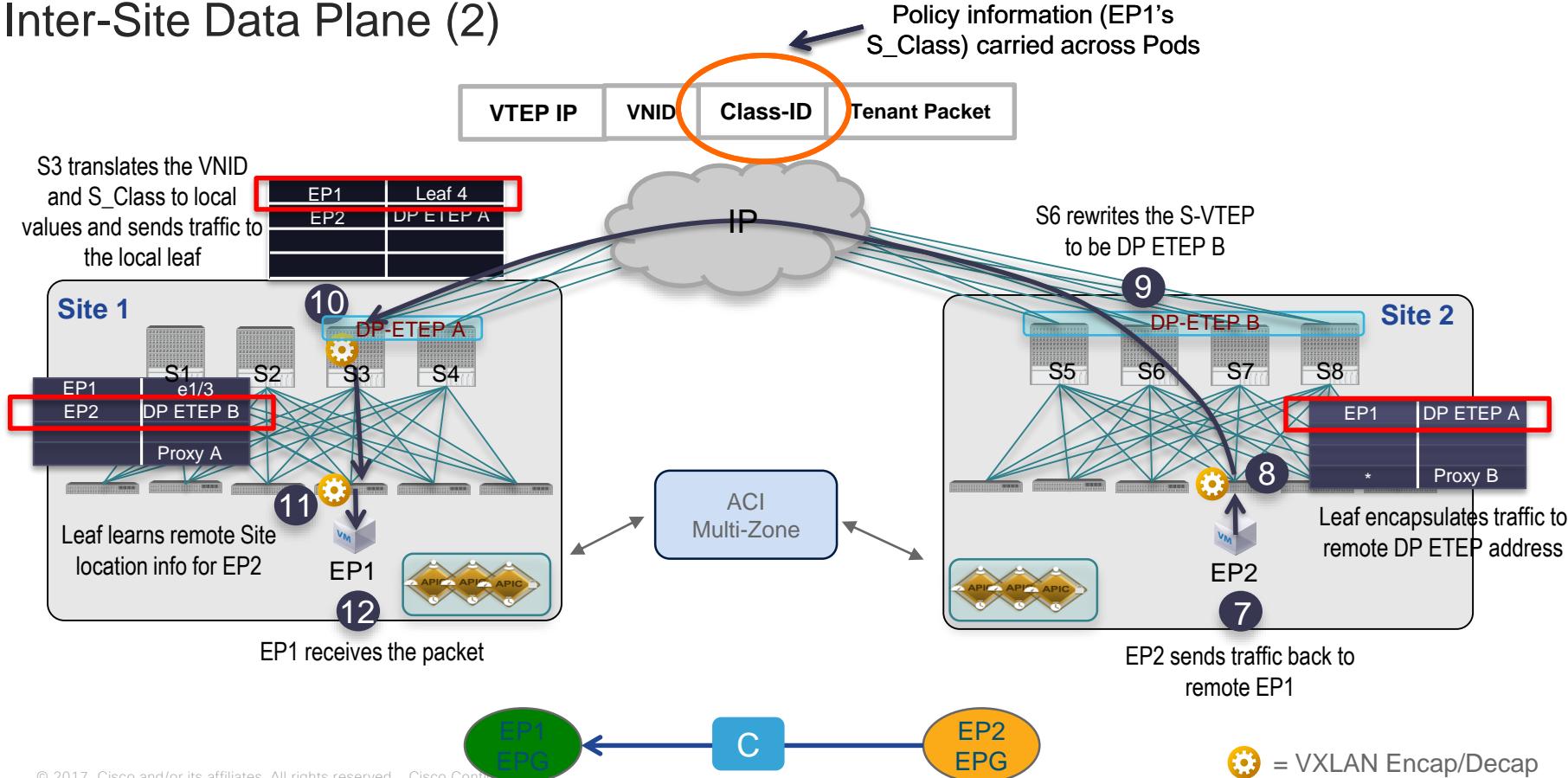
- Inter-Site policies defined on the ACI Multi-Zone are pushed to the respective APIC domains
- Policies are enforced at the ingress leaf node, once it has learned on the data plane info for remote endpoint



ACI Multi-Site Inter-Sites Unicast Data Plane



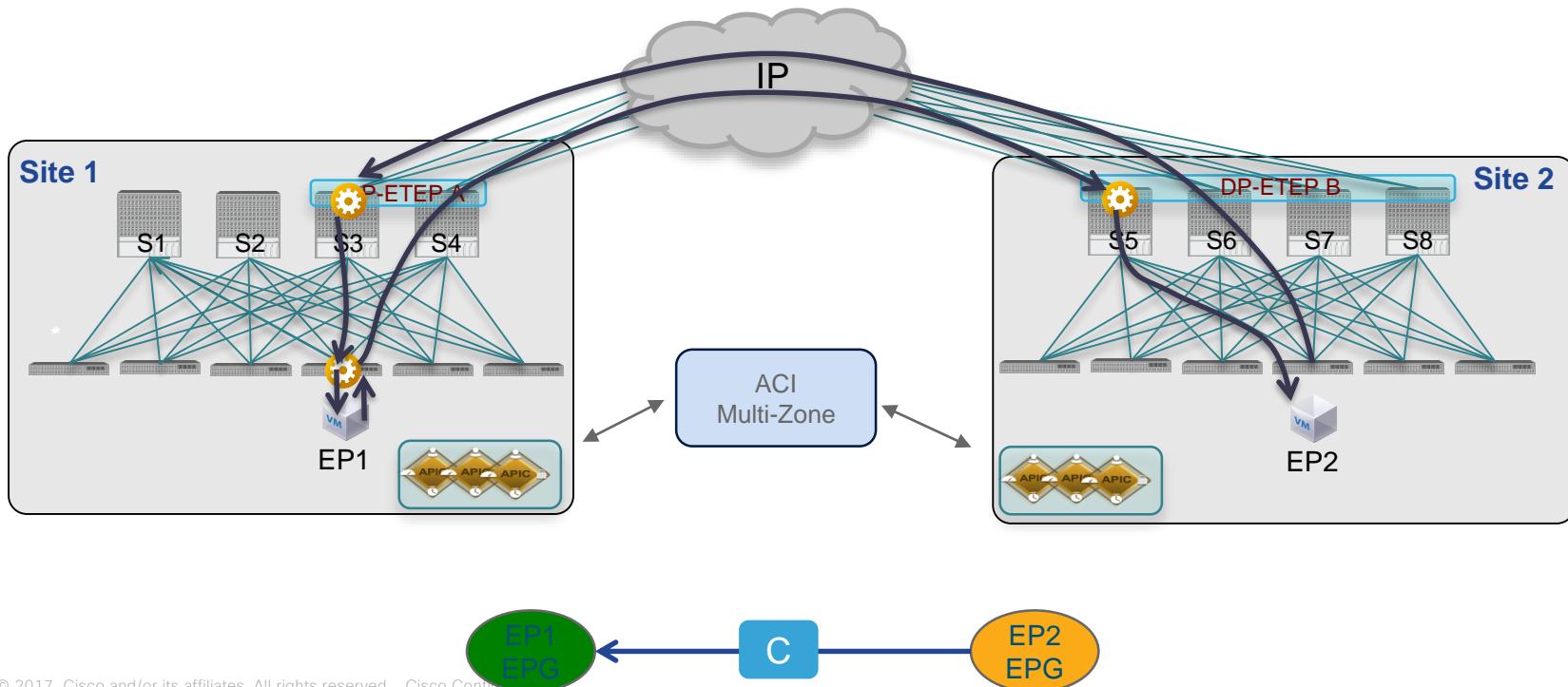
ACI Multi-Site Inter-Site Data Plane (2)



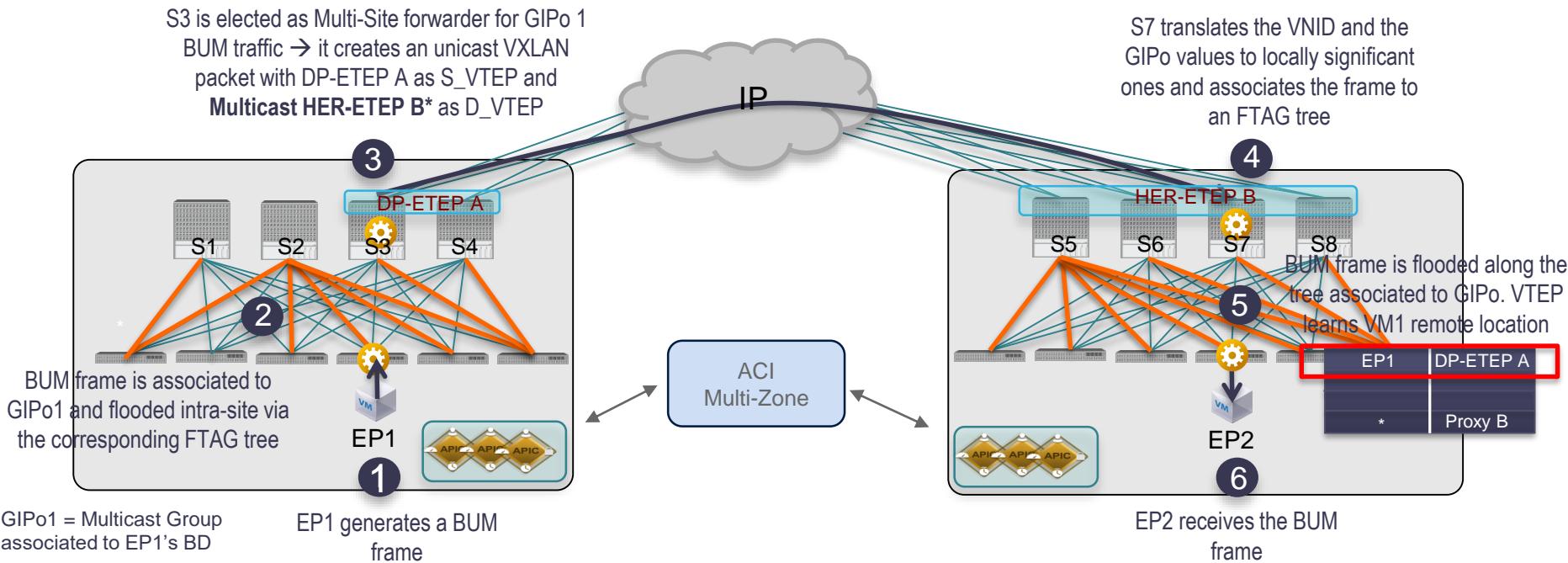
ACI Multi-Site Inter-Site Data Plane (3)

Orange gear = VXLAN Encap/Decap

From this point EP1 to EP2 communication is encapsulated Leaf to Remote Spine DP ETEPs in both directions



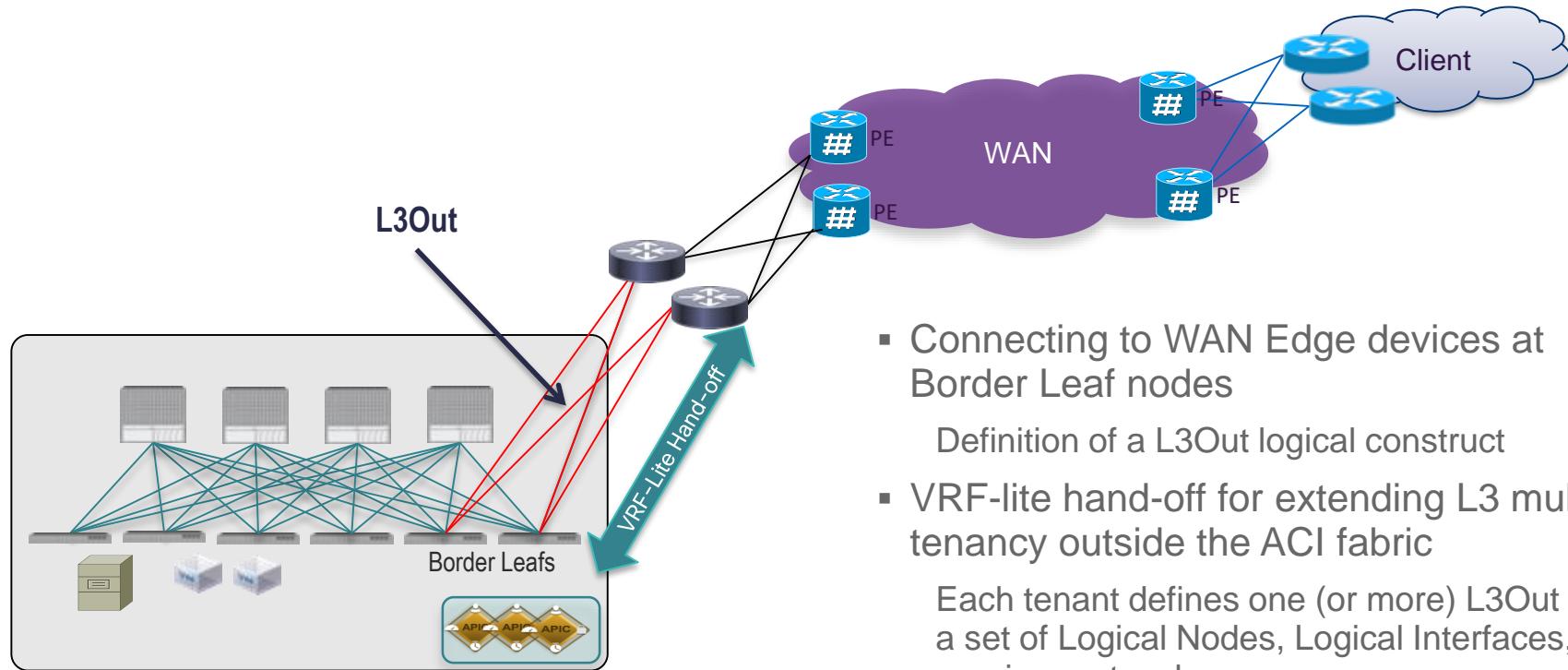
ACI Multi-Site Layer 2 BUM Traffic Data Plane



*This is a different ETEP address than the one used for inter-site L3 unicast communication

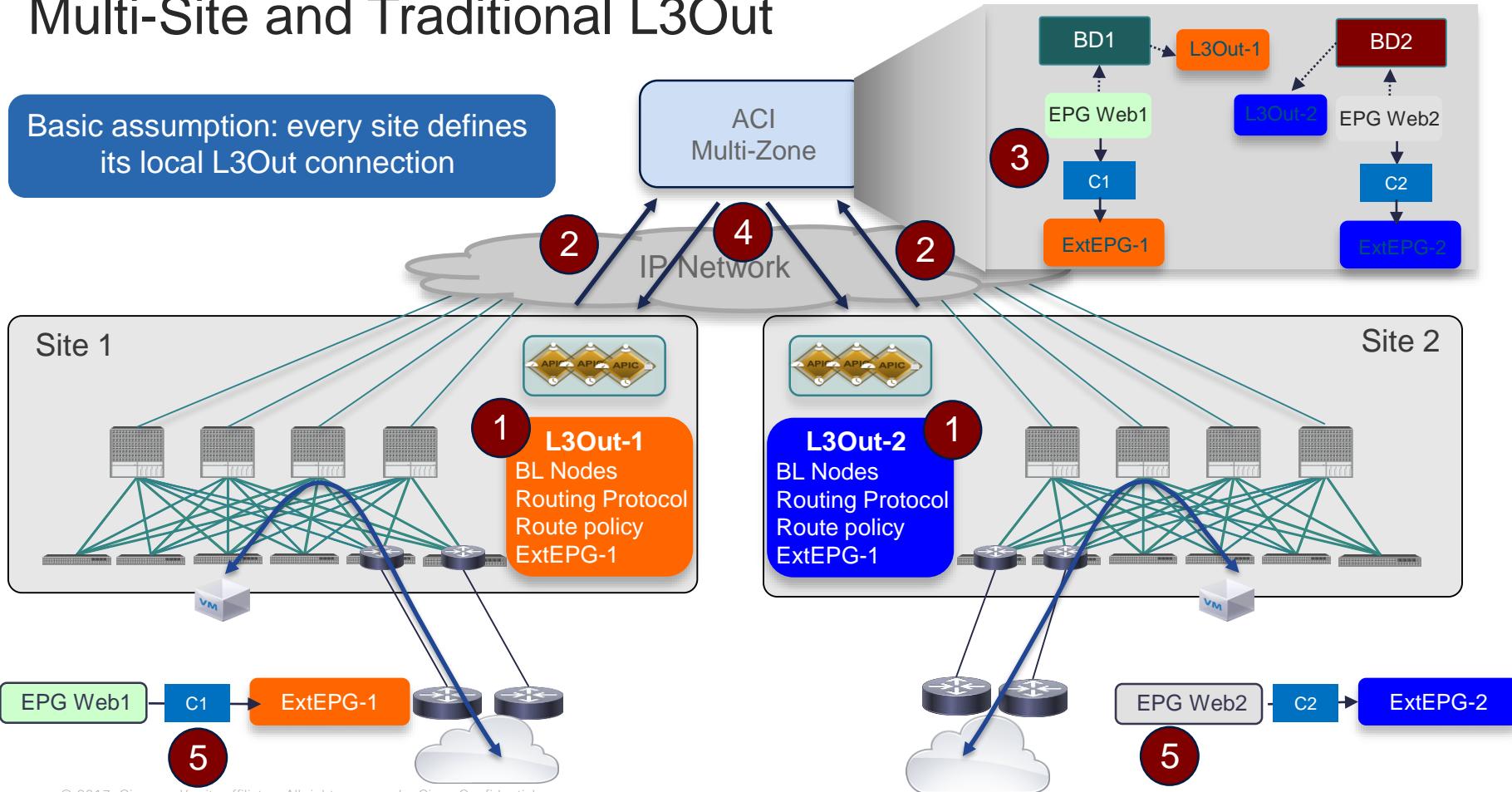
Connecting ACI to Layer 3 Domain

'Traditional' L3Out on the BL Nodes

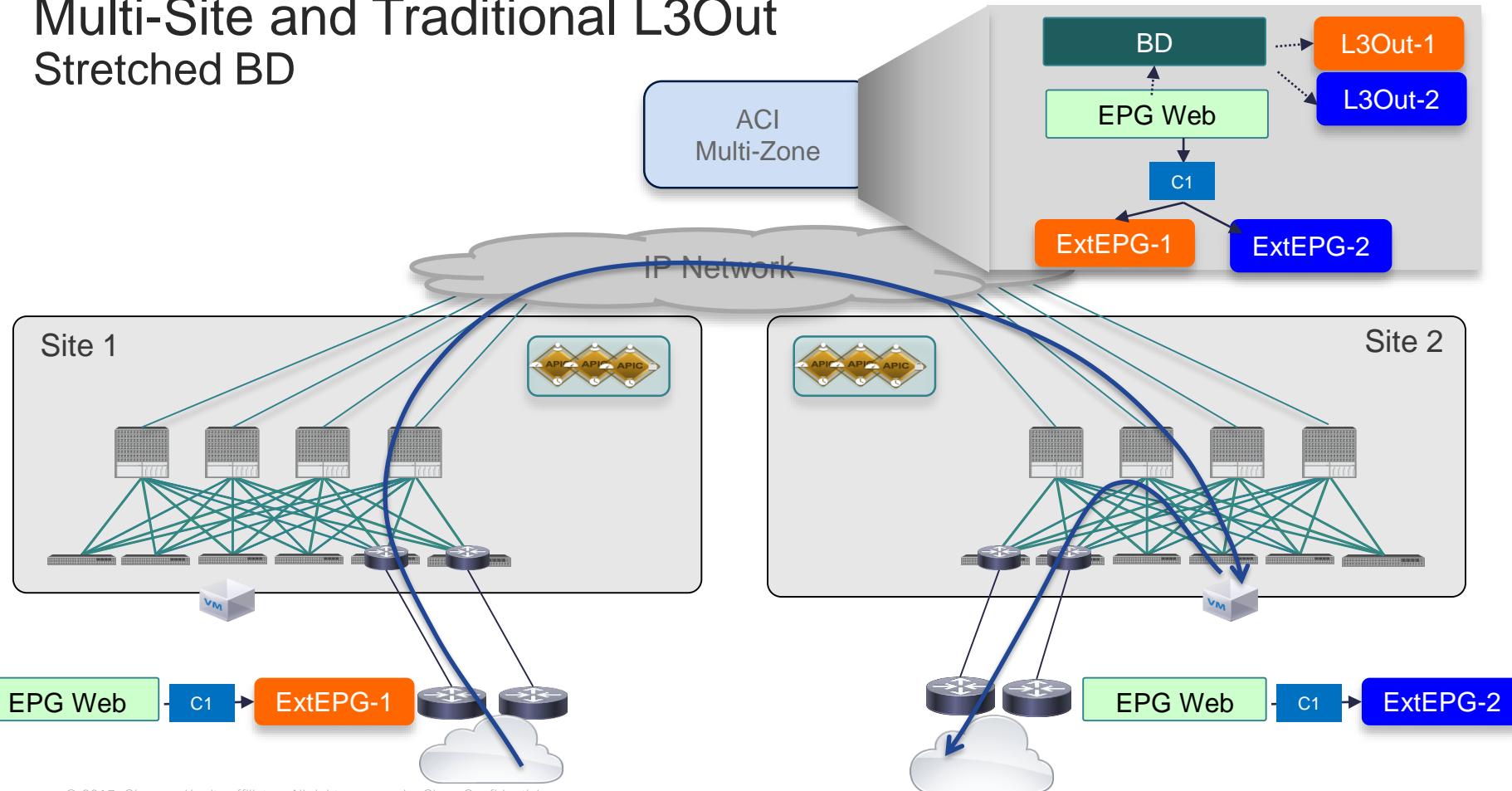


Multi-Site and Traditional L3Out

Basic assumption: every site defines its local L3Out connection



Multi-Site and Traditional L3Out Stretched BD



ACI Multi-Site Handling Silent Hosts

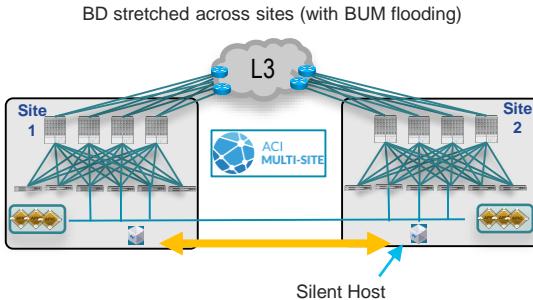
- While silent hosts are rare in real production network, be aware of this behavior when performing lab tests (CPOC, etc.)
- Silent Hosts cannot currently be discovered when the destination endpoint is in a **separate site** and part of a different BD than the source endpoint



- ARP Glean not supported across sites in 3.0 and 3.1 , so the destination host cannot be forced to reply to ARP request and get discovered (Note: ARP Glean works instead intra-site)
- Support for discovery of silent hosts is available since 3.2 release

ACI Multi-Site Handling Silent Hosts

- Silent Hosts can instead be discovered when the destination endpoint is in a separate site but part of the **same BD** of the source endpoint (**BD stretched with BUM flooding enabled**)



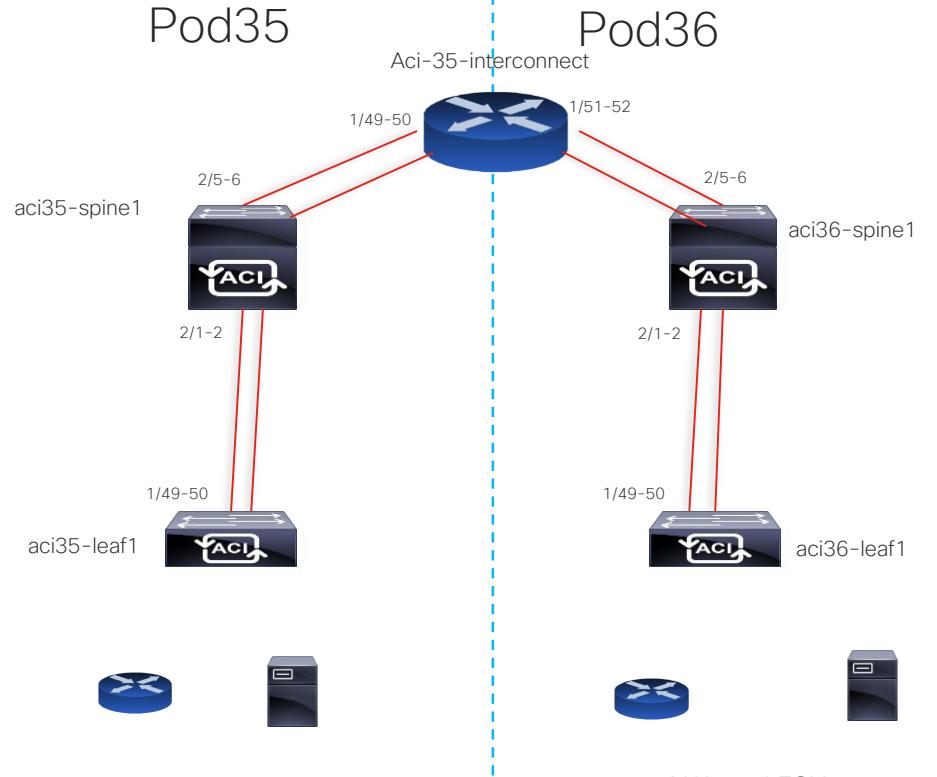
- ARP request generated by the endpoint in site 1 is flooded to site 2 and forces the silent host to reply and hence get discovered

ACI Silent Host discovery (Only in 3.2)

- From 3.2 onwards, we will be supporting silent host discovery across sites using Multisite-Glean (MS-Glean) path
- The following traffic types will support MS-Glean on the spine, when the EP lookup is a miss on the spine
 - Unknown L3 Unicast traffic with Destination BD deployed in proxy mode
 - L2 ARP traffic with ARP configured in unicast mode
- TX (local to remote-site)
 - Unknown L3/L2 ARP Unicast Traffic sourced from the local site is Head-end replicated (HREP) to each remote site from the Spine. DIPO is rewritten to a unicast address called as Unicast ETEP IP of the remote site. SIPO is rewritten with the Unicast ETEP IP of the local site.
- RX (remote to local site)
 - Incoming traffic destined to the local site's unicast ETEP IP goes through vnid and sclass translations. The receiving spine looks up the route for destination EP, which will result in a miss and trigger the existing Glean path to reach all the TOR in the local site, who will deliver the L3Unicast/ARP traffic to the silent-host.

Multi-Site Detailed Use cases

Lab Layout



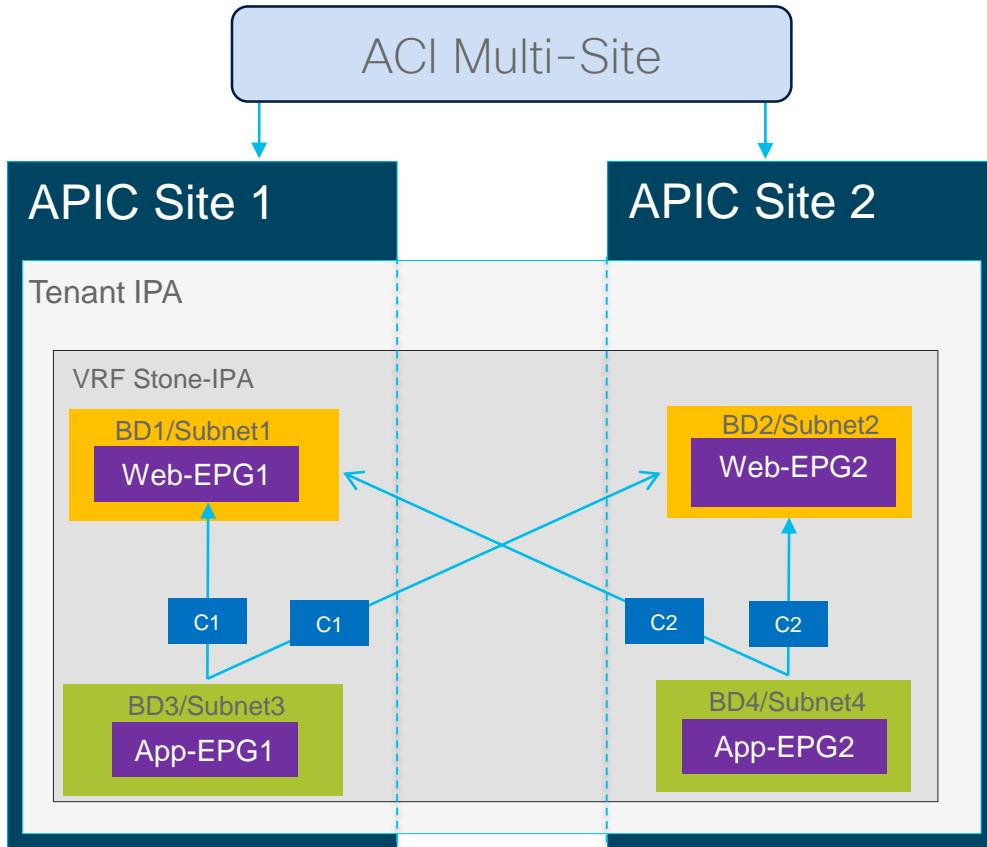
ACI Multi-Site

Common Functions Supported across All Use Cases in Phase 1

- Multi-Site User Authentication and RBAC rules
- Dashboard and health scores for inter-site policies
- Provisioning of inter-sites BGP-EVPN control plane and Spines' TEPS reachability
- Inter-Site Policies and Objects:
 - Tenant, AP, EPG, Contracts, Filters BD, VRF, Micro-Segmentation
- Operational view before pushing the policies to the sites
- Visualizer for inter-site policies
- Change management: Scope of change at site(s) level
- L4-L7 services and stitching are site local
- **L3Out are site local for GOLF or Border Leaf**
- **APIC will continue to configure and manage site local features (access policies, VMM domain,...)**

Multi-Site Use Case 1.1

Stretched VRF with Inter-Site Contracts

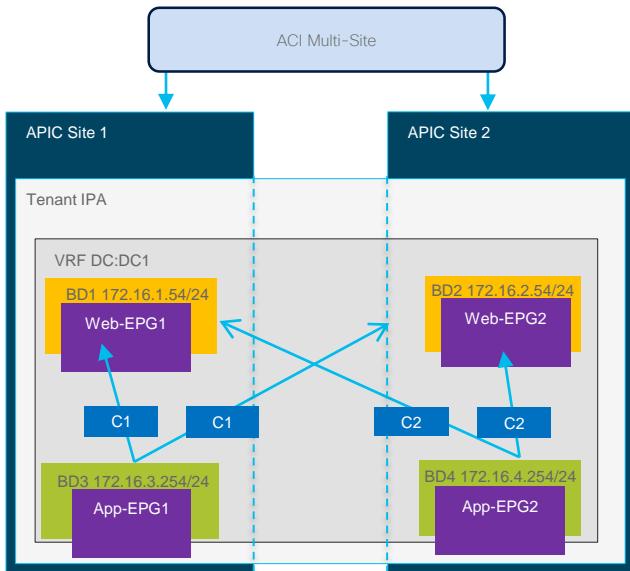
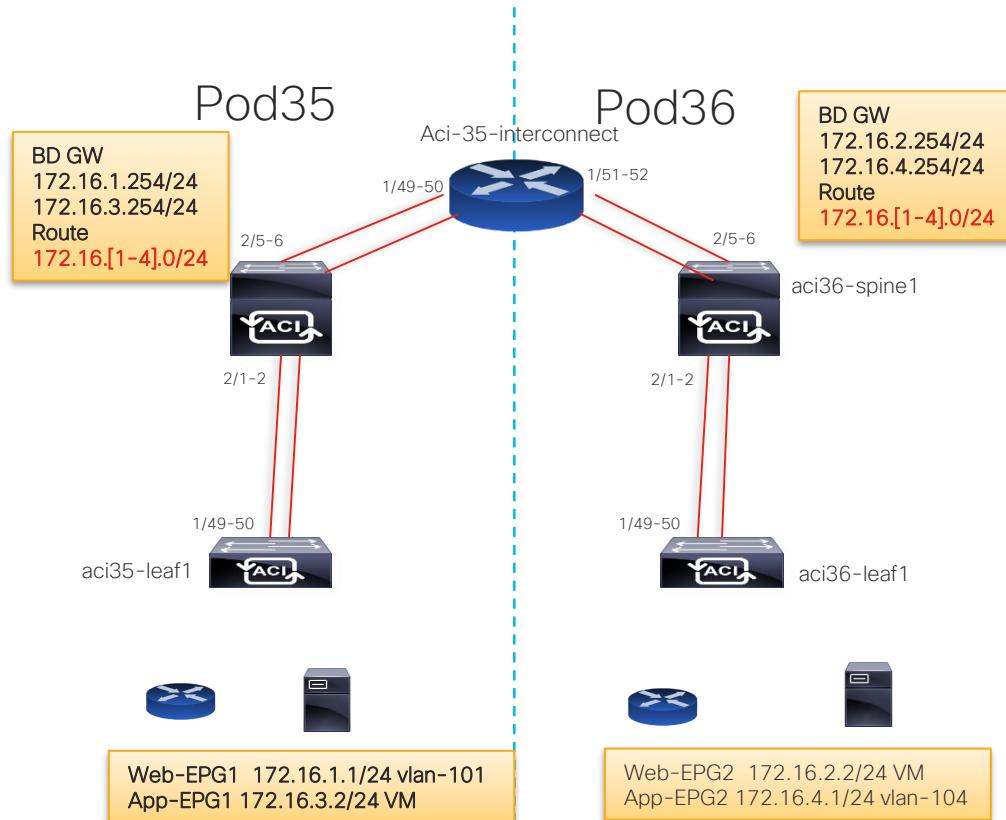


Use Case Properties

- Common ('stretched') objects across sites:
 - Tenant ID
 - VRF context
- Separate templates defined on Multi-Site for defining 'stretched' and 'site-local' objects
- Layer 3 only communication across sites
- Localization of fault domain and flooding
- Contracts are offered within a site or across sites to all consumer EPGs

1. Template 1: Tenant/VRF in template 1, push
2. Template 2: Then define site1 BDs/Subnets/EPGs and contract C1, push
3. Template 3: Then define site2 BDs/Subnets/EPGs and contract C2, push
4. Go back to template 2, reference C2 for cross-site contract, push
5. Go back to template 3, reference C1 for cross-site contract, push
6. Domain association is site level

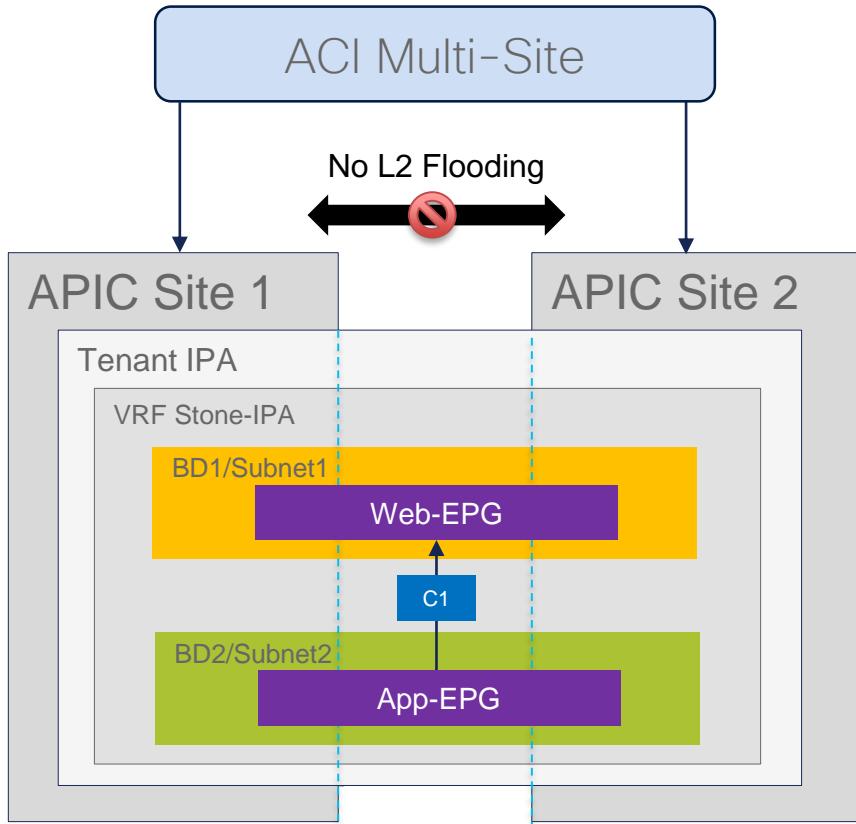
Use case 1.1 – lab VRF DC:DC1



Multi-Site Use Case 2

Stretched BD without L2 Broadcast Extension

Supported
At FCS



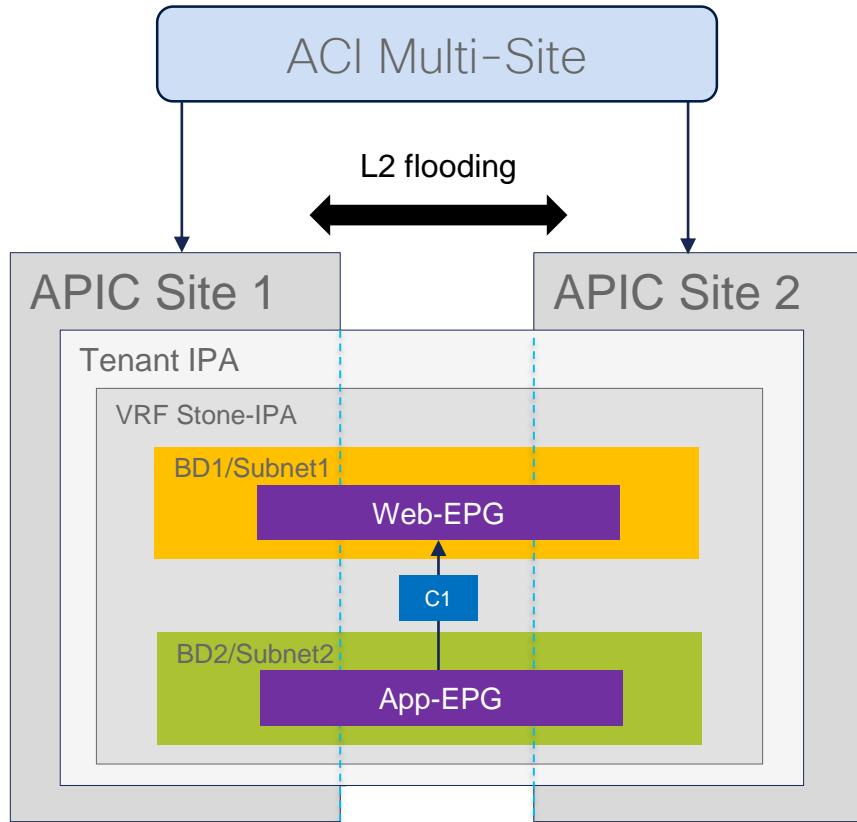
Use Case Properties

- Inter-Site IP Mobility for Disaster Recovery
- Objects stretched across sites:
 - Tenant ID
 - VRF context
 - BD/Subnet
 - Provider and Consumer EPGs
 - Policy between EPGs
- L2 flooding is disabled at the BD level
 - No L2 broadcast/multicast across sites → no clustering across sites
 - Unknown unicast traffic takes spine Proxy path for inter-site and in local site
 - ARP unicast mode for inter-site intra-subnet communication
- 'Cold' VM Migration supported

Multi-Site Use Case 3

Stretched BD with L2 Broadcast Extension

Supported
At FCS



Use Case Properties

- Active/Active deployment with inter-site Layer 2 extension
- Objects stretched across sites:
 - Tenant ID
 - VRF context
 - BD/Subnet
 - Provider and Consumer EPGs
 - Policy between EPGs
- L2 flooding enabled at the BD level
 - L2 BUM traffic forwarded over head-end replicated VXLAN tunnels
- L2 application clustering and 'live' VM migration

ACI Multi-Site

Caveats and Restrictions

- Each site must deploy a local L3Out connection
 - A site cannot provide transit routing services for a different site
 - Possible to share a pair of WAN edge routers across sites (traditional L3Out on Border Leaf nodes)
- Service chaining (service graph) cannot be deployed across sites
- Multi-Pod fabric is not initially supported as a site (supported since 3.2)
- Domain (VMM, Physical) definition and association is done at the site level
- Policies pushed to a site from Multi-Site can be locally modified in APIC

Translation troubleshooting Object Model

Multisite Translation Requirements

- In a typical APIC multisite environment, a particular site may expose certain services which other sites may desire to consume.
 - An example would be a shared DNS service provided by one of the sites which other sites want to use.
- The other common use case is to allow peer-to-peer communication between a pair of EPGs across sites.
 - This requires context stretching between the sites to give an illusion of same-context communication between the sites.

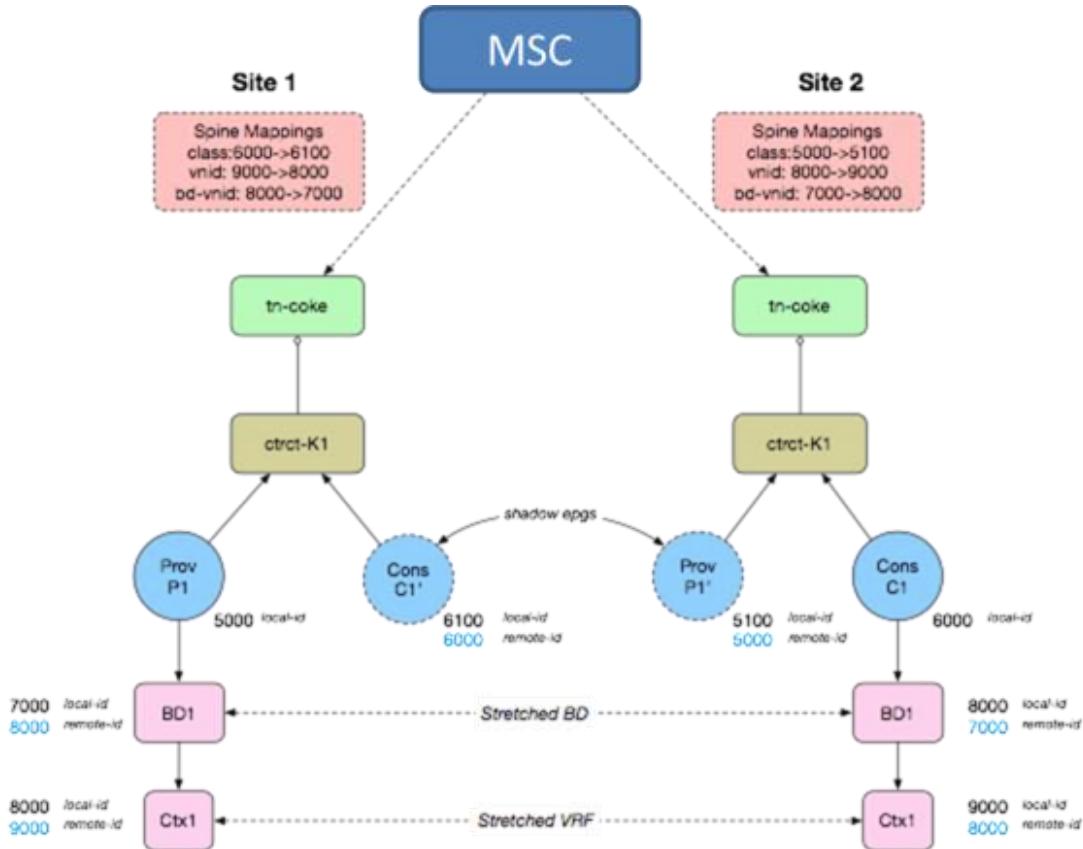
Translation Mechanism

- The policy mechanism we define to achieve this is *stretched EPgs, BDs, Contexts and InstPs*.
- MSC creates identical contracts on multiple sites with similar policy hierarchies and filters or make sure that the contract is replicated from one site to other sites.
- For the multisite controller, a possible deployment model could involve the user sanitizing the contract first on one site, and then using the controller to deploy the contract onto multiple sites.

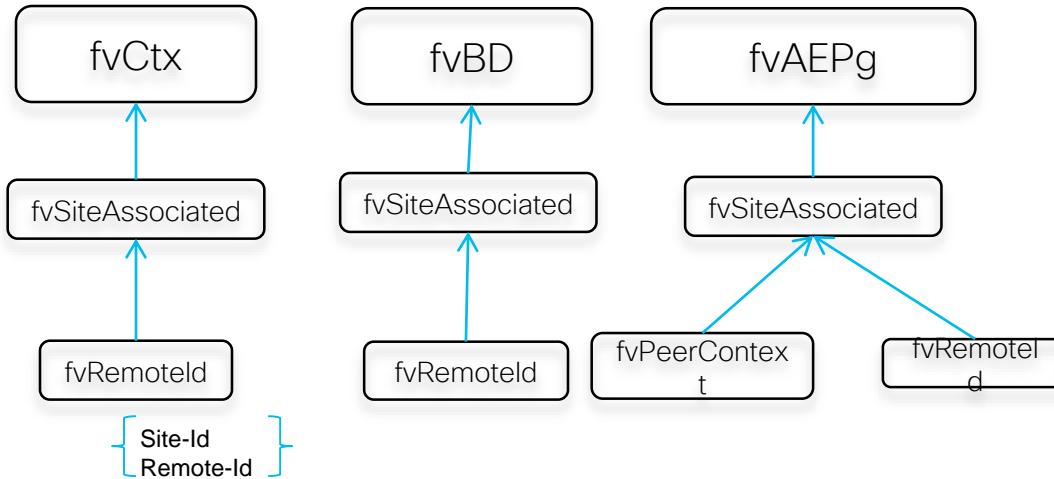
Translation Mechanism

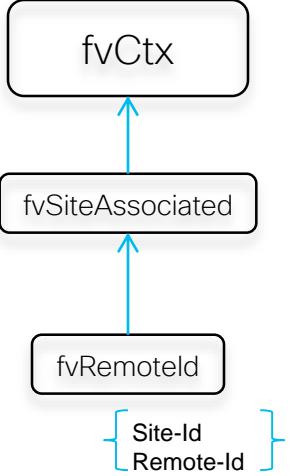
- The controller needs to be told which EPgs need to be stretched to other sites. It will then create identical EPgs on remote sites and associate them with the contract and configure the class ID mappings under it.
- Conceptually this can be viewed as *stretching* an EPg onto remote sites, though this EPg may not be physically deployed onto the ToR switches on remote sites.
- Note: Contracts should exist but EPGs specific actrlRules would exist on the ToR depending on Deployment Immediacy flag set to Immediate or Lazy.
- MSC has the flag present under static port deployment of EPGs.

BD Stretching



Logical Model





```

bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/ctx-L2/stAsc
Total Objects shown: 1

# fv.SiteAssociated
childAction   :
descr        :
dn           : uni/tn-RD-L2/ctx-L2/stAsc
lcOwn        : local
modTs        : 2018-05-03T03:14:39.572+00:00
monPolDn     : uni/tn-common/monepg-default
name          : msc-local
nameAlias    :
ownerKey    :
ownerTag    :
rn           : stAsc
siteId       : 1
status        :
uid          : 15374

```

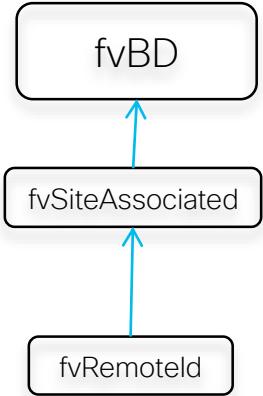
```

bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/ctx-L2/stAsc/site-2
Total Objects shown: 1

# fv.RemoteId
siteId       : 2
childAction  :
descr        :
dn           : uni/tn-RD-L2/ctx-L2/stAsc/site-2
lcOwn        : local
modTs        : 2018-05-03T03:14:40.895+00:00
monPolDn     : uni/tn-common/monepg-default
name          :
nameAlias    :
ownerKey    :
ownerTag    :
remoteCtxPcTag : 32770
remotePcTag  : 2162688
rn           : site-2
status        :
uid          : 15374

```

Logical BD – site 1



Site 2 BD

```
dsol-aci36-apic1# moquery -c fvBD -f 'fv.BD.seg  
="15073234"' | egrep "dn|scope|seg"  
dn : uni/tn-RD-L2/BD-Web  
scope : 2162688  
eg : 15073234
```

```
bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/BD-Web/stAsc/  
Total Objects shown: 1
```

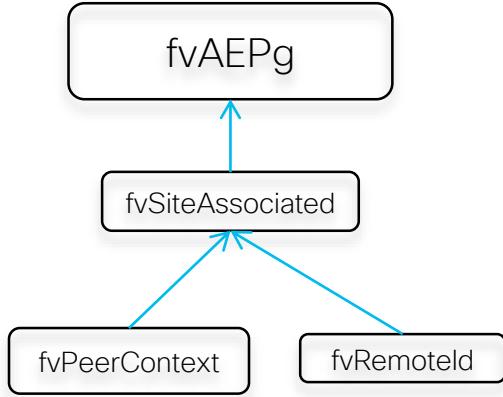
```
# fv.SiteAssociated  
childAction :  
descr :  
dn : uni/tn-RD-L2/BD-Web/stAsc/stAsc  
lcOwn : local  
modTs : 2018-05-03T03:14:39.572+00:00  
monPolDn : uni/tn-common/monepg-default  
name : msc-local  
nameAlias :  
ownerKey :  
ownerTag :  
rn : stAsc  
siteId : 1  
status :  
uid : 15374
```

```
bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/BD-Web/stAsc/site-2  
Total Objects shown: 1
```

```
# fv.RemoteId  
siteId : 2  
childAction :  
descr :  
dn : uni/tn-RD-L2/BD-Web/stAsc/site-2  
lcOwn : local  
modTs : 2018-05-03T03:14:40.895+00:00  
monPolDn : uni/tn-common/monepg-default  
name :  
nameAlias :  
ownerKey :  
ownerTag :  
remoteCtxPcTag : any  
remotePcTag : 15073234 Actually remote BD VNID  
rn : site-2  
status :  
uid : 15374
```

Logical AEPg – site

fvPeerContext only in case of shared services



Site 2 BD

```
bdsol-aci36-apic1# moquery -c fvAEPg -f 'fv.AEPg.pcTag=="49155"' | egrep "dn|scope|pcTag"
dn          : uni/tn-RD-L2/ap-App/epg-Web
pcTag       : 49155
scope        : 2162688
```

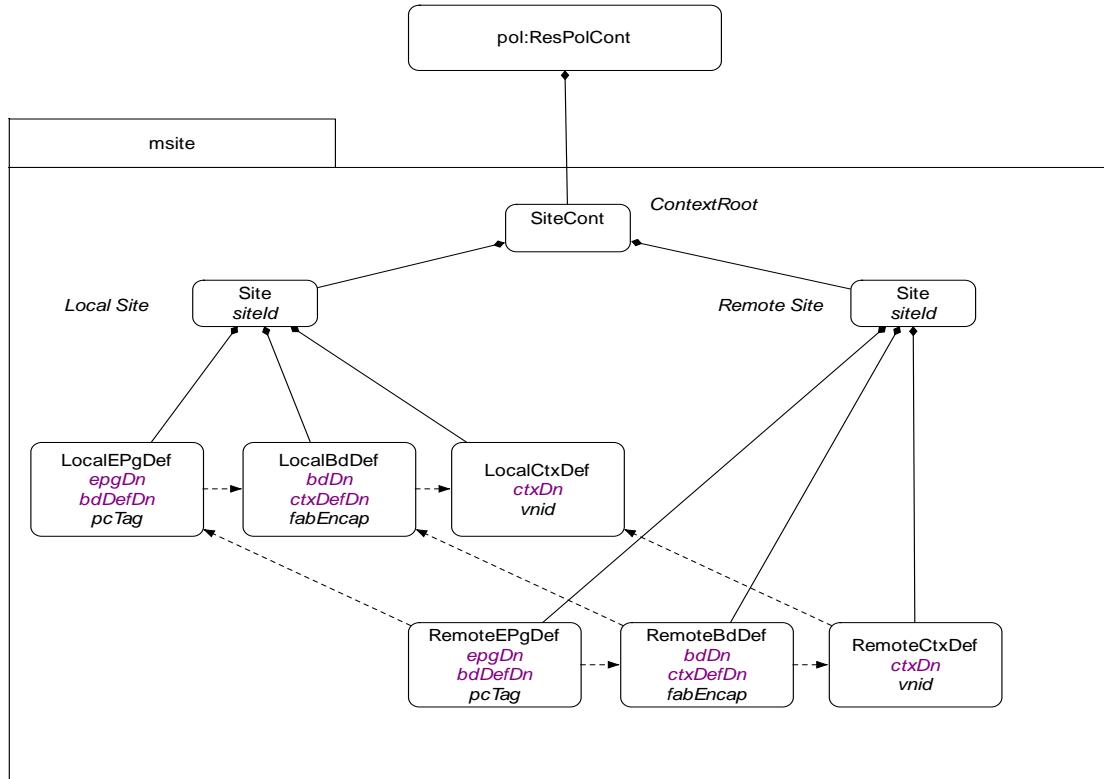
```
bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/ap-App/epg-Web/stAsc/
Total Objects shown: 1

# fv.SiteAssociated
childAction  :
descr        :
dn          : uni/tn-RD-L2/ap-App/epg-Web/stAsc/stAsc
lcOwn       : local
modTs        : 2018-05-03T03:14:39.572+00:00
monPolDn    : uni/tn-common/monepg-default
name         : msc-local
nameAlias   :
ownerKey    :
ownerTag    :
rn          : stAsc
siteId      : 1
status       :
uid         : 15374
```

```
bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/ap-App/epg-
Web/stAsc/site-2
Total Objects shown: 1

# fv.RemoteId
siteId      : 2
childAction  :
descr        :
dn          : uni/tn-RD-L2/ap-App/epg-Web/stAsc/site-2
lcOwn       : local
modTs        : 2018-05-03T03:14:40.895+00:00
monPolDn    : uni/tn-common/monepg-default
name         :
nameAlias   :
ownerKey    :
ownerTag    :
remoteCtxPcTag : any
remotePcTag  : 49155
rn          : site-2
status       :
uid         : 15374
```

Logical Model



```
# fv.Site
siteId      : 1
childAction :
dn          : resPolCont/sitecont/site-1/site-1
lcOwn       : local
modTs       : 2018-03-30T05:50:38.595+00:00
name        : msc-local
rn          : site-1
```

```
# fv.Site
siteId      : 2
childAction :
dn          : resPolCont/sitecont/site-2/site-2
lcOwn       : local
modTs       : 2018-03-30T05:50:38.595+00:00
name        :
rn          : site-2
```

```
# fv.LocalEPgDef
moDn        : uni/tn-RD-L2/ap-App/epg-Web
LocalBdDefDn : resPolCont/sitecont/site-1/localbddef-[uni/tn-RD-L2/BD-Web]
LocalCtxDefDn : resPolCont/sitecont/site-1/localctxdef-[uni/tn-RD-L2/ctx-L2]
LocalDefDn   : resPolCont/sitecont/site-1/localepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
childAction  :
dn          : resPolCont/sitecont/site-1/localepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
lcOwn       : local
modTs       : 2018-05-03T03:14:40.903+00:00
pcTag        : 32771
rn          : localepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
```

```
# fv.RemoteEPgDef
moDn        : uni/tn-RD-L2/ap-App/epg-Web
LocalDefDn  : resPolCont/sitecont/site-1/localepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
RemoteBdDefDn : resPolCont/sitecont/site-2/remotebddef-[uni/tn-RD-L2/BD-Web]
RemoteCtxDefDn : resPolCont/sitecont/site-2/remotectxdef-[uni/tn-RD-L2/ctx-L2]
childAction  :
dn          : resPolCont/sitecont/site-2/remoteepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
lcOwn       : local
modTs       : 2018-05-03T03:14:40.903+00:00
pcTag        : 49155
rn          : remoteepgdef-[uni/tn-RD-L2/ap-App/epg-Web]
```

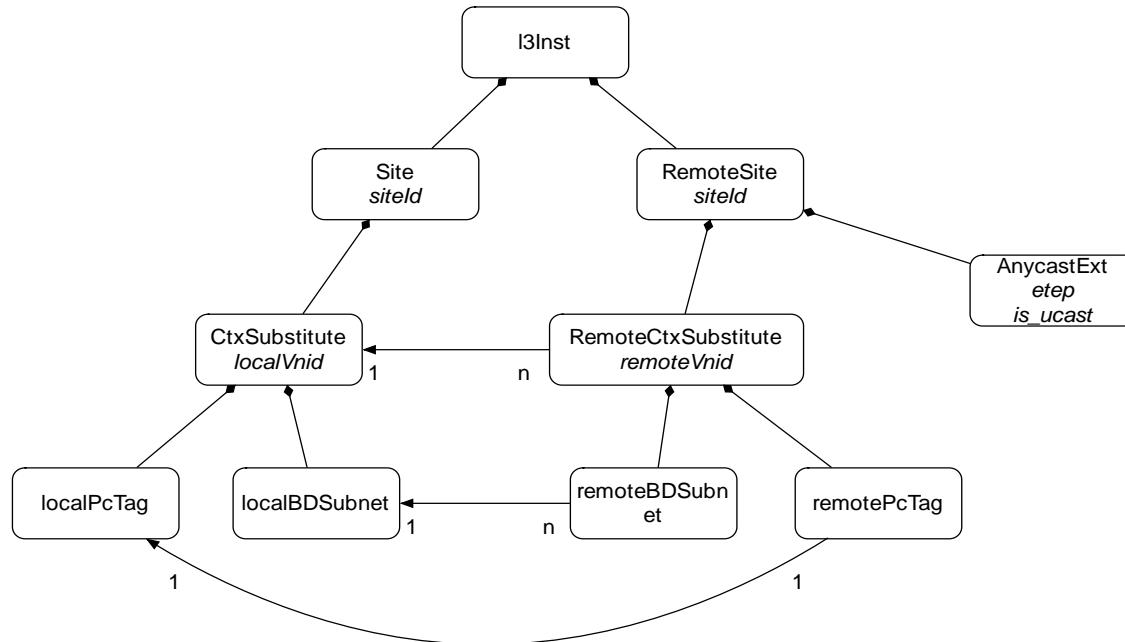
```
# fv.LocalBdDef
moDn        : uni/tn-RD-L2/BD-Web
LocalCtxDefDn : resPolCont/sitecont/site-1/localctxdef-[uni/tn-RD-L2/ctx-L2]
LocalDefDn   : resPolCont/sitecont/site-1/localbddef-[uni/tn-RD-L2/BD-Web]
bcastP      : 225.0.216.80
childAction  :
dn          : resPolCont/sitecont/site-1/localbddef-[uni/tn-RD-L2/BD-Web]
intersiteBumTrafficAllow : yes
intersiteL2Stretch : yes
lcOwn       : local
modTs       : 2018-05-03T03:14:40.903+00:00
pcTag        : 15204288
rn          : localbddef-[uni/tn-RD-L2/BD-Web]
```

```
# fv.RemoteBdDef
moDn        : uni/tn-RD-L2/BD-Web
LocalDefDn  : resPolCont/sitecont/site-1/localbddef-[uni/tn-RD-L2/BD-Web]
RemoteCtxDefDn : resPolCont/sitecont/site-2/remotectxdef-[uni/tn-RD-L2/ctx-L2]
childAction  :
dn          : resPolCont/sitecont/site-2/remotebddef-[uni/tn-RD-L2/BD-Web]
lcOwn       : local
modTs       : 2018-05-03T03:14:40.903+00:00
pcTag        : 15073234
rn          : remotebddef-[uni/tn-RD-L2/BD-Web]
status      :
```

```
# fv.LocalCtxDef
moDn        : uni/tn-RD-L2/ctx-L2
LocalDefDn  : resPolCont/sitecont/site-1/localctxdef-[uni/tn-RD-L2/ctx-L2]
childAction  :
dn          : resPolCont/sitecont/site-1/localctxdef-[uni/tn-RD-L2/ctx-L2]
lcOwn       : local
modTs       : 2018-05-03T03:14:40.901+00:00
pcTag        : 2457600
rn          : localctxdef-[uni/tn-RD-L2/ctx-L2]
```

```
# fv.RemoteCtxDef
moDn        : uni/tn-RD-L2/ctx-L2
LocalDefDn  : resPolCont/sitecont/site-1/localctxdef-[uni/tn-RD-L2/ctx-L2]
childAction  :
dn          : resPolCont/sitecont/site-2/remotectxdef-[uni/tn-RD-L2/ctx-L2]
lcOwn       : local
modTs       : 2018-05-03T03:14:40.901+00:00
pcTag        : 2162688
rn          : remotectxdef-[uni/tn-RD-L2/ctx-L2]
status      :
```

Concrete Model



```

# dci.LocalSite
id      : 1
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-
1/localSite-1
lcOwn   : local
modTs   : 2018-03-30T05:50:38.558+00:00
name    : msc-local
rn      : localSite-1

# 13.LocalCtxSubstitute
FabEncap  : vxlan-2457600
DnName    : uni/tn-RD-L2/ctx-L2
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/localSite-1/localCtxSubstitute-
[vxlan-2457600]
lcOwn   : local
mcastEncap : 0.0.0.0
modTs   : 2018-05-03T03:14:40.863+00:00
rn      : localCtxSubstitute-[vxlan-2457600]
status   :

# 13.RtToLocalCtxSubstitute
tDn      : topology/pod-1/node-201/sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-
[vxlan-2162688]
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/localSite-1/localCtxSubstitute-
[vxlan-2457600]/rttToLocalCtxSubstitute-[topology/pod-1/node-201/sys/inst-overlay-
1/remoteSite-2/remoteCtxSubstitute-[vxlan-2162688]]
rn      : rttoLocalCtxSubstitute-[topology/pod-1/node-201/sys/inst-overlay-
1/remoteSite-2/remoteCtxSubstitute-[vxlan-2162688]]
tC1     : 13RemoteCtxSubstitute

# 12.LocalBdSubstitute
FabEncap  : vxlan-15204288
DnName    : uni/tn-RD-L2/BD-Web
childAction :
ctrl     : bum-traffic
dn      : topology/pod-1/node-201/sys/inst-overlay-1/localSite-1/localCtxSubstitute-
[vxlan-2457600]/localBdSubstitute-[vxlan-15204288]
lcOwn   : local
mcastEncap : 225.0.216.80
rn      : localBdSubstitute-[vxlan-15204288]

# 12.LocalPcTagSubstitute
pcTag    : 32771
DnName    : uni/tn-RD-L2/ap-App/epg-Web
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/localSite-
1/localCtxSubstitute-[vxlan-2457600]/localPcTagSubstitute-32771
lcOwn   : local
modTs   : 2018-05-03T03:14:40.868+00:00
rn      : localPcTagSubstitute-32771

# dci.RemoteSite
id      : 2
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-
1/remoteSite-2
lcOwn   : local
modTs   : 2018-03-30T05:50:38.558+00:00
name    :
rn      : remoteSite-2

# 13.RemoteCtxSubstitute
FabEncap  : vxlan-2162688
DnName    : uni/tn-RD-L2/ctx-L2
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/remoteSite-
2/remoteCtxSubstitute-[vxlan-2162688]
lcOwn   : local
modTs   : 2018-05-03T03:14:40.863+00:00
rn      : remoteCtxSubstitute-[vxlan-2162688]
status   :

# 12.RemoteBdSubstitute
FabEncap  : vxlan-15073234
DnName    : uni/tn-RD-L2/BD-Web
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/remoteSite-
2/remoteCtxSubstitute-[vxlan-2162688]/remoteBdSubstitute-[vxlan-15073234]
lcOwn   : local
modTs   : 2018-05-03T03:14:40.868+00:00
rn      : remoteBdSubstitute-[vxlan-15073234]

# 12.RemotePcTagSubstitute
pcTag    : 49155
DnName    : uni/tn-RD-L2/ap-App/epg-Web
childAction :
dn      : topology/pod-1/node-201/sys/inst-overlay-1/remoteSite-
2/remoteCtxSubstitute-[vxlan-2162688]/remotePcTagSubstitute-49155
lcOwn   : local
modTs   : 2018-05-03T03:14:40.868+00:00
rn      : remotePcTagSubstitute-49155

```

Child of local VRF Concrete.
 tDn points to remote VRF concrete
 

Troubleshooting

- Check any Failed Translations.

Filter

Class or DN: fvFailedXlate

Property: Op: == Val1: Val2:

Run Query

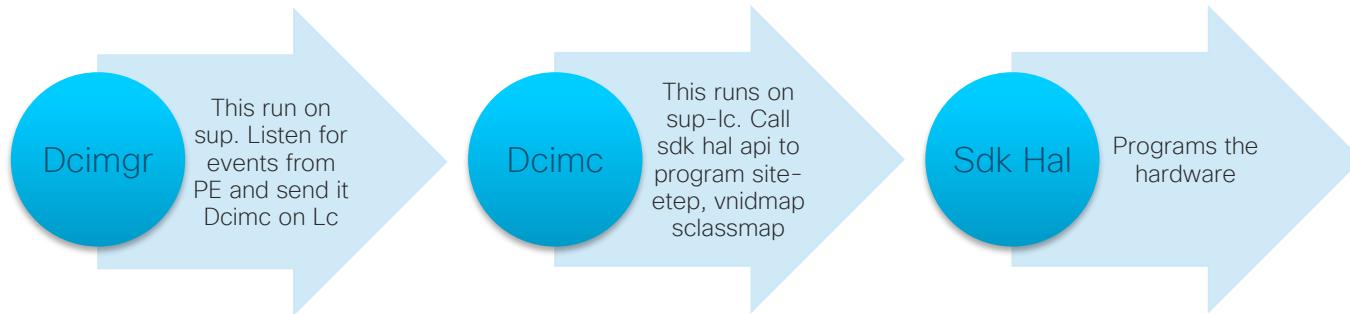
[Display URI of last query](#)

[Display last response](#)

fvFailedXlate	
childAction	
descr	
dn	uni/failedxlatecont/failedxlate-[resPolCont/sitecont/site-2/remoteepgdef-[uni/tn-coke1/ap-ap0/epg-epg1]]
lcOwn	local
moDn	resPolCont/sitecont/site-2/remoteepgdef-[uni/tn-coke1/ap-ap0/epg-epg1]
modTs	2017-08-07T23:07:27.441+00:00
name	
nameAlias	
ownerKey	
ownerTag	
status	

Translation troubleshooting
on switch (nxos) – dcimgr

Process involved for vnidmap/sclass/site-ete



Dcimgr/dcimc/sdkTraces for sclass/vnid map

- Dcimgr (on sup)
 - show dcimgr internal event-history events

And log file :

```
pod35-spine1# ls -al /var/sysmgr/tmp_logs/dcimgr.log
-rw-rw-rw- 1 root root 3162338 May  2 16:37 /var/sysmgr/tmp_logs/dcimgr.log
pod35-spine1#
```

- HAL CLI :

```
module-2# show platform internal hal objects dci ?
      all          Dump All HAL objects
      remotesite   Remotesite or wan instance
      remotesiteetep Unicast etep that belongs to this remotesite
      remotevrfvnid Vrf for remotesite object
      sclassmap    Sclass mapping for remotesite vrf
      vnidmap      Vnid mapping for remotesite object
```

Dcimgr trace

Dcimgr consume concrete object to create translation

```
pod35-spine1# show dcimgr internal event-history events | egrep -A 1 Event
1) Event:E_DEBUG, length:150, at 135850 usecs after Tue Apr 24 14:40:18 2018
   [1835623268] gr_objstore_bdvnid_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-2490368]/localBdSubstitute-[vxlan-15040468]
---
2) Event:E_DEBUG, length:278, at 135813 usecs after Tue Apr 24 14:40:18 2018
   [1835623268] gr_objstore_bdvnid_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-2490368]/localBdSubstitute-[vxlan-15040468]/rttoLocalBdSubstitute-[sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2392064]/remoteBdSubstitute-[vxlan-15335344]]
---
3) Event:E_DEBUG, length:265, at 277412 usecs after Tue Apr 24 14:38:04 2018
   [1835623268] gr_objstore_sclass_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-2490368]/localPcTagSubstitute-16386/rttoLocalPcTagSubstitute-[sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2392064]/remotePcTagSubstitute-16386]
---
4) Event:E_DEBUG, length:206, at 276933 usecs after Tue Apr 24 14:38:04 2018
   [1835623268] gr_objstore_vnid_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-2490368]/rttoLocalCtxSubstitute-[sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2392064]]
---
5) Event:E_DEBUG, length:122, at 275994 usecs after Tue Apr 24 14:38:04 2018
   [1835623268] gr_objstore_remote_vrf_vnid_mts_hdlr: [Create]: sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2392064]
---
6) Event:E_DEBUG, length:150, at 622469 usecs after Mon Apr 23 14:48:38 2018
   [1835623268] gr_objstore_bdvnid_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-3014656]/localBdSubstitute-[vxlan-16056262]
---
7) Event:E_DEBUG, length:278, at 622447 usecs after Mon Apr 23 14:48:38 2018
   [1835623268] gr_objstore_bdvnid_map_mts_hdlr: [Create]: sys/inst-overlay-1/localSite-1/localCtxSubstitute-[vxlan-3014656]/localBdSubstitute-[vxlan-16056262]/rttoLocalBdSubstitute-[sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]/remoteBdSubstitute-[vxlan-15794151]]
```

DCI mgr – xlate with hex translation !

Vnid translate (vrf and bd)

```
pod36-spine1# show dcimgr repo vnid-maps detail
```

site	Remote		Local		
	Vrf	Bd	Vrf	Bd	Rel-state
<hr/>					
1	2981888 0x2d8000		2293760 0x230000		[formed]
1	2981888 0x2d8000	16678778 0xfe7f7a	2293760 0x230000	16154554 0xf67fba	[formed]
1	3014656 0x2e0000		2457600 0x258000		[formed]

pcTag (sclass) translate

```
pod36-spine1# show dcimgr repo sclass-maps detail
```

site	Remote		Local		
	Vrf	PcTag	Vrf	PcTag	Rel-state
1	2981888 0x2d8000	49153 0xc001	2293760 0x230000	49153 0xc001	[formed]
1	2981888 0x2d8000	49154 0xc002	2293760 0x230000	49155 0xc003	[formed]
1	2981888 0x2d8000	16387 0x4003	2293760 0x230000	16386 0x4002	[formed]
1	3014656 0x2e0000	49153 0xc001	2457600 0x258000	49153 0xc001	[formed]
1	3014656 0x2e0000	16387 0x4003	2457600 0x258000	32772 0x8004	[formed]

HAL - dci vnid Xlate

Tabular view

```
module-2# show platform internal hal dci vnidmap
```

Non-Sandbox Mode

Sandbox_ID: 0 Asic Bitmap: 0x0

Site ID	POD ID	isBD	Local		Remote		----EPG table----		-BDState Table	
			vnid	vnid	vnid	idx	Localvnid	idx	isBD	
2	2	0	2457600	2162688	15372	2457600	15372	0		
2	2	1	16285610	16482195	15377	16285610	15377	1		
2	2	1	16056263	16121790	15364	16056263	15364	1		
2	2	1	15925206	16220082	15366	15925206	15366	1		
2	2	1	15040468	15335344	15371	15040468	15371	1		
2	2	0	2162689	2949121	15376	2162689	15376	0		

Detail object

```
module-2# show platform internal hal objects dci vnidmap
## Get Objects for dci vnidmap for Asic 0
```

OBJECT 1:

Handle	:	339894
isbdvnid	:	Enabled
localvnid	:	0xf87faa
localgipo	:	225.1.36.0/32
remotevnid	:	0xfb7f93
remotevrfnid	:	0x2d0001
islocalbdctrl	:	Enabled
siteid	:	0x2

HAL - dci sclass translate

Tabular view

```
module-2# show platform internal hal dci sclassmap  
Non-Sandbox Mode
```

```
Sandbox_ID: 0 Asic Bitmap: 0x0
```

--- DCI Sclass table ---						
Site ID	Remote Vnid	Local Sclass	Remote Sclass	Remote Sclass	Local Sclass	Local Scope
2	2293760	16387	16386	16386	16387	1
2	2490368	32772	16389	16389	32772	6
2	2162688	49154	16386	16386	49154	5
2	2392064	16386	16386	16386	16386	4
2	2293760	49154	49155	49155	49154	1
2	2686976	32770	49153	49153	32770	3

Detail object

```
module-2# show platform internal hal objects dci sclassmap  
## Get Objects for dci sclassmap for Asic 0
```

```
OBJECT 0:  
Handle : 38469  
localsclass : 0x4003  
remotesclass : 0x4002  
remotevnid : 0x230000  
siteid : 0x2
```

BGP route exchange detail

BGP VNI

- Route Exchange issues can be seen either in the source or on the remote site.
 - Check if the BGP MOs are created for VNIs/RTs and RDs correctly. These MOs are created only on spines in every site. These MOs are created when the VRF/BD/EPGs are stretched or the contracts are created at EPG level
- Following shows mapping of BGP VNIDs and what routes are requested from COOP and why they are used:

	VNID	Request Route Type from COOP	What routes are advertised	Why
1	VRF	IP Only with reserved MAC (0200.0000.0002)	Only Loop Back/SVI IPs	For inband management
2	BD	MAC-IP routes MAC routes only if L2 is Stretched	All EPs	For EP-EP communication

Spine VNI directory MO

- Under /mit/sys/bgp/inst/encapgroupevi-1 (class bgp.EncapGroupEvi)
- We have one MO (directory for each translation VRF to VRF or VRF/BD to VRF/BD)

```
pod35-spine1# pwd  
/mit/sys/bgp/inst/encapgroupevi-1  
pod35-spine1# ls -a | egrep vni  
vni-bd-vrf-[vxlan-2162689]-bd-[vxlan-16285610]-epg-[unknown]  
vni-bd-vrf-[vxlan-2228225]-bd-[vxlan-16449431]-epg-[unknown]  
vni-bd-vrf-[vxlan-2228225]-bd-[vxlan-16547723]-epg-[unknown]  
vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-14712828]-epg-[unknown]  
vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-15204288]-epg-[unknown]  
vni-bd-vrf-[vxlan-2490368]-bd-[vxlan-15040468]-epg-[unknown]  
vni-bd-vrf-[vxlan-2981888]-bd-[vxlan-16678778]-epg-[unknown]  
vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-15925206]-epg-[unknown]  
vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16056262]-epg-[unknown]  
vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16056263]-epg-[unknown]  
vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]  
vni-bd-vrf-[vxlan-3112960]-bd-[vxlan-16449430]-epg-[unknown]  
vni-bd-vrf-[vxlan-3112960]-bd-[vxlan-16613250]-epg-[unknown]  
vni-vrf-vrf-[vxlan-2162689]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-2228225]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-2457600]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-2490368]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-2981888]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-3014656]-bd-[unknown]-epg-[unknown]  
vni-vrf-vrf-[vxlan-3112960]-bd-[unknown]-epg-[unknown]
```

BGP MO for VNI - VRF only (site 1 and site 2)

```
pod35-spine1# moquery -d sys/bgp/inst/encapgroupevi-1/vni-vrf-vrf-[vxlan-3014656]-bd-[unknown]-epg-[unknown]
```

```
Total Objects shown: 1
```

```
# bgp.Vni
type          : vrf
vrfVnid       : vxlan-3014656
bdVnid        : unknown
epgVnid       : unknown
bgpCfgFailedBmp   :
bgpCfgFailedTs  : 00:00:00:00.000
bgpCfgState     : 0
childAction     :
dn             : sys/bgp/inst/encapgroupevi-1/vni-vrf-vrf-[vxlan-3014656]-bd-[unknown]-epg-[unknown]
l2Stretch       : disabled
lcOwn          : local
modTs          : 2018-04-11T04:06:17.167+00:00
name           :
rd             : rd:as2-nn4:1:19791872
rn             : vni-vrf-vrf-[vxlan-3014656]-bd-[unknown]-epg-[unknown]
```

```
pod36-spine1# moquery -d sys/bgp/inst/encapgroupevi-1/vni-vrf-vrf-[vxlan-2457600]-bd-[unknown]-epg-[unknown]
```

```
Total Objects shown: 1
```

```
# bgp.Vni
type          : vrf
vrfVnid       : vxlan-2457600
bdVnid        : unknown
epgVnid       : unknown
bgpCfgFailedBmp   :
bgpCfgFailedTs  : 00:00:00:00.000
bgpCfgState     : 0
childAction     :
dn             : sys/bgp/inst/encapgroupevi-1/vni-vrf-vrf-[vxlan-2457600]-bd-[unknown]-epg-[unknown]
l2Stretch       : disabled
lcOwn          : local
modTs          : 2018-04-11T04:06:11.695+00:00
name           :
rd            : rd:as2-nn4:1:36012032
rn             : vni-vrf-vrf-[vxlan-2457600]-bd-[unknown]-epg-[unknown]
```

BGP MO for VNI – BD/VRF (site 1 and site 2)

BGP EVI contains the RD used to send Prefix for that BD /VRF

```
pod35-spine1# moquery -d sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]
Total Objects shown: 1

# bgp.Vni
type          : bd
vrfVnid        : vxlan-3014656
bdVnid         : vxlan-16351138
epgVnid        : unknown
bgpCfgFailedBmp :
bgpCfgFailedTs : 00:00:00:00.000
bgpCfgState    : 0
childAction    :
dn             : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]
l2Stretch       : enabled
lcOwn          : local
modTs          : 2018-04-11T04:28:21.600+00:00
name           :
rd             : rd:as2-nn4:1:33128354
rn             : vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]
status         :
```

```
pod36-spine1# moquery -d sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-16121791]-epg-[unknown]
Total Objects shown: 1

# bgp.Vni
type          : bd
vrfVnid        : vxlan-2457600
bdVnid         : vxlan-16121791
epgVnid        : unknown
bgpCfgFailedBmp :
bgpCfgFailedTs : 00:00:00:00.000
bgpCfgState    : 0
childAction    :
dn             : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-16121791]-epg-[unknown]
l2Stretch       : enabled
lcOwn          : local
modTs          : 2018-04-11T04:28:16.142+00:00
name           :
rd             : rd:as2-nn4:1:49676223
rn             : vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-16121791]-epg-[unknown]
status         :
```

BGP Route Target (same Stretched BD as previous slide)

Pod 35 spine Import RT

```
# bgp.RttEntry
rtt      : route-target:as2-nn4:136:49676223
childAction :
dn       : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]/rtp-import/ent-route-target:as2-nn4:136:49676223
lcOwn   : local
modTs   : 2018-04-11T04:28:21.600+00:00
rn      : ent-route-target:as2-nn4:136:49676223
status  :
```

Pod 36 spine Import RT

```
# bgp.RttEntry
rtt      : route-target:as2-nn4:135:33128354
childAction :
dn       : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-16121791]-epg-[unknown]/rtp-import/ent-route-target:as2-nn4:135:33128354
lcOwn   : local
modTs   : 2018-04-11T04:28:16.142+00:00
rn      : ent-route-target:as2-nn4:135:33128354
status  :
```

Pod 35 spine export RT

```
# bgp.RttEntry
rtt      : route-target:as2-nn4:135:33128354
childAction :
dn       : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-3014656]-bd-[vxlan-16351138]-epg-[unknown]/rtp-export/ent-route-target:as2-nn4:135:33128354
lcOwn   : local
modTs   : 2018-04-11T04:28:21.600+00:00
rn      : ent-route-target:as2-nn4:135:33128354
status  :
```

Pod 36 spine export RT

```
# bgp.RttEntry
rtt      : route-target:as2-nn4:136:49676223
childAction :
dn       : sys/bgp/inst/encapgroupevi-1/vni-bd-vrf-[vxlan-2457600]-bd-[vxlan-16121791]-epg-[unknown]/rtp-export/ent-route-target:as2-nn4:136:49676223
lcOwn   : local
modTs   : 2018-04-11T04:28:16.142+00:00
rn      : ent-route-target:as2-nn4:136:49676223
status  :
```

BGP EVI check (NXOS) – BD on site 1

Use `show bgp internal evi xx` to verify
RD and RT exp/import (where xx is BD VNID)
(kind of similar to show bgp process for GOLF)

```
pod35-spine1# show bgp internal evi 16351138

...
*****
BGP L2VPN/EVPN RD Information for 1:33128354
    L2VNI ID : 16351138 (vni_16351138)
    #Prefixes Local/BRIB : 2 / 2
    #Paths L3VPN->EVPN/EVPN->L3VPN : 0 / 0
*****
=====

BGP Configured VNI Information:
    VNI ID (Index) : 16351138 (0)
    RD : 1:33128354
    Export RTs : 1
        Export RT cfg list: 135:33128354(refcount:1)
    Import RTs : 1
        Import RT cfg list: 136:49676223(refcount:1)
    Topo Id : 16351138
    VTEP IP : 0.0.0.0
    VTEP VPC IP : 0.0.0.0
    Enabled : Yes
    Delete Pending : No
    RD/Import RT/Export RT : Yes/Yes/Yes
    Type : 3
    Usage : 2
    L2 stretch enabled : 1
    VRF Vnid : 3014656
    Refcount : 00000003
    Encap : VxLAN
=====
*****
```

```
+++++
BGP VNI Information for vni_16351138
    L2VNI ID : 16351138 (vni_16351138)
    RD : 1:33128354
    VRF Vnid : 3014656
    Prefixes (local/total) : 2/2
    VNIID registered with COOP : Yes
    Enabled : Yes
    Delete pending : 0
    Stale : No
    Import pending : 0
    Import in progress : 0
    Encap : VxLAN
    Topo Id : 16351138
    VTEP IP : 0.0.0.0
    VTEP VPC IP : 0.0.0.0
    Active Export RTs : 1
    Active Export RT list : 135:33128354
    Config Export RTs : 1
        Export RT cfg list: 135:33128354(refcount:1)
    Export RT chg/chg-pending : 0/0
    Active Import RTs : 1
    Active Import RT list : 136:49676223
    Config Import RTs : 1
        Import RT cfg list: 136:49676223(refcount:1)
    Import RT chg/chg-pending : 0/0
    IMET Reg/Unreg from L2RIB : 1/0
    MAC Reg/Unreg from L2RIB : 1/0
    MAC IP Reg/Unreg from L2RIB : 1/0
    IP-only Reg/Unreg from L2RIB : 0/0
    SMAD Reg/Unreg from L2RIB : 1/0
    IMET Add/Del from L2RIB : 0/0
    MAC Add/Del from L2RIB : 3/2
    MAC IP Add/Del from L2RIB : 3/2
    SMAD Add/Del from L2RIB : 0/0
    IMET Dnld/Wdraw to L2RIB : 0/0
    IMET Dnld/Wdraw to L2RIB failures : 0/0
    MAC Dnld/Wdraw to L2RIB : 0/0
    MAC Dnld/Wdraw to L2RIB failures : 0/0
    SMAD Dnld/Wdraw to L2RIB : 0/0
    SMAD Dnld/Wdraw to L2RIB failures : 0/0
=====
*****
```

Note the EVI number if the BD VNID we are looking for

BGP EVI check (NXOS) – BD on site 2

Use `show bgp internal evi xx` to verify
RD and RT exp/import (where xx is BD VNID)
(kind of similar to show bgp process for GOLF)

```
pod36-spine1# show bgp internal evi 16121791

.
*****
BGP L2VPN/EVPN RD Information for 1:49676223
    L2VNI ID : 16121791 (vni_16121791)
    #Prefixes Local/BRIB : 0 / 2
    #Paths L3VPN->EVPN/EVPN->L3VPN : 0 / 0
*****
=====
BGP Configured VNI Information:
    VNI ID (Index) : 16121791 (0)
    RD : 1:49676223
    Export RTs : 1
        Export RT cfg list: 136:49676223(refcount:1)
    Import RTs : 1
        Import RT cfg list: 135:33128354(refcount:1)
    Topo Id : 16121791
    VTEP IP : 0.0.0.0
    VTEP VPC IP : 0.0.0.0
    Enabled : Yes
    Delete Pending : No
    RD/Import RT/Export RT : Yes/Yes/Yes
    Type : 3
    Usage : 2
    L2 stretch enabled : 1
    VRF Vnid : 2457600
    Refcount : 00000003
    Encap : VxLAN
```

```
+++++
BGP VNI Information for vni_16121791
    L2VNI ID : 16121791 (vni_16121791)
    RD : 1:49676223
    VRF Vnid : 2457600
    Prefixes (local/total) : 0/2
    VNIID registered with COOP : Yes
    Enabled : Yes
    Delete pending : 0
    Stale : No
    Import pending : 0
    Import in progress : 0
    Encap : VxLAN
    Topo Id : 16121791
    VTEP IP : 0.0.0.0
    VTEP VPC IP : 0.0.0.0
    Active Export RTs : 1
    Active Export RT list : 136:49676223
    Config Export RTs : 1
        Export RT cfg list: 136:49676223(refcount:1)
        Export RT chg/chg-pending : 0/0
        Active Import RTs : 1
        Active Import RT list : 135:33128354
        Config Import RTs : 1
            Import RT cfg list: 135:33128354(refcount:1)
            Import RT chg/chg-pending : 0/0
            IMET Reg/Unreg from L2RIB : 1/0
            MAC Reg/Unreg from L2RIB : 1/0
            MAC IP Reg/Unreg from L2RIB : 1/0
            IP-only Reg/Unreg from L2RIB : 0/0
            SMAD Reg/Unreg from L2RIB : 1/0
            IMET Add/Del from L2RIB : 0/0
            MAC Add/Del from L2RIB : 0/0
            MAC IP Add/Del from L2RIB : 0/0
            SMAD Add/Del from L2RIB : 0/0
            IMET Dnld/Wdraw to L2RIB : 0/0
            IMET Dnld/Wdraw to L2RIB failures : 0/0
            MAC Dnld/Wdraw to L2RIB : 11/10
            MAC Dnld/Wdraw to L2RIB failures : 0/0
            SMAD Dnld/Wdraw to L2RIB : 0/0
            SMAD Dnld/Wdraw to L2RIB failures : 0/0
            MAC-IP/SMAD Msite-RD routes : 2
            MAC-IP WAN-RD routes : 0
            MAC-IP network host routes : 0
            Type : 3
```

Global EVI (BGP COOP registration)

- Global EVI is EVI 1
- All BGP path are imported in l2 vni1 which is the one registered with coop (l2rib)

```
pod35-spine1# show bgp internal evi 1
*****
Global EVI : 1
Number of EVI : 1
L2RIB bound / VNI Req to L2RIB : Yes / 1
VNI Adds / Dels from L2RIB : 34 / 13
Topo global/mpod/wan/avs/msite reg pending: 0/0/0/0/0
Topo global/mpod/wan/avs/msite registered: 1/0/0/0/1
L2RIB is up/registered/local-req: 1/1
L2RIB down: in-prg/up-defer: 0/0
L2RIB register/failures: 1/0
L2RIB deregister/failures: 0/0
L2RIB flow control (#enabled/#disabled): Disabled (0/0)
*****
...
=====

BGP Configured VNI Information:
  VNI ID (Index)          : 1 (0)
  RD                      : 10.10.35.101:135
  Export RTs              : 1
    Export RT cfg list: 5:16(refcount:1
  Import RTs              : 1
    Import RT cfg list: 5:16(refcount:1
  Topo Id                : 134217729
  VTEP IP               : 10.10.35.101
  VTEP VPC IP            : 0.0.0.0
  Enabled                 : Yes
  Delete Pending          : No
  RD/Import RT/Export RT : Yes/Yes/Yes
  Type                   : 1
  Usage                  : 4
  L2 stretch enabled      : 1
  VRF Vnid               : 60
  Refcount               : 00000003
  Encap                  : VxLAN
```

Other BGP useful command

- *show bgp internal node-role*
- *show bgp internal mpod-msite-map*

```
pod35-spine1# show bgp internal node-role
Node role : : MSITE_SPEAKER
pod35-spine1# show bgp internal mpod-msite-map
Anycast Mpod->Msite map:
Mpod Address          Msite Address          Mpod ID
10.10.35.101           10.10.35.101           0
```

Coop internal info l2rib

- There is a COOP topology for each VNID registered to coop
- Highlighted here **0xf97fa2 = 16351138** (VNID from the BD in site 1 in previous example slides.)

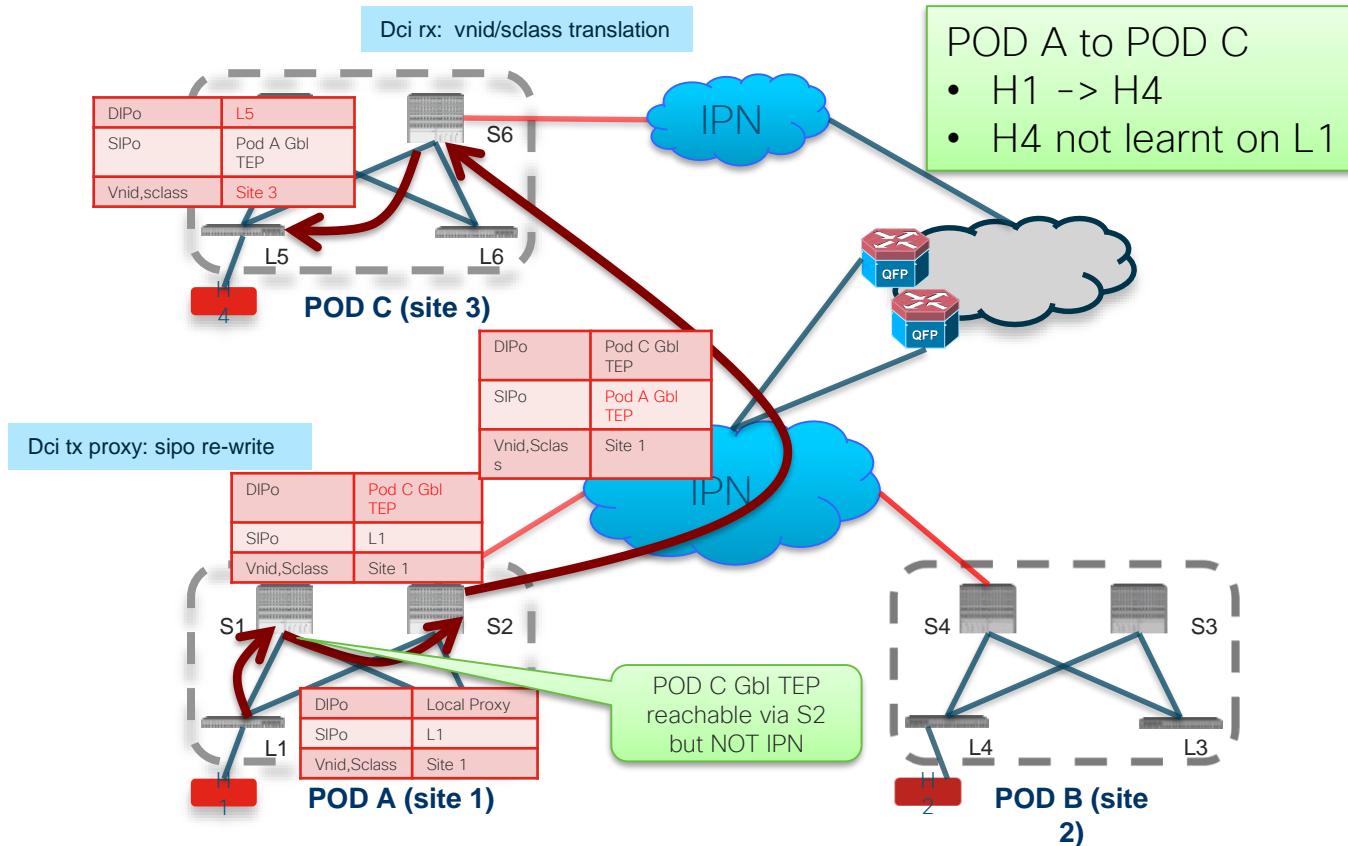
```
pod35-spine1# show coop internal info l2rib
ESTEP : 10.10.35.101 MSITE ESTEP : 10.10.35.101
MPOD shard - L2RIB shard [0, ffffffff]
MSITE shard - L2RIB shard [0, ffffffff]
GOLF shard - L2RIB shard [0, 0]
EXT_SRC shard - L2RIB shard [0, 0]
L2RIB topologies registered in COOP:
  Topo    1000000 route type 0x          40
  Topo    260000  route type 0x          440
  Topo    f4ffc6  route type 0x          145
  Topo    e57fd4  route type 0x          145
  Topo    faff97  route type 0x          144
...
  Topo    fe7f7a  route type 0x          145
  Topo    f97fa2 route type 0x          145

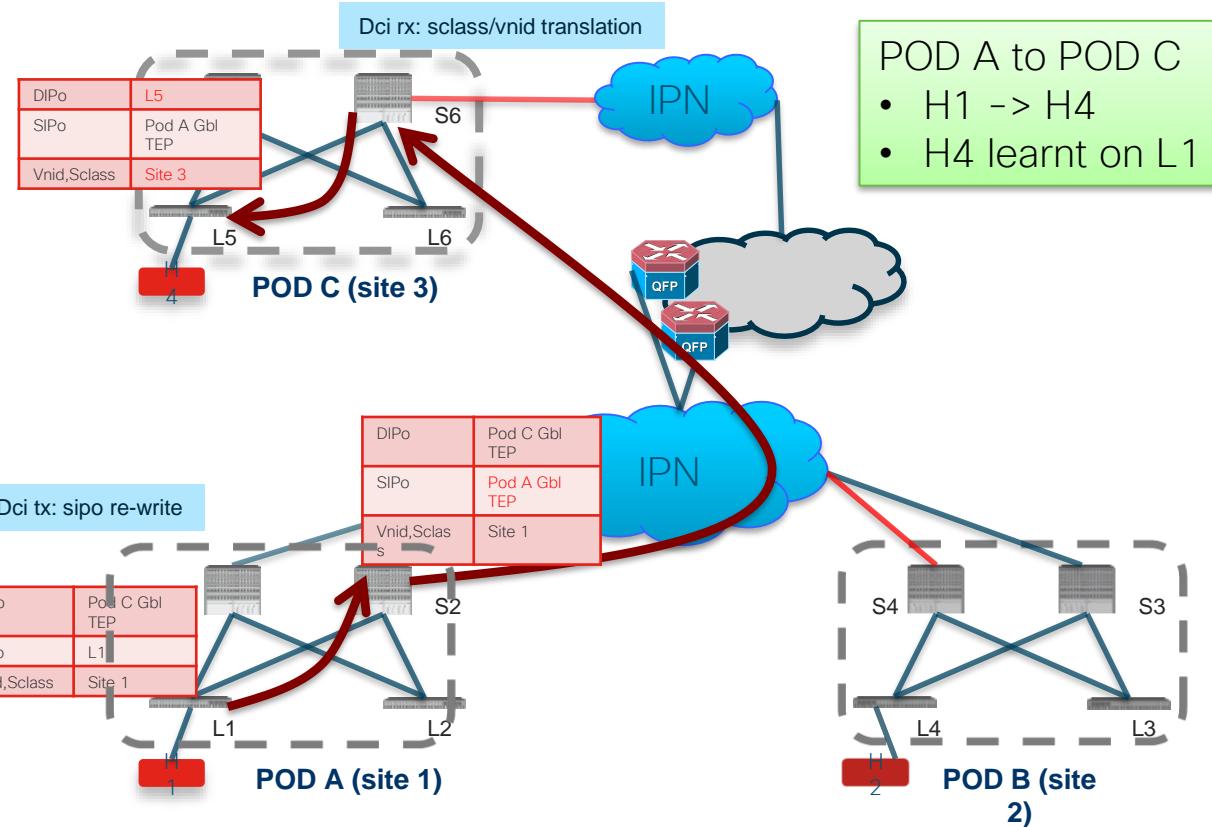
L2RIB topologies importing to COOP:
  Topo    1000000
  Topo    260000
  Topo    f4ffc6
  Topo    e57fd4
  Topo    faff97
  Topo    e7ffc0
...
  Topo    f97fa2
```

Unicast forwarding across site

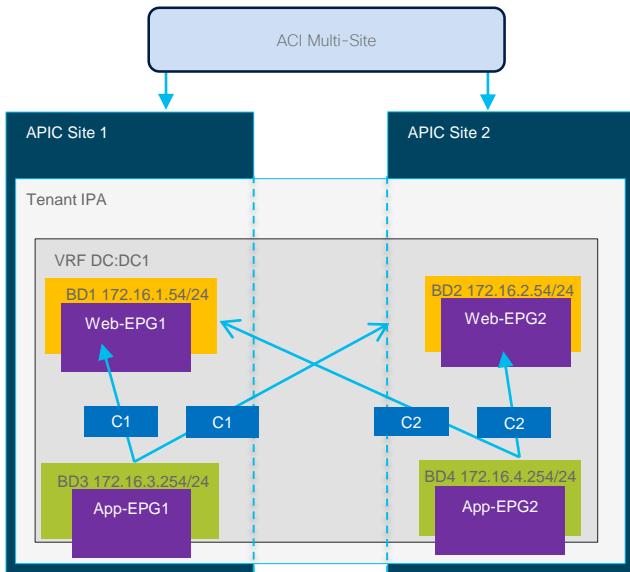
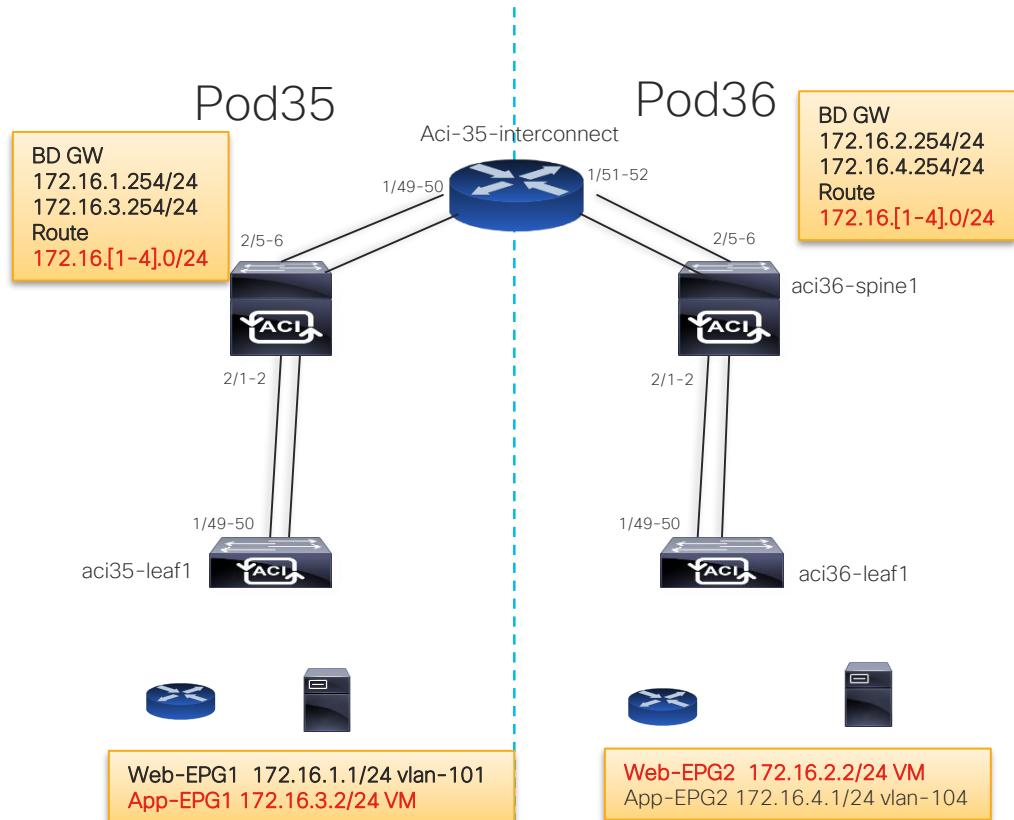
Overview

- Unicast TX proxy/(local to remote site)
 - Leaf has not learned the remote site ep. Leaf sends the traffic to local spine proxy. Local spine looks up the route. The route for remote site Ep is programmed with next hop of remote site's ETEP. Dipo is re-written with remote site ETEP. Sipo is re-written with local site ETEP
- Unicast TX (local to remote site)
 - Leaf has learned the remote site ep against remote site ETEP. Leaf sends the traffic to remote site ETEP. Local site spine will intercept this packet and re-write the sipo with Local site ETEP
- Unicast RX (remote to local site)
 - Incoming traffic destined to the local site's unicast ETEP goes through vnid and sclass translations. The receiving spine looks up the route for destination ep and sends the traffic to correct leaf.





Use case 1.1 – lab VRF DC:DC1



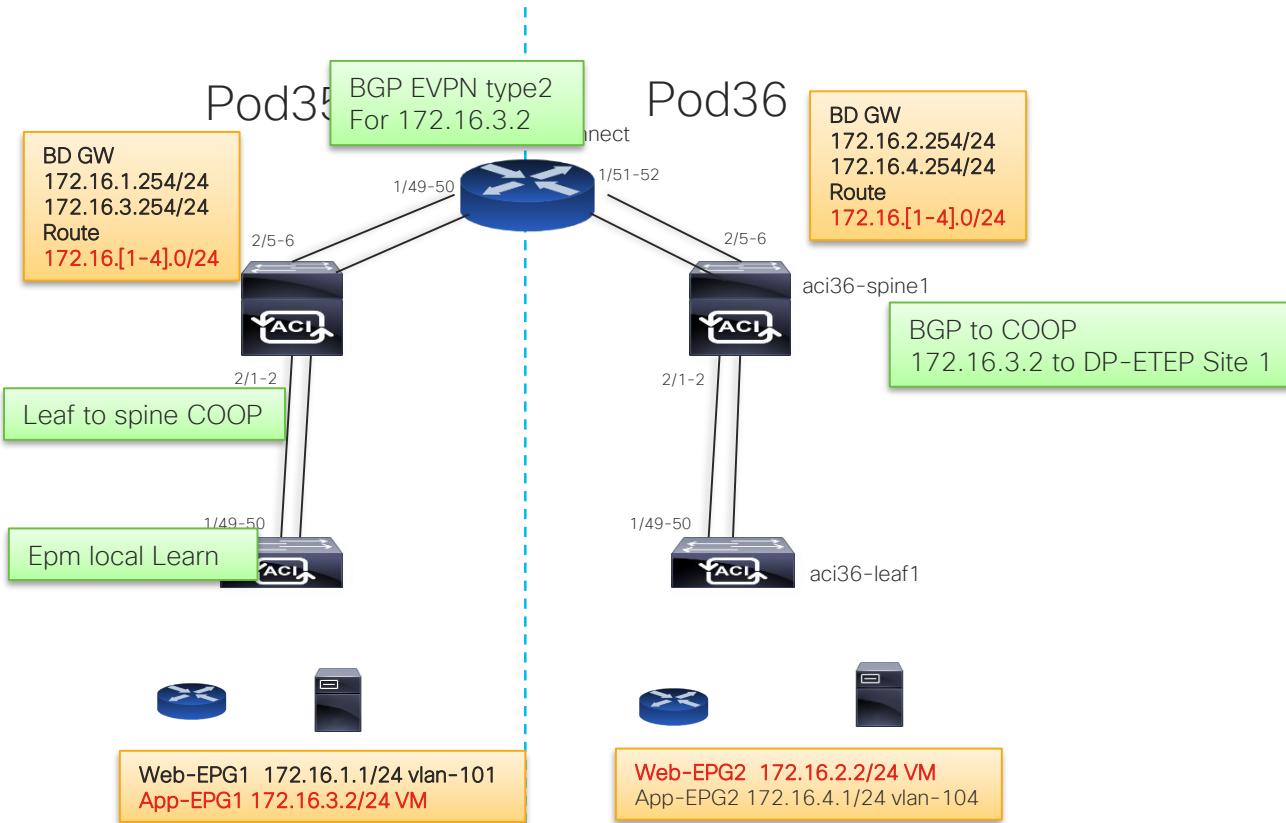
Test :
172.16.3.2 to 172.16.2.2

Src Site Leaf routing table

```
pod35-leaf1# show ip route vrf DC:DC1
IP Route Table for VRF "DC:DC1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

172.16.1.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.88.66%overlay-1, [1/0], 3d00h, static, tag 4294967295
172.16.1.254/32, ubest/mbest: 1/0, attached, pervasive
  *via 172.16.1.254, vlan13, [1/0], 3d00h, local, local
172.16.2.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.88.66%overlay-1, [1/0], 00:33:15, static, tag 4294967295
172.16.3.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.88.66%overlay-1, [1/0], 01w08d, static, tag 4294967295
172.16.3.254/32, ubest/mbest: 1/0, attached, pervasive
  *via 172.16.3.254, vlan21, [1/0], 00:33:15, local, local
172.16.4.0/24, ubest/mbest: 1/0, attached, direct, pervasive
  *via 10.0.88.66%overlay-1, [1/0], 01w08d, static, tag 4294967295
```

Control Plane



Control plane EP in Pod1

Local EPM

```
pod35-leaf1# show system internal epm endpoint ip 172.16.3.2

MAC : 0050.56b1.4b52 :: Num IPs : 1
IP# 0 : 172.16.3.2 :: IP# 0 flags :
Vlan id : 22 :: Vlan vnid : 8259 :: VRF name : DC:DC1
BD vnid : 16351138 :: VRF vnid : 3014656
Phy If : 0x1a001000 :: Tunnel If : 0
Interface : Ethernet1/2
Flags : 0x80004c04 :: sclass : 32770 :: Ref count : 5
EP Create Timestamp : 04/19/2018 07:02:49.606635
EP Update Timestamp : 04/24/2018 05:08:19.642826
EP Flags : local|IP|MAC|sclass|timer|
:::::
```

Extract BGP table site 1

```
pod35-spine1# show bgp 12vpn evpn vrf overlay-1 | egrep "Route Dis|172.16.3.2\["
Route Distinguisher: 1:133128354 (L2VNI 16351138)
*>1[2]:[0]:[16351138]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272
Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
*>1[2]:[0]:[16351138]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272
```

Extract BGP table site 2

```
pod36-spine1# show bgp 12vpn evpn vrf overlay-1 | egrep "Route Dis|172.16.3.2\["
Route Distinguisher: 1:133128354
*>e[2]:[0]:[0]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272
Route Distinguisher: 1:49676223 (L2VNI 16121791)
*>e[2]:[0]:[16121791]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272
Route Distinguisher: 10.10.35.102:136 (L2VNI 1)
*>e[2]:[0]:[16121791]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272
```

Local COOP site 1

```
pod35-spine1# show coop internal info ip-db key 3014656 172.16.3.2

IP address : 172.16.3.2
Vrf : 3014656
Flags : 0
EP bd vnid : 16351138
EP mac : 00:50:56:B1:4B:52
Publisher Id : 10.0.112.64
Record timestamp : 04 24 2018 04:29:16 958852767
Publish timestamp : 04 24 2018 04:29:16 959066592
Seq No: 0
Remote publish timestamp: 01 01 1970 00:00:00 0
URIB Tunnel Info
Num tunnels : 1
Tunnel address : 10.0.112.64
Tunnel ref count : 1
```

Remote COOP entry site 2

```
pod36-spine1# show coop internal info ip-db | egrep -A
15 -B 1 "172.16.3.2$"
-----
IP address : 172.16.3.2
Vrf : 2457600
Flags : 0x4
EP bd vnid : 16121791
EP mac : 00:50:56:B1:4B:52
Publisher Id : 10.10.35.101
Record timestamp : 01 01 1970 00:00:00 0
Publish timestamp : 01 01 1970 00:00:00 0
Seq No: 0
Remote publish timestamp: 04 24 2018 05:05:25 371412024
URIB Tunnel Info
Num tunnels : 1
Tunnel address : 10.10.35.101
Tunnel ref count : 1
```

Control plane BGP path from site 1 to site 2 (detail on site 1)

Entry with BGP EVI 16351138
Which is BD vnid of the EP

```
pod35-spine1# show bgp internal evi 16351138 | egrep "RT list"
Active Export RT list      : 135:33128354
Active Import RT list     : 136:49676223
```

```
pod35-spine1# show bgp 12vpn evpn 172.16.3.2 vrf overlay-1
Route Distinguisher: 1:33128354 (L2VNI 16351138)
BGP routing table entry for
[2]:[0]:[16351138]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272,
version 79 dest ptr 0xa95a74d4
MSITE RD: 1:33128354 (L2VNI 16351138)
Local Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
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, path is valid, is best
path
```

```
AS-Path: NONE, path locally originated
10.10.35.101 (metric 0) from 0.0.0.0 (10.10.35.111)
Origin IGP, MED not set, localpref 100, weight 32768
Received label 16351138 3014656
Extcommunity:
RT:5:16
```

```
Path-id 1 advertised to peers:
10.10.35.112
```

```
Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
BGP routing table entry for
[2]:[0]:[16351138]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272,
version 79 dest ptr 0xa95a74d4
MSITE RD: 1:33128354 (L2VNI 16351138)
Local Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
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, path is valid, is best path
AS-Path: NONE, path locally originated
10.10.35.101 (metric 0) from 0.0.0.0 (10.10.35.111)
Origin IGP, MED not set, localpref 100, weight 32768
Received label 16351138 3014656
Extcommunity:
RT:5:16

Path-id 1 advertised to peers:
10.10.35.112
```

Control plane BGP path from site 1 to site 2 (detail on site 2)

Rx path from site 1

```
pod36-spine1# show bgp 12vpn evpn 172.16.3.2 vrf overlay-1
Route Distinguisher: 1:33128354
BGP routing table entry for
[2]:[0]:[0]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272, version
828 dest ptr Oxa9603fa8
Paths: (1 available, best #1)
Flags: (0x000202 00000000) on xmit-list, is not in rib/evpn, is
locked
Multipath: eBGP iBGP

Advertised path-id 1
Path type: external 0x40000028 0x82040 ref 1, path is valid, is
best path, remote site path, remote nh not installed
AS-Path: 135 , path sourced external to AS
  10.10.35.101 (metric 3) from 10.10.35.111 (10.10.35.111)
    Origin IGP, MED not set, localpref 100, weight 0
  Received label 16351138 3014656
  Received path-id 1
  Extcommunity:
    RT:135:33128354
    SOO:135:33554415
    COST:pre-bestpath:166:2684354560
    ENCAP:8
    Router MAC:0200.0a0a.2365

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

```
pod36-spine1# show bgp internal evi 16121791 | egrep "RT list"
 Active Export RT list      : 136:49676223
 Active Import RT list     : 135:33128354
```

Route Distinguisher: 1:49676223 (L2VNI 16121791)
BGP routing table entry for
[2]:[0]:[16121791]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272, version 841 dest
ptr Oxa9604646
MSITE RD: 1:49676223 (L2VNI 16121791)
Local Route Distinguisher: 10.10.35.102:136 (L2VNI 1)
Paths: (1 available, best #1)
Flags: (0x00021a 0x00000a) on xmit-list, is in rib/evpn, is in 12rib msite
shard, is in 12rib
Multipath: eBGP iBGP

In rib/evpn and 12rib msite shard – means it is in coop and this spine is shard ownerFor coop

Advertised path-id 1
Path type: external 0xc0000028 0xa0040 ref 56506, path is valid, is best path,
remote site path, remote nh not installed
Imported from
1:33128354:[2]:[0]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/144
AS-Path: 135 , path sourced external to AS
 10.10.35.101 (metric 3) from 10.10.35.111 (10.10.35.111)
 Origin IGP, MED not set, localpref 100, weight 0
 Received label 16351138 3014656
 Received path-id 1
 Extcommunity:
 RT:135:33128354
 SOO:135:33554415
 COST:pre-bestpath:166:2684354560
 ENCAP:8
 Router MAC:0200.0a0a.2365
Route Distinguisher: 10.10.35.102:136 (L2VNI 1)
BGP routing table entry for
[2]:[0]:[16121791]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/272, version 841 dest
ptr Oxa9604646
MSITE RD: 1:49676223 (L2VNI 16121791)
Local Route Distinguisher: 10.10.35.102:136 (L2VNI 1)
Paths: (1 available, best #1)
Flags: (0x00021a 0x00000a) on xmit-list, is in rib/evpn, is in 12rib msite
shard, is in 12rib
Multipath: eBGP iBGP

Advertised path-id 1
Path type: external 0xc0000028 0xa0040 ref 56506, path is valid, is best path,
remote site path, remote nh not installed
Imported from
1:33128354:[2]:[0]:[48]:[0050.56b1.4b52]:[32]:[172.16.3.2]/144
AS-Path: 135 , path sourced external to AS
 10.10.35.101 (metric 3) from 10.10.35.111 (10.10.35.111)
 Origin IGP, MED not set, localpref 100, weight 0
 Received label 16351138 3014656
 Received path-id 1
 Extcommunity:
 RT:135:33128354

Imported path
To local RD

Coop site 2

- Checking COOP in local VRF VNID in site 2 we see it is indeed in coop DB
- 2457600 is the local vnid for the VRF

```
pod36-spine1# show coop internal info ip-db key  
2457600 172.16.3.2  
  
IP address : 172.16.3.2  
Vrf : 2457600  
Flags : 0x4  
EP bd vnid : 16121791  
EP mac : 00:50:56:B1:4B:52  
Publisher Id : 10.10.35.101  
Record timestamp : 01 01 1970 00:00:00 0  
Publish timestamp : 01 01 1970 00:00:00 0  
Seq No: 0  
Remote publish timestamp: 05 02 2018 05:11:56  
694567370  
URIB Tunnel Info  
Num tunnels : 1  
    Tunnel address : 10.10.35.101  
    Tunnel ref count : 1
```

DCI Mgr on spine pod 35 (site 1)

Remote Site
DP-ETEP and
Mcast ETEP(dcimgr and Object model)

```
pod35-spine1# show dcimgr repo eteps

Remote site=2 :
Rem Etep=10.10.35.102/32, is_ucast=yes
Rem Etep=10.10.35.122/32, is_ucast=no
```

```
pod35-spine1# moquery -c dci.AnycastExtn
Total Objects shown: 2

# dci.AnycastExtn
etep          : 10.10.35.102/32
childAction   :
dn            : sys/inst-overlay-1/remoteSite-2/anycastExtn-[10.10.35.102/32]
is_ucast      : yes
lcOwn        : local
modTs         : 2018-03-30T05:50:38.558+00:00
rn            : anycastExtn-[10.10.35.102/32]
status        :

# dci.AnycastExtn
etep          : 10.10.35.122/32
childAction   :
dn            : sys/inst-overlay-1/remoteSite-2/anycastExtn-[10.10.35.122/32]
is_ucast      : no
lcOwn        : local
modTs         : 2018-03-30T05:50:38.558+00:00
rn            : anycastExtn-[10.10.35.122/32]
status        :
```

DCI Mgr on spine pod 35 (site 1) – VNID MAP

Vrf vnid map in Obj Model

DCI mgr vnid map

```
pod35-spine1# show dcimgr repo vnid-maps
```

Remote site	Remote Vrf	Remote Bd	Local Vrf	Local Bd	Rel-state
2	2457600		3014656		[formed]
2	2457600	16121790	3014656	16056263	[formed]
2	2457600	16121791	3014656	16351138	[formed]
2	2457600	16220082	3014656	15925206	[formed]
2	2457600	15794151	3014656	16056262	[formed]

```
pod35-spine1# moquery -d sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]
Total Objects shown: 1
```

```
# 13.RemoteCtxSubstitute
FabEncap      : vxlan-2457600
DnName        : uni/tn-DC/ctx-DC1
childAction   :
dn           : sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]
lcOwn        : local
modTs         : 2018-04-11T04:06:17.167+00:00
rn           : remoteCtxSubstitute-[vxlan-2457600]
status        :
```

BD vnid map in Obj Model

```
pod35-spine1# moquery -d sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]/remoteBdSubstitute-[vxlan-16121790]
Total Objects shown: 1

# 12.RemoteBdSubstitute
FabEncap      : vxlan-16121790
DnName        : uni/tn-DC/BD-BD4
childAction   :
dn           : sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]/remoteBdSubstitute-[vxlan-16121790]
lcOwn        : local
modTs         : 2018-04-11T04:27:30.868+00:00
rn           : remoteBdSubstitute-[vxlan-16121790]
status        :
```

DCI Mgr on spine pod 35 (site 1) – SCLASS MAP

DCI mgr vnid map

```
pod35-spine1# show dcimgr repo sclass-maps
```

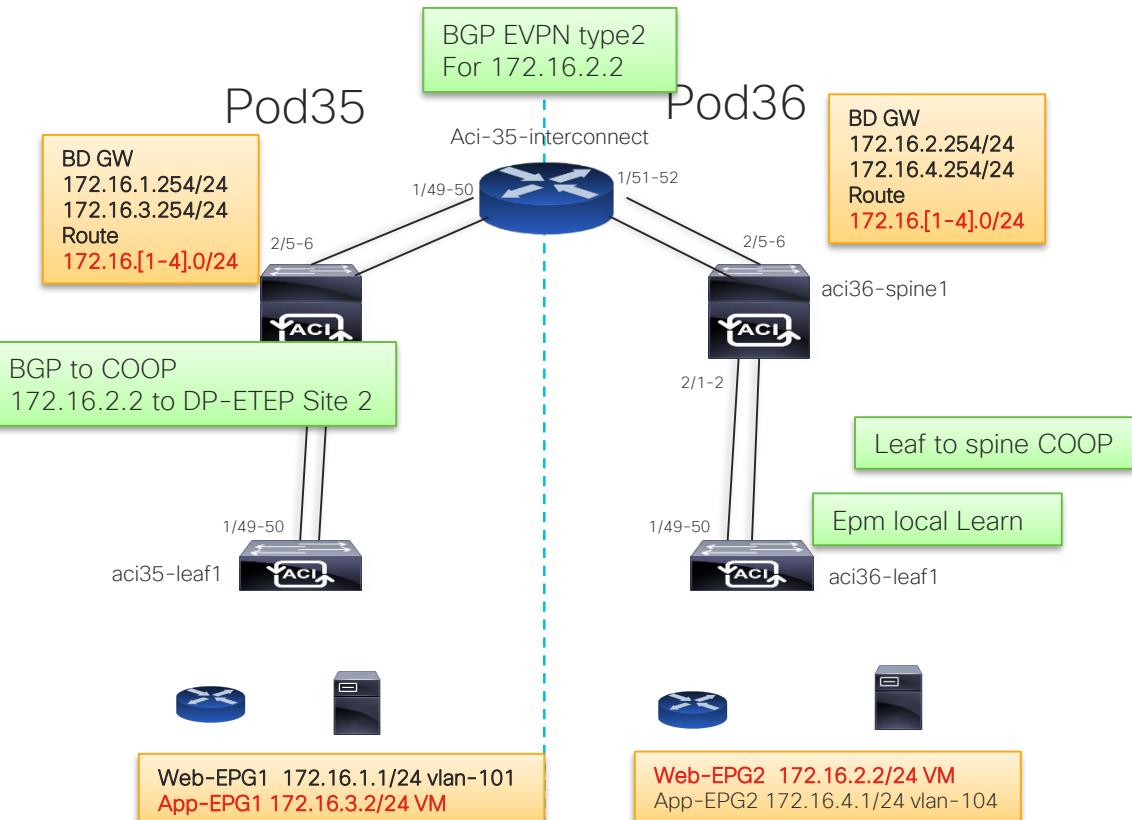
site	Remote		Local		
	Vrf	PcTag	Vrf	PcTag	Rel-state
2	2457600	49153		3014656	49153 [formed]
2	2457600	16387		3014656	16388 [formed]
2	2457600	32772		3014656	16387 [formed]
2	2457600	16390		3014656	32770 [formed]
2	2457600	32771		3014656	32772 [formed]

Sclass translate in object model

```
pod35-spine1# moquery -d sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]/remotePcTagSubstitute-16387
Total Objects shown: 1

# 12 .RemotePcTagSubstitute
pcTag      : 16387
DnName     : uni/tn-DC/ap-App2/epg-App-EPG2
childAction :
dn         : sys/inst-overlay-1/remoteSite-2/remoteCtxSubstitute-[vxlan-2457600]/remotePcTagSubstitute-16387
lcOwn      : local
modTs      : 2018-04-11T04:27:30.868+00:00
rn         : remotePcTagSubstitute-16387
status     :
```

Control Plane EP -in site 2 to COOP in site 1



Control plane EP in Site 2

Local EPM

```
pod36-leaf1# show system internal epm endpoint ip 172.16.2.2

MAC : 0050.56b1.4403 :: Num IPs : 1
IP# 0 : 172.16.2.2 :: IP# 0 flags :
Vlan id : 21 :: Vlan vnid : 8194 :: VRF name : DC:DC1
BD vnid : 16220082 :: VRF vnid : 2457600
Phy If : 0x1a001000 :: Tunnel If : 0
Interface : Ethernet1/2
Flags : 0x80004c04 :: sclass : 32771 :: Ref count : 5
EP Create Timestamp : 04/19/2018 07:03:23.999543
EP Update Timestamp : 05/02/2018 02:33:29.507208
EP Flags : local|IP|MAC|sclass|timer|
:::::
```

Extract BGP table site 2

```
pod36-spine1# show bgp l2vpn evpn vrf overlay-1 | egrep "Route Dis|172.16.2.2\["
Route Distinguisher: 1:49774514 (L2VNI 16220082)
*>1[2]:[0]:[16220082]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272
Route Distinguisher: 10.10.35.102:136 (L2VNI 1)
*>1[2]:[0]:[16220082]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272
```

Extract BGP table site 1

```
pod35-spine1# show bgp l2vpn evpn vrf overlay-1 | egrep "Route Dis|172.16.2.2\["
Route Distinguisher: 1:49774514
*>e[2]:[0]:[0]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272
Route Distinguisher: 1:32702422 (L2VNI 15925206)
*>e[2]:[0]:[15925206]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272
Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
*>e[2]:[0]:[15925206]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272
```

Local COOP site 2

```
pod36-spine1# show coop internal info ip-db key 2457600 172.16.2.2

IP address : 172.16.2.2
Vrf : 2457600
Flags : 0
EP bd vnid : 16220082
EP mac : 00:50:56:B1:44:03
Publisher Id : 10.1.48.64
Record timestamp : 05 02 2018 02:29:12 339899902
Publish timestamp : 05 02 2018 02:29:12 340145880
Seq No: 0
Remote publish timestamp: 01 01 1970 00:00:00 0
URIB Tunnel Info
Num tunnels : 1
    Tunnel address : 10.1.48.64
    Tunnel ref count : 1
```

Remote COOP entry site 1

```
pod35-spine1# show coop internal info ip-db | egrep -A
15 -B 1 "172.16.2.2$"
-----
IP address : 172.16.2.2
Vrf : 3014656
Flags : 0x4
EP bd vnid : 15925206
EP mac : 00:50:56:B1:44:03
Publisher Id : 10.10.35.102
Record timestamp : 01 01 1970 00:00:00 0
Publish timestamp : 01 01 1970 00:00:00 0
Seq No: 0
Remote publish timestamp: 04 24 2018 05:05:34 611613733
URIB Tunnel Info
Num tunnels : 1
    Tunnel address : 10.10.35.102
    Tunnel ref count : 1
```

Control plane BGP path from site 2 to site 1 (detail on site 2)

```
pod36-spine1# show bgp l2vpn evpn 172.16.2.2 vrf overlay-1
Route Distinguisher: 1:49774514      (L2VNI 16220082)
BGP routing table entry for
[2]:[0]:[16220082]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272,
version 69 dest ptr 0xa9603608
MSITE RD: 1:49774514      (L2VNI 16220082)
Local Route Distinguisher: 10.10.35.102:136      (L2VNI 1)
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, path is valid, is best
path
AS-Path: NONE, path locally originated
10.10.35.102 (metric 0) from 0.0.0.0 (10.10.35.112)
Origin IGP, MED not set, localpref 100, weight 32768
Received label 16220082 2457600
Extcommunity:
    RT:5:16

Path-id 1 advertised to peers:
10.10.35.111
```

```
Route Distinguisher: 10.10.35.102:136      (L2VNI 1)
BGP routing table entry for
[2]:[0]:[16220082]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272,
version 69 dest ptr 0xa9603608
MSITE RD: 1:49774514      (L2VNI 16220082)
Local Route Distinguisher: 10.10.35.102:136      (L2VNI 1)
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, path is valid, is best path
AS-Path: NONE, path locally originated
10.10.35.102 (metric 0) from 0.0.0.0 (10.10.35.112)
Origin IGP, MED not set, localpref 100, weight 32768
Received label 16220082 2457600
Extcommunity:
    RT:5:16

Path-id 1 advertised to peers:
10.10.35.111
```

Control plane BGP path from site 2 to site 1 (detail on site 1)

Rx path from site 2

```
pod35-spine1# show bgp l2vpn evpn 172.16.2.2 vrf overlay-1
Route Distinguisher: 1:49774514
BGP routing table entry for
[2]:[0]:[0]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272, version
529 dest ptr Oxa95a7a3e
Paths: (1 available, best #1)
Flags: (0x000202 00000000) on xmit-list, is not in rib/evpn, is
locked
Multipath: eBGP iBGP
```

```
Advertised path-id 1
Path type: external 0x40000028 0x82040 ref 1, path is valid, is
best path, remote site path, remote nh not installed
AS-Path: 136 , path sourced external to AS
  10.10.35.102 (metric 3) from 10.10.35.112 (10.10.35.112)
    Origin IGP, MED not set, localpref 100, weight 0
  Received label 16220082 2457600
  Received path-id 1
  Extcommunity:
    RT:136:49774514
    SOO:136:50331631
    COST:pre-bestpath:166:2684354560
    ENCAP:8
    Router MAC:0200.0a0a.2366
Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
```

Path-id 1 not advertised to any peer

```
Route Distinguisher: 1:32702422 (L2VNI 15925206)
BGP routing table entry for
[2]:[0]:[15925206]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272, version 542 dest
ptr Oxa95a8646
MSITE RD: 1:32702422 (L2VNI 15925206)
Local Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
Paths: (1 available, best #1)
Flags: (0x00021a 0x00000a) on xmit-list, is in rib/evpn, is in l2rib msite
shard, is in l2rib
Multipath: eBGP iBGP
```

```
Advertised path-id 1
Path type: external 0xc0000028 0xa0040 ref 56506, path is valid, is best path,
remote site path, remote nh not installed
Imported from
```

```
1:49774514:[2]:[0]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/144
```

```
AS-Path: 136 , path sourced external to AS
```

```
  10.10.35.102 (metric 3) from 10.10.35.112 (10.10.35.112)
```

```
    Origin IGP, MED not set, localpref 100, weight 0
```

```
  Received label 16220082 2457600
```

```
  Received path-id 1
```

```
  Extcommunity:
```

```
    RT:136:49774514
```

```
    SOO:136:50331631
```

```
    COST:pre-bestpath:166:2684354560
```

```
    ENCAP:8
```

```
    Router MAC:0200.0a0a.2366
```

```
Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
```

```
BGP routing table entry for
```

```
[2]:[0]:[15925206]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/272, version 542 dest
```

```
ptr Oxa95a8646
```

```
MSITE RD: 1:32702422 (L2VNI 15925206)
```

```
Local Route Distinguisher: 10.10.35.101:135 (L2VNI 1)
```

```
Paths: (1 available, best #1)
```

```
Flags: (0x00021a 0x00000a) on xmit-list, is in rib/evpn, is in l2rib msite
```

```
shard, is in l2rib
```

```
Multipath: eBGP iBGP
```

```
Advertised path-id 1
```

```
Path type: external 0xc0000028 0xa0040 ref 56506, path is valid, is best path,
remote site path, remote nh not installed
Imported from
```

```
1:49774514:[2]:[0]:[48]:[0050.56b1.4403]:[32]:[172.16.2.2]/144
```

```
AS-Path: 136 , path sourced external to AS
```

```
  10.10.35.102 (metric 3) from 10.10.35.112 (10.10.35.112)
```

```
    Origin IGP, MED not set, localpref 100, weight 0
```

```
  Received label 16220082 2457600
```

```
  Received path-id 1
```

```
  Extcommunity:
```

```
    RT:136:49774514
```

DCI Mgr on spine pod 36 (site 2)

Remote Site
DP-ETEP and
Mcast ETEP(dcimgr and Object model)

```
pod36-spine1# show dcimgr repo eteps

Remote site=1 :
Rem Etep=10.10.35.101/32, is_ucast=yes
Rem Etep=10.10.35.121/32, is_ucast=no
pod36-spine1#
```

```
pod36-spine1# moquery -c dciAnycastExtn
Total Objects shown: 2

# dci.AnycastExtn
etep          : 10.10.35.101/32
childAction   :
dn            : sys/inst-overlay-1/remoteSite-1/anycastExtn-
[10.10.35.101/32]
is_ucast      : yes
lcOwn        : local
modTs         : 2018-03-30T05:50:34.562+00:00
rn            : anycastExtn-[10.10.35.101/32]
status        :

# dci.AnycastExtn
etep          : 10.10.35.121/32
childAction   :
dn            : sys/inst-overlay-1/remoteSite-1/anycastExtn-
[10.10.35.121/32]
is_ucast      : no
lcOwn        : local
modTs         : 2018-03-30T05:50:34.562+00:00
rn            : anycastExtn-[10.10.35.121/32]
status        :
```

DCI Mgr on spine pod 36 (site 2) – VNID MAP

DCI mgr vnid map

```
pod35-spine1# show dcimgr repo vnid-maps
```

Remote site	Vrf	Bd	Local Vrf	Bd	Rel-state
1	3014656		2457600		[formed]
1	3014656	16056263	2457600	16121790	[formed]
1	3014656	16351138	2457600	16121791	[formed]
1	3014656	15925206	2457600	16220082	[formed]
1	3014656	16056262	2457600	15794151	[formed]

Vrf vnid map in Obj Model

```
pod36-spine1# moquery -c 13RemoteCtxSubstitute  
  
# 13.RemoteCtxSubstitute  
FabEncap : vxlan-3014656  
DnName : uni/tn-DC/ctx-DC1  
childAction :  
dn : sys/inst-overlay-1/remoteSite-1/remoteCtxSubstitute-[vxlan-3014656]  
lcOwn : local  
modTs : 2018-04-11T04:06:11.695+00:00  
rn : remoteCtxSubstitute-[vxlan-3014656]  
status :
```

BD vnid map in Obj Model

```
pod36-spine1# moquery -d sys/inst-overlay-1/remoteSite-1/remoteCtxSubstitute-[vxlan-3014656]/remoteBdSubstitute-[vxlan-15925206]  
Total Objects shown: 1  
  
# 12.RemoteBdSubstitute  
FabEncap : vxlan-15925206  
DnName : uni/tn-DC/BD-BD2  
childAction :  
dn : sys/inst-overlay-1/remoteSite-1/remoteCtxSubstitute-[vxlan-3014656]/remoteBdSubstitute-[vxlan-15925206]  
lcOwn : local  
modTs : 2018-04-11T04:28:16.146+00:00  
rn : remoteBdSubstitute-[vxlan-15925206]  
status :
```

DCI Mgr on spine pod 36 (site 2) – SCLASS MAP

DCI mgr vnid map

```
pod36-spine1# show dcimgr repo sclass-maps
```

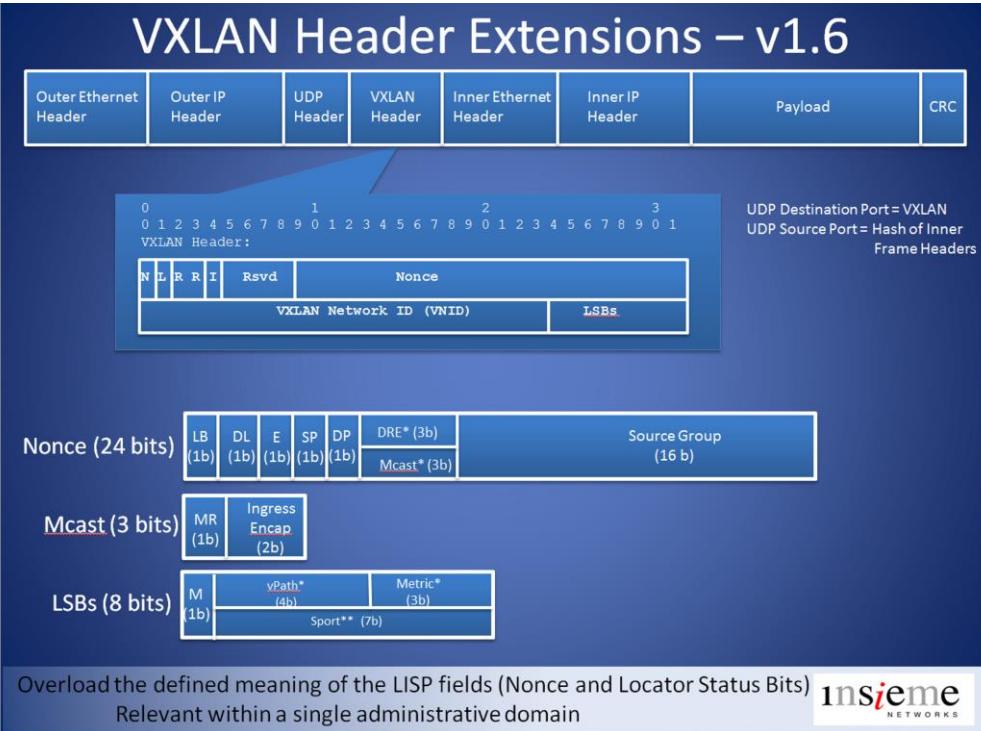
site	Remote		Local		
	Vrf	PcTag	Vrf	PcTag	Rel-state
1	3014656	49153	1	2457600	[formed]
1	3014656	16387	1	2457600	[formed]
1	3014656	16388	1	2457600	[formed]
1	3014656	32770	1	2457600	[formed]
1	3014656	32772	1	2457600	[formed]

Sclass translate in object model

```
pod36-spine1# moquery -c 12RemotePcTagSubstitute
Total Objects shown: 12

# 12.RemotePcTagSubstitute
pcTag      : 16388
DnName     : uni/tn-DC/ap-App2/epg-App-EPG2
childAction :
dn         : sys/inst-overlay-1/remoteSite-1/remoteCtxSubstitute-[vxlan-3014656]/remotePcTagSubstitute-16388
lcOwn      : local
modTs     : 2018-04-11T04:27:26.433+00:00
rn         : remotePcTagSubstitute-16388
status     :
```

vxlan header review



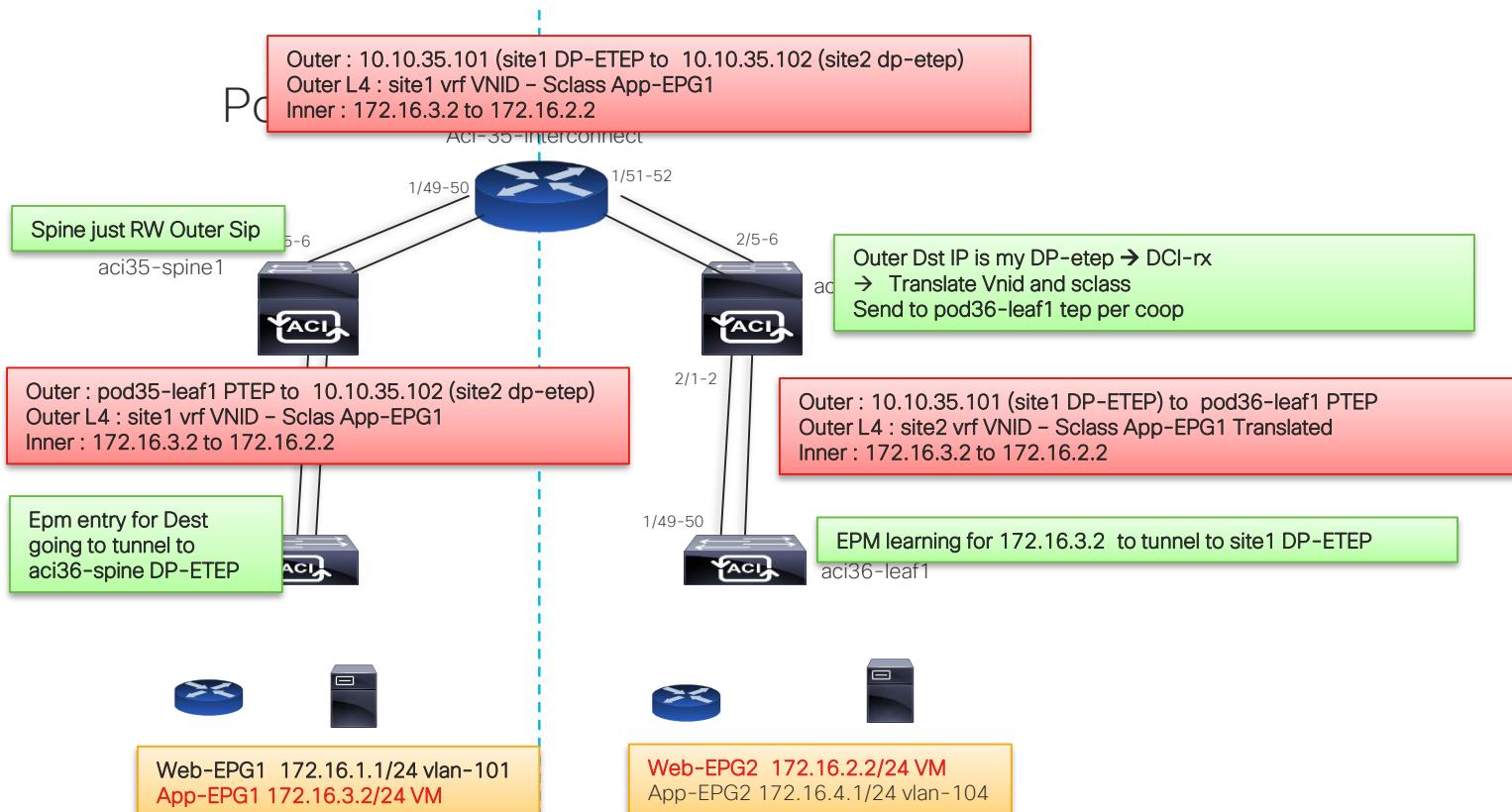
Note in Outer L4 header you can Get :
VNID (BD or VRF)
Sclass (src sclass) as part of Nounce field (last 4 nibble):

Ex :

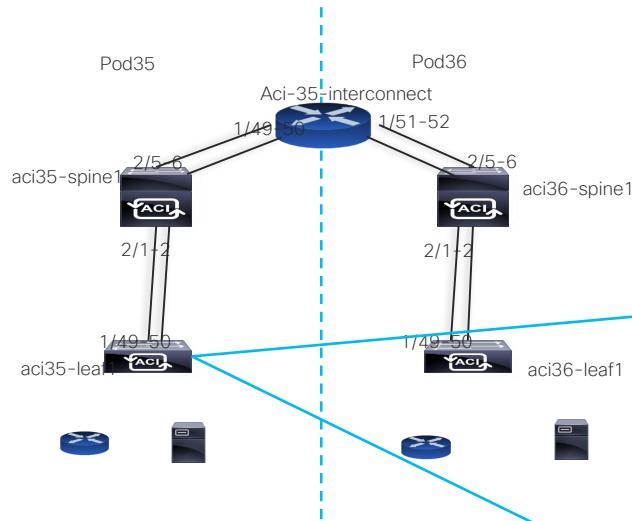
`hom_elam_in_14v_tn.tn_nonce_info: 0x188002`

`Sclass of Rx frame is 0x8002`

Data path - known EP Site 1 to Site 2 (Known unicast on ingress leaf)



Ingress Leaf Known EP

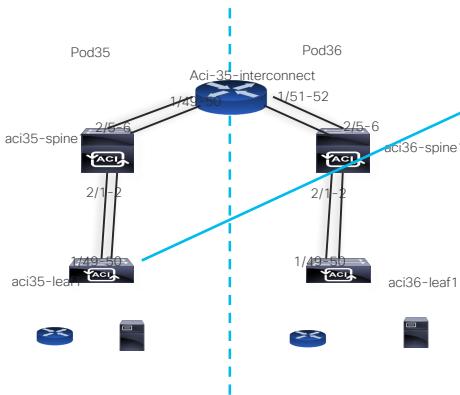


```
pod35-leaf1# show system internal epm endpoint ip 172.16.2.2

MAC : 0000.0000.0000 :: Num IPs : 1
IP# 0 : 172.16.2.2 :: IP# 0 flags :
Vlan id : 0 :: Vlan vnid : 0 :: VRF name : DC:DC1
BD vnid : 0 :: VRF vnid : 3014656
Phy If : 0 :: Tunnel If : 0x18010007
Interface : Tunnel17
Flags : 0x80004400 :: sclass : 32772 :: Ref count : 3
EP Create Timestamp : 04/24/2018 05:05:32.831665
EP Update Timestamp : 04/25/2018 04:58:50.374323
EP Flags : IP|sclass|timer|
:::::
```

```
pod35-leaf1# show interface tunnel 7
Tunnel17 is up
MTU 9000 bytes, BW 0 Kbit
Transport protocol is in VRF "overlay-1"
Tunnel protocol/transport is ivxlan
Tunnel source 10.0.112.64/32 (lo0)
Tunnel destination 10.10.35.102/32
Last clearing of "show interface" counters never
Tx
0 packets output, 1 minute output rate 0 packets/sec
Rx
0 packets input, 1 minute input rate 0 packets/sec
```

ELAM Ingress leaf Site 1 - EP known



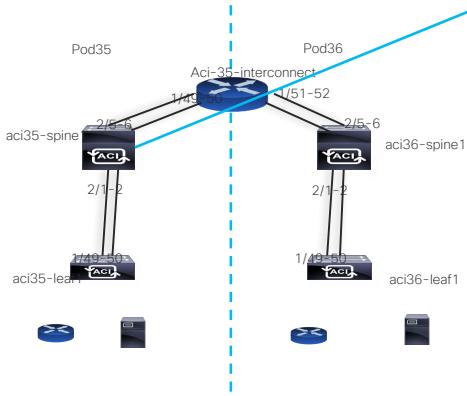
```
module-1# debug platform internal roc elam asic 0
module-1# trigger init in-select 6 out-select 1
module-1# set outer ipv4 src ip 172.16.3.2 dst ip 172.16.2.2
```

```
module-1(DBG-elam-insel16) # report | egrep "SRCID|ovector|encap"
    SRCID: 22
    hom_lurw_vec.encap_12_idx: 0x2
    hom_lurw_vec.encap_pcid: 0x0
    hom_lurw_vec.encap_ids: 0x1
    hom_lurw_vec.encap_vld: 0x1
    hom_elam_out_sidebnd_no_spare_vec.ovector_idx: 0x78
```

```
module-1(DBG-elam-insel6)#
show platform internal hal tunnel rtep apd
=====
ifId      IP          HwVrfId   BDXlate  SrcTepIdx DstInfoIdx  RwEncapIdx ECMPIdx  ECMPMbrIdx Num    L2Index  RwDmacIdx
=====
18010007  10.10.35.102 4        1        3aa3       2800        1           0           0           2           1a030000 2
                                         1a031000 3
```

module-1(DBG-elam-insel6) # show platform internal hal 12 port gpd																																	
IfId	Ifname	Uc		Uc		Reprogram														Rep													
		I	PC	P	Cfg	MbrID	As	AP	S1	Sp	Ss	Ovec	S	L	R	I	R	D	R	U	U	X	L	Xla	Ovx	N	NI	Vif	RwV	Ing	Egr	V	R
1a001000	Eth1/2	0	77	58	0	12	0	11	22	22	1	0	0	0	0	0	0	0	0	0	0	0	D-116	-	0	0	0	0	24	0			
1a030000	Eth1/49	0	1	42	0	3d	0	3c	78	78	1	0	0	0	0	0	0	0	0	1	6	4	2	2	D-16a	-	400	0	0	0	19	0	

ELAM Ingress LC Spine Site 1 - EP known



```
module-2# debug platform internal roc elam asic 0
module-2(DBG-elam)# trigger init in-select 14 out-select 1
module-2(DBG-elam-insel15)# set inner ipv4 src_ip 172.16.3.2 dst_ip 172.16.2.2
```

```
#####
##### HOMWOOD ELAM REPORT START #####
Dumping report for asic type 8 inst 0 slice 0 a_to_d 1 insel 15 outsel 1
LUA captured data with :
SRCID: 20
*** Parsed Outer 12 vector
hom_elam_in_12v_da_sa_qtag0_qtag0_vlan: 0x2
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv4_da: 0xA0A2366      - 10.0.35.102 (Dp-ETEP site2)
hom_elam_in_13v_ipv4_sa: 0xA007040      - 10.0.112.64 (leaf1 pod35 PTEP)
*** Parsed Outer 14 vector
hom_elam_in_14v_tn_tn_seg_id: 0x2E0000          - 3014656
hom_elam_in_14v_tn_tn_nonce_info: 0x8002

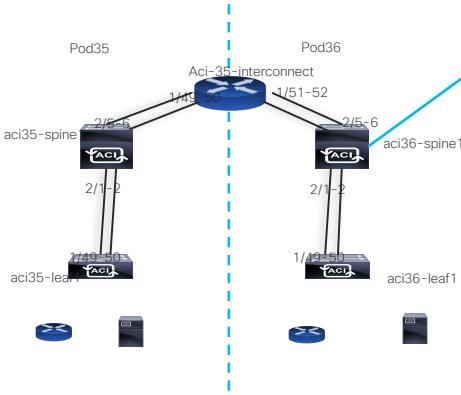
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_rx: 0x0
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_tx: 0x0
```

```
module-2(DBG-elam-insel15)# show platform internal hal 12 port gpd | egrep "Eth2/1==|IfId|Uc|Xla"
```

IfId	Ifname	Uc		Uc		Reprogram						RwV	Ing	Egr	Rep		PROF	H																	
		I	PC	P	Cfg	MbrID	As	AP	S1	Sp	Ss				Ovec	S				R	I	D	R	U	X		L	Xla	Ovx	N	NI	Vif			
		P	Cfg								S		P	P	P	S	P	S	Sp	C	M	L		3	Idx	Idx	L3	L3	Tid	Tid					
1a080000	Eth2/1	0	9a	28	0	11	0	10	20	20	1		0	0	0	0	0	0	0	0	0	0		1	1	1	1	D-f3	D-61	100	0	0	0	4	0

```
pod35-spine1# show lldp neighbors | egrep "Eth2/1"
pod35-leaf1      Eth2/1      120      BR      Eth1/49
pod35-spine1#
```

ELAM Ingress LC Spine Site 2 - Proxy



```
module-2# debug platform internal roc elam asic 0
module-2(DBG-elam)# trigger reset
module-2(DBG-elam)# trigger init in-select 15 out-select 1
module-2(DBG-elam-insel15)# set inner ipv4 src_ip 172.16.3.2 dst_ip 172.16.2.2
```

```
#####
##### HOMWOOD ELAM REPORT START #####
Dumping report for asic type 8 inst 0 slice 0 a_to_d 1 insel 15 outsel 1
LUA captured data with :
SRCID: 0
*** Parsed Outer 12 vector
hom_elam_in_12v_da_sa_qtag0.qtag0_vlan: 0x4
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv6_da_only.da: 0x0000000000000000A0A2366 - 10.10.35.102 (site2 - DP-EPEP)
hom_elam_in_13v_ipv6_da_only.sa: 0xA0A2365 - 10.10.35.101 (site1 - DP-EPEP)
*** Parsed Outer 14 vector
hom_elam_in_14v_tn.tn_nonce_info: 0x188002 - Rx sclass is 0x8002 = 16387
hom_elam_in_14v_tn.tn_seg_id: 0x2E0000 - 3014656 (vnid before rewrite)

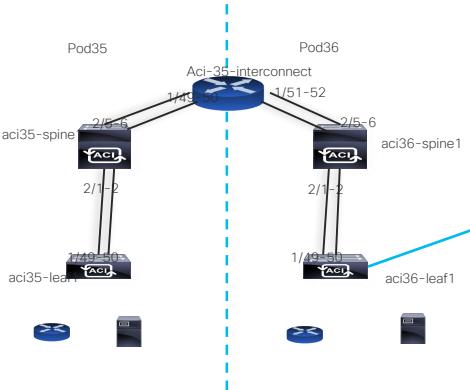
hom_elam_out_sidebnd_no_spare_vec.ovector_idx: 0x78 (useless internal port to FC)
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_rx: 0x1
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_tx: 0x0

hom_lurw_vec.info.ifabric_spine.vnid: 0x258000 - Vnid after rewrite = 2457600
hom_lurw_vec.info.ifabric_spine.sclass: 0x4006 - rewritten Sclass is 16390

===== lux_fwd_mode = 0x09516040
LUX_FWD_MODE: ISPIINE_LC      bit is set      ingress LC
LUX_FWD_MODE: ISPINE_DCI     bit is set
..
```

```
pod36-spine1# show dcimgr repo sclass-maps | egrep "3014656.*16387"
1 3014656 16387 | 2457600 32772 [formed]
pod36-spine1# show dcimgr repo vnid-maps | egrep 3014656
1 3014656 | 2457600 [formed]
```

ELAM egress leaf Site 2 – EP known



```
module-1# debug platform internal roc elam asic 0
module-1(DBG-elam)# trigger reset
module-1(DBG-elam)# trigger init in-select 15 out-select 1
module-1(DBG-elam-insel15)# set inner ipv4 src_ip 172.16.3.2 dst_ip 172.16.2.2
```

```
#####
##### HOMWOOD ELAM REPORT START #####
Dumping report for asic type 8 inst 0 slice 0 a_to_d 1 insel 15 outsel 1
LUA captured data with :
SRCID: 80
*** Parsed Outer 12 vector
hom_elam_in_12v_da_sa_qtag0.qtag0_vlan: 0x2
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv6_da_only.da: 0x0000000000000000A013040 - 10.1.48.64 (pod36-leaf1 PTEP)
hom_elam_in_13v_ipv6_da_only.sa: 0xA0A2365 - 10.10.35.101 (site 1 - DP-EPEP)
*** Parsed Outer 14 vector
hom_elam_in_14v_tn.tn_nonce_info: 0x384006 - sclass in packet is 0x4006
hom_elam_in_14v_tn.tn_seg_id: 0x258000 - vnid is the rewritten vnid

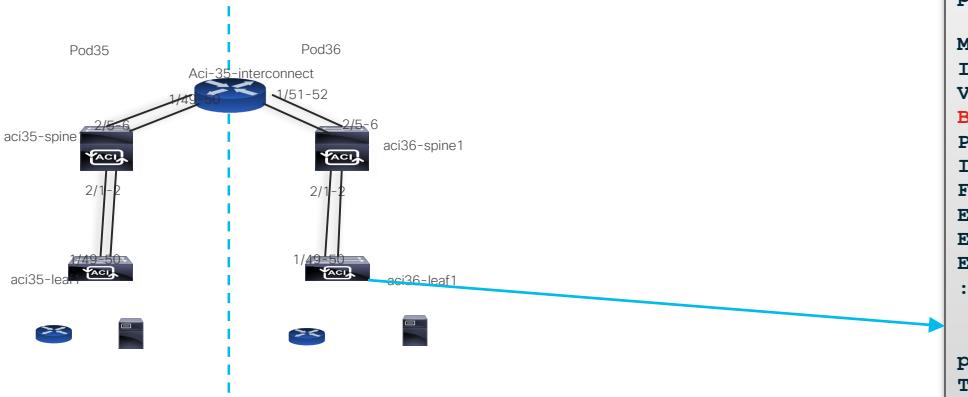
hom_lurw_vec.info.ifabric_leaf.dclass: 0x8003
hom_lurw_vec.info.ifabric_leaf.sclass: 0x4006
```

```
module-2(DBG-elam-insel15)# show platform internal hal 12 port gpd | egrep "Eth2/1==|IfId|Uc|Xla"
```

IfId	Ifname	Uc		Uc		Reprogram						Rep																							
		I	PC	P	Cfg	MbrID	As	AP	S1	Sp	Ss	Ovec	S	P	P	P	S	P	S	P	Sp	C	M	L	3	Idx	Idx	L3	L3	Tid	RwV	Ing	Egr	V	R
1a084000	Eth2/5	0	9e	24	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	b	b	1	1	D-19a	D-2ee	300	0	1	0	2	0	

```
pod35-spine1# show lldp neighbors | egrep "Eth2/1"
pod35-leaf1      Eth2/1      120      BR      Eth1/49
pod35-spine1#
```

EPM dest leaf in site 2

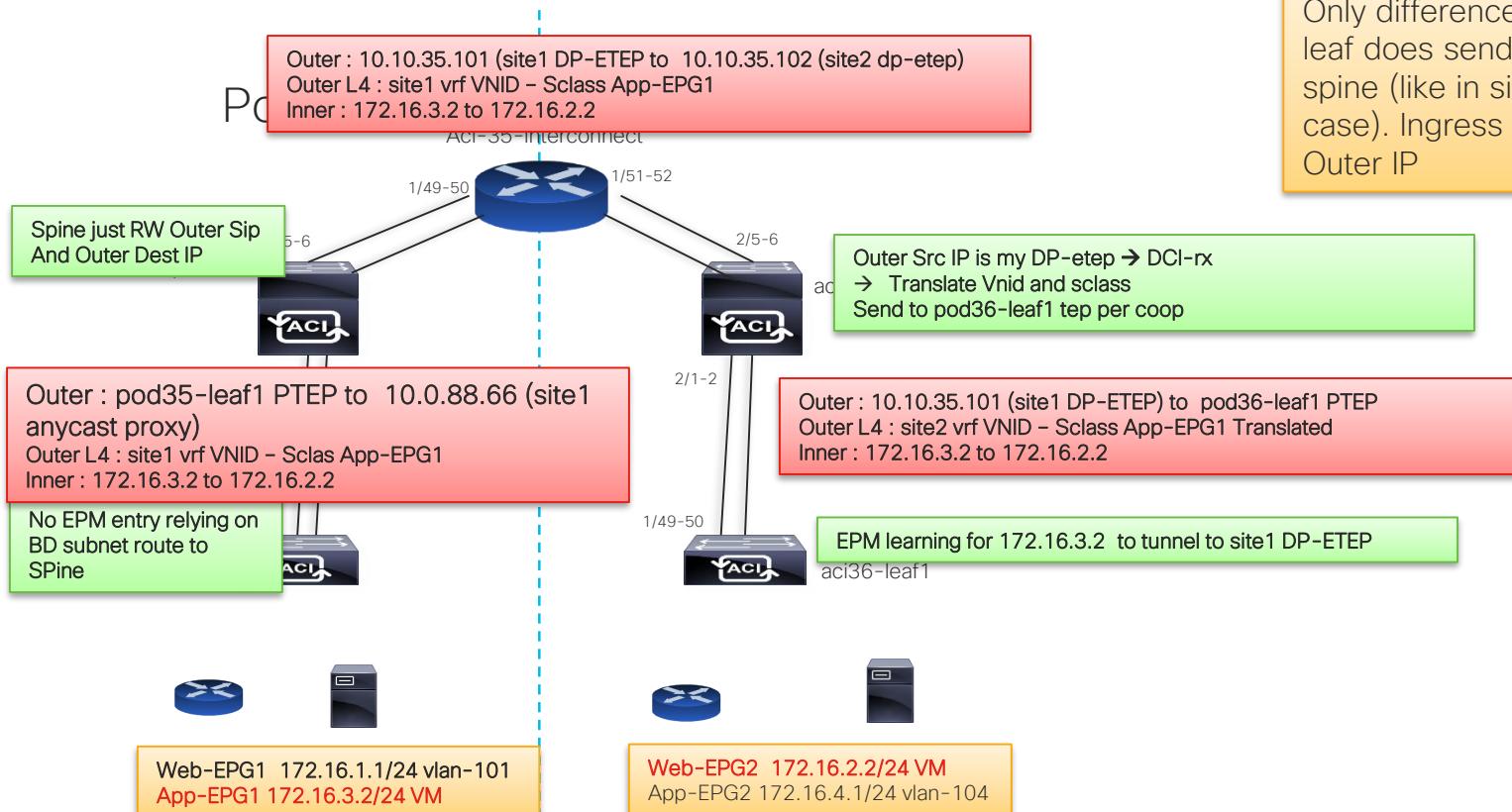


```
pod36-leaf1# show system internal epm endpoint ip 172.16.3.2

MAC : 0000.0000.0000 :: Num IPs : 1
IP# 0 : 172.16.3.2 :: IP# 0 flags :
Vlan id : 0 :: Vlan vniid : 0 :: VRF name : DC:DC1
BD vniid : 0 :: VRF vniid : 2457600
Phy If : 0 :: Tunnel If : 0x18010007
Interface : Tunnel7
Flags : 0x80004400 :: sclass : 16390 :: Ref count : 3
EP Create Timestamp : 04/24/2018 05:05:25.720716
EP Update Timestamp : 04/25/2018 05:22:30.240899
EP Flags : IP|sclass|timer|
::::

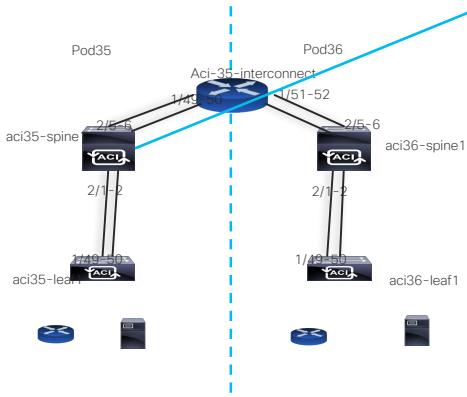
pod36-leaf1# show interface tunnel 7
Tunnel7 is up
MTU 9000 bytes, BW 0 Kbit
Transport protocol is in VRF "overlay-1"
Tunnel protocol/transport is ivxlan
Tunnel source 10.1.48.64/32 (lo0)
Tunnel destination 10.10.35.101/32
Last clearing of "show interface" counters never
Tx
0 packets output, 1 minute output rate 0 packets/sec
Rx
0 packets input, 1 minute input rate 0 packets/sec
```

Data path - unknown EP on leaf Site 1 to Site 2 - Proxy



Only differences is that ingress leaf does send to Local Proxy spine (like in single site pod case). Ingress spine does Rw Outer IP

ELAM Ingress LC Spine Site 1 – Proxy



```
module-2# debug platform internal roc elam asic 0
module-2(DBG-elam)# trigger init in-select 14 out-select 1
module-2(DBG-elam-insel15)# set inner ipv4 src_ip 172.16.3.2 dst_ip 172.16.2.2
```

```
#####
##### HOMWOOD ELAM REPORT START #####
Dumping report for asic type 8 inst 0 slice 0 a_to_d 1 insel 15 outsel 1
LUA captured data with :
SRCID: 20
*** Parsed Outer 12 vector
hom_elam_in_12v_da_sa_qtag0_qtag0_vlan: 0x2
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv4_da: 0xA005842      - 10.0.88.66 (spine proxy site 1)
hom_elam_in_13v_ipv4_sa: 0xA007040      - 10.0.112.64  (leaf1 pod35 PTEP)
*** Parsed Outer 14 vector
hom_elam_in_14v_tn_tn_seg_id: 0x2E0000          - 3014656
hom_elam_in_14v_tn_tn_nonce_info: 0x8002

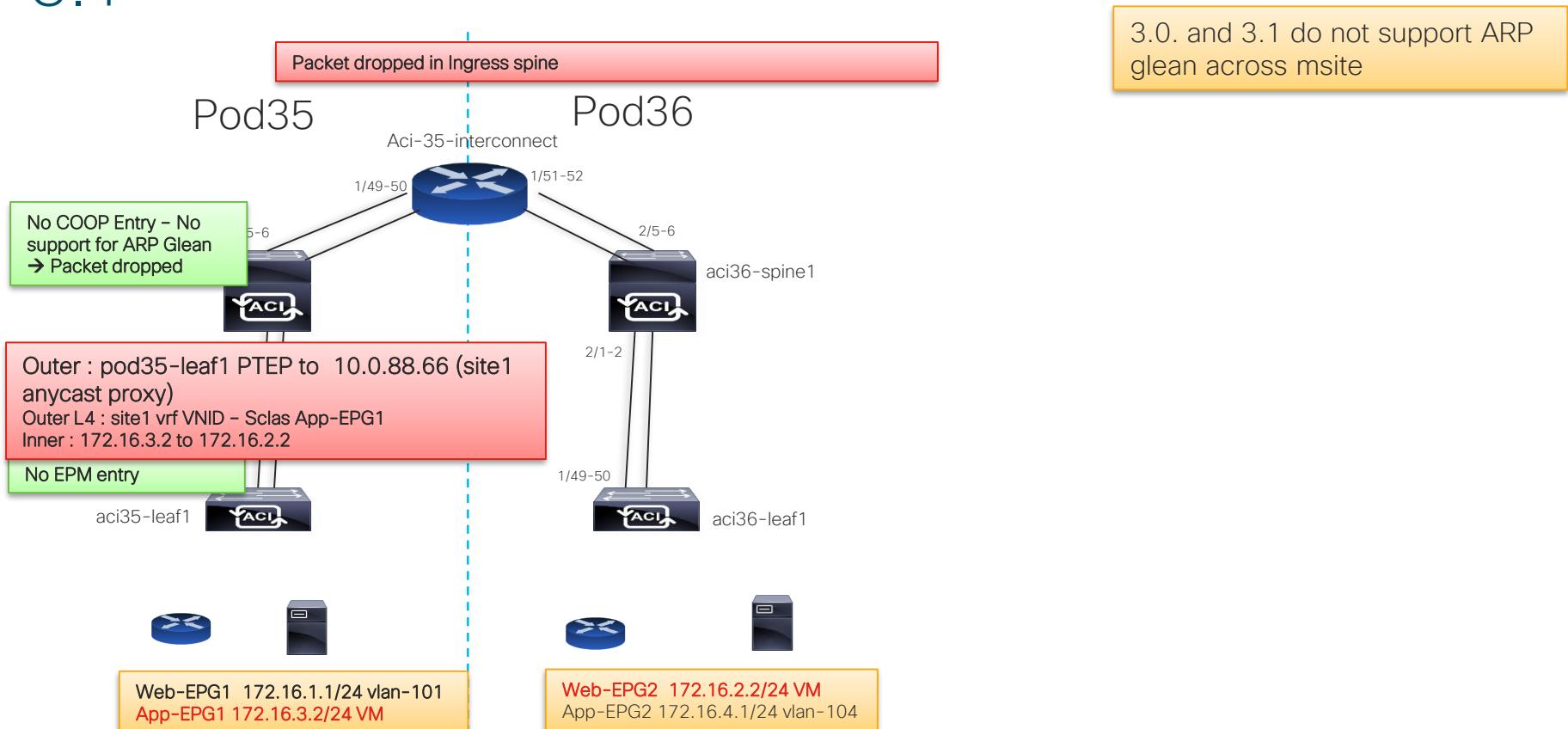
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_rx: 0x0
hom_lua_latch_results_vec.lua4_1.lux_ispine_dci_tx: 0x0
```

```
module-2(DBG-elam-insel15)# show platform internal hal 12 port gpd | egrep "Eth2/1==|IfId|Uc|Xla"
```

IfId	Ifname	Uc		Uc		Reprogram												RwV	Ing	Egr	Rep		PROF	H												
		I	PC	P	Cfg	MbrID	As	AP	S1	Sp	Ss	Ovec	S		R	I	D				R	U			U	X		L	Xla	Ovx	N	NI	Vif	Tid	Tid	Lbl
1a080000	Eth2/1	0	9a	28	0	11	0	10	20	20	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	D-f3	D-61	100	0	0	0	4	0			

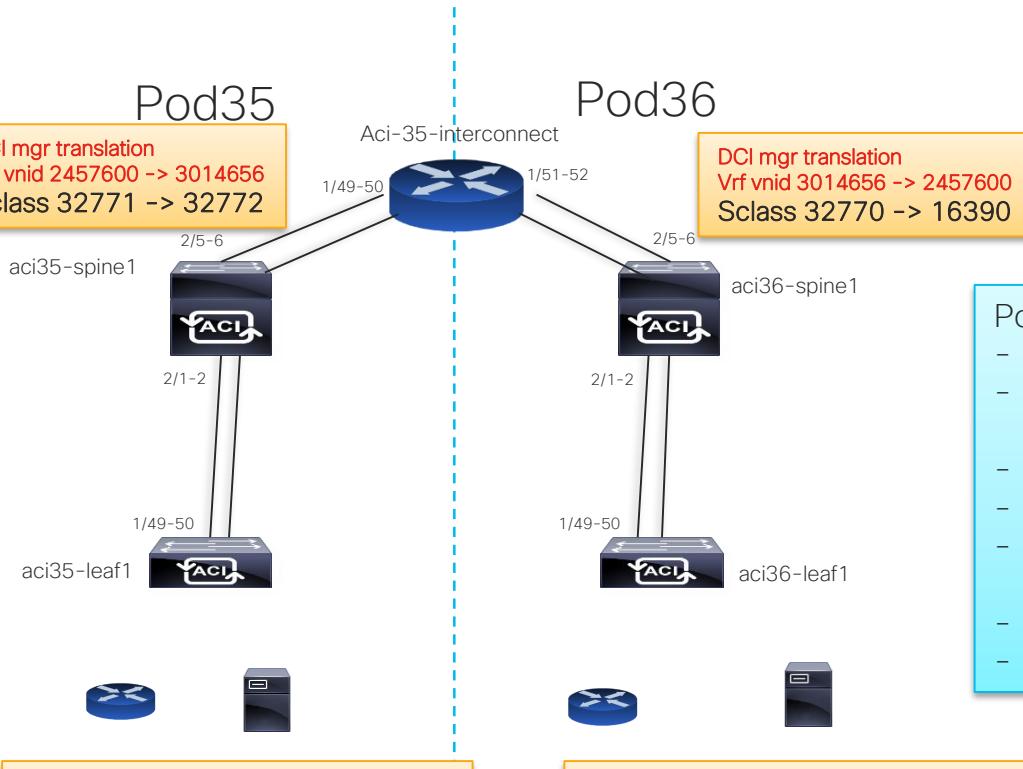
```
pod35-spine1# show lldp neighbors | egrep "Eth2/1"
pod35-leaf1           Eth2/1           120          BR          Eth1/49
pod35-spine1#
```

Data path - unknown EP Site 1 to Site 2 – 3.0 or 3.1



Policy enforcement

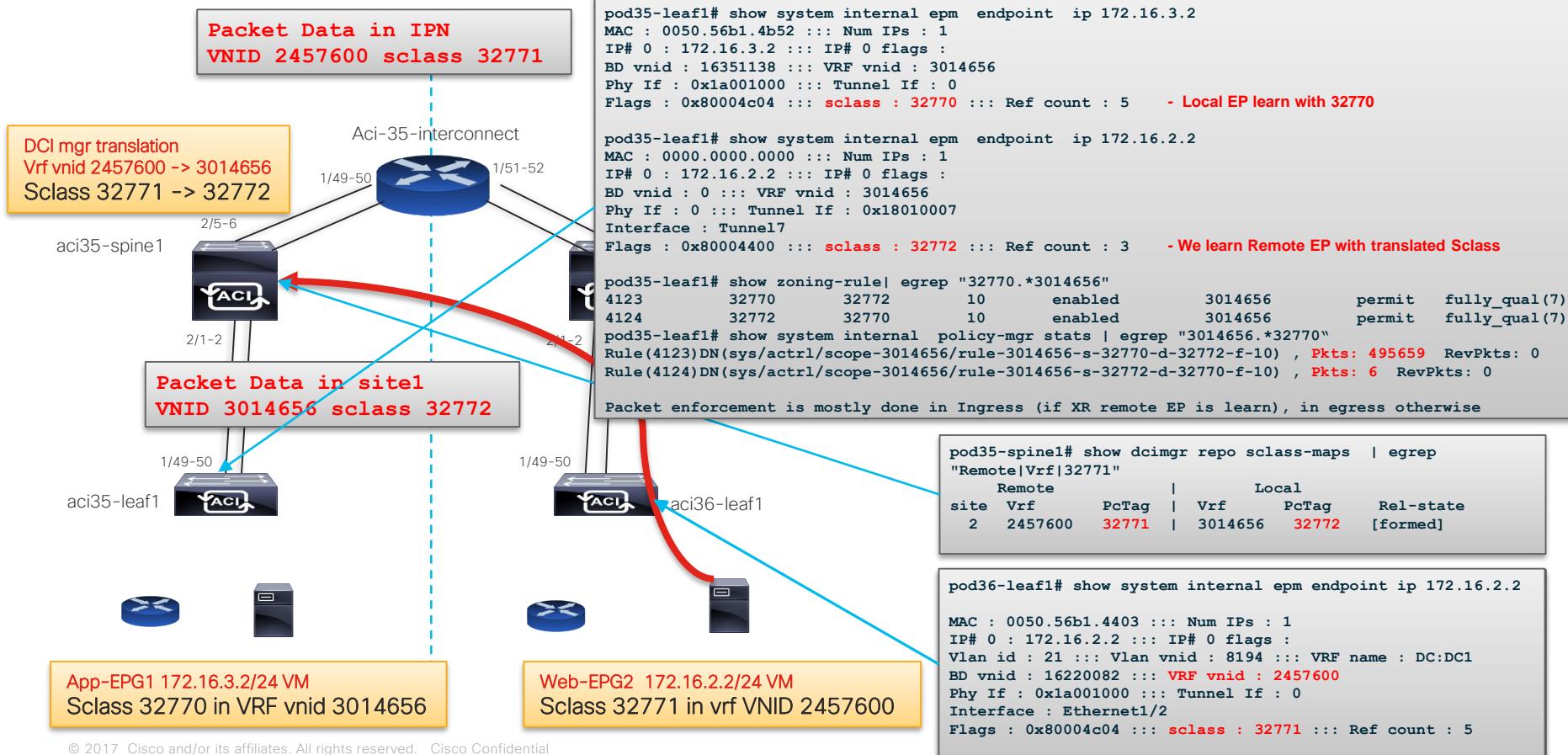
Sclass Translationg



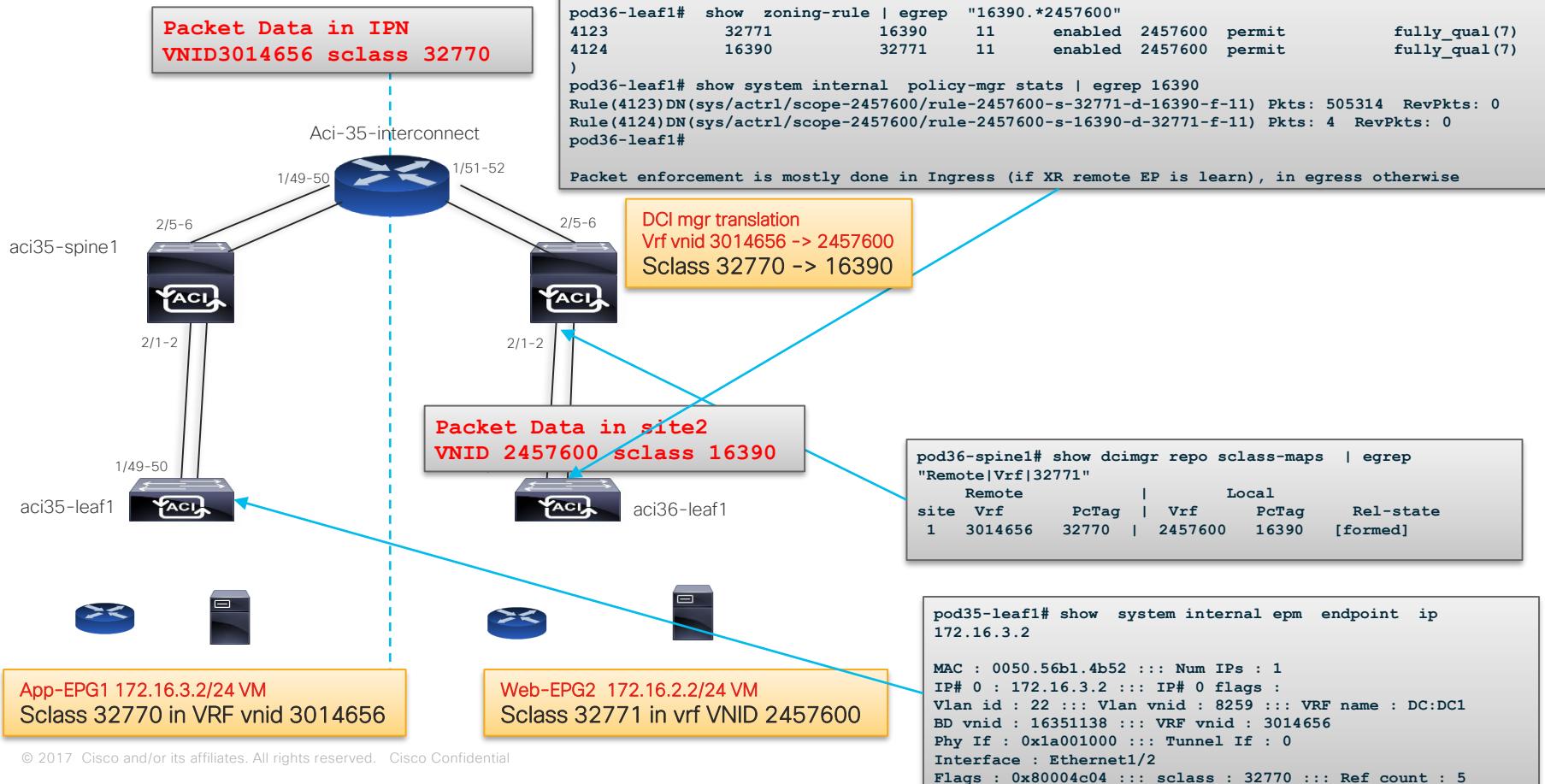
Policy Enforcement

- Ingress leaf derives sclass and vnid based on local EPM
- If Remote EPM is populated – Enforce Policy (as usual)
- Transmit to Remote Spine Site
- Remote spine site translate sclass and VNId
- - sent it to Dest leaf
- Dest leaf learn remote EP entry in translated sclass
- Enforce policy if not done on ingress

Policy from Site 2 (172.16.2.2) to Site 1 (172.16.3.2)



Policy from Site 1 (172.16.3.2) to Site 2 (172.16.3.2)



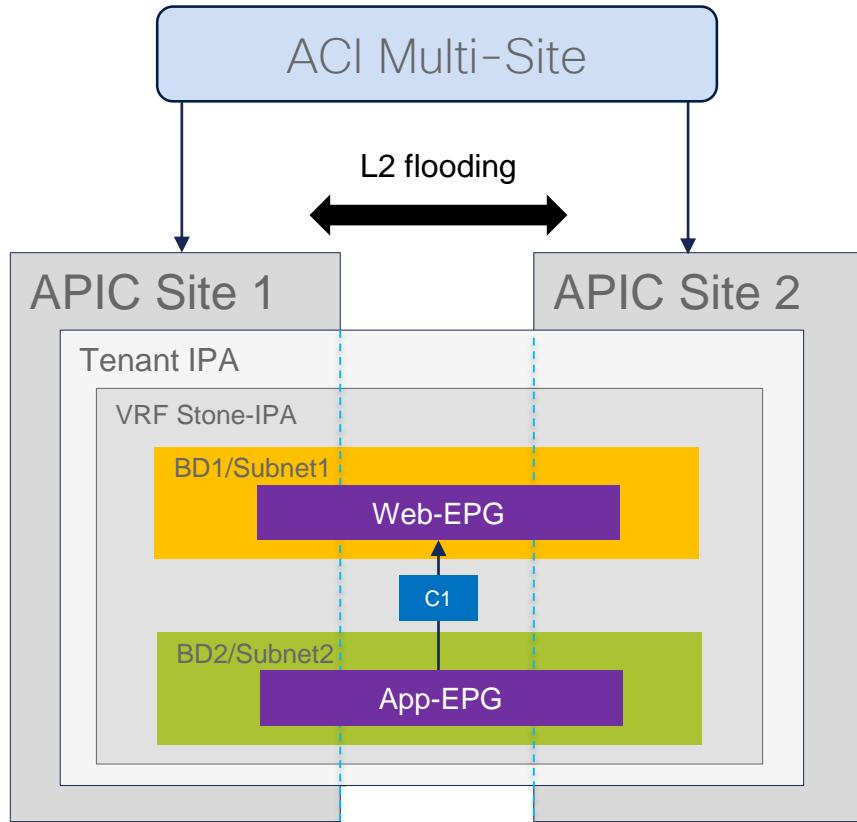
Multicast Multisite

Overview - Layer 2 BUM traffic across Sites

- TX (local to remote site)
 - GIPo (BUM) traffic sourced from the local site is Head-end replicated (HREP) to each remote site from the Spine. DIPo is rewritten to a unicast address called as Multicast HREP TEP IP (also called Multicast DP-TEP IP) of the remote site. SIPo is rewritten with the Unicast ETEP IP
- RX (remote to local site)
 - Incoming traffic destined to the local site's Multicast HREP TEP IP gets translated, derives the local site's BD-GIPo, and follows the regular GIPo lookup path from there

Multi-Site

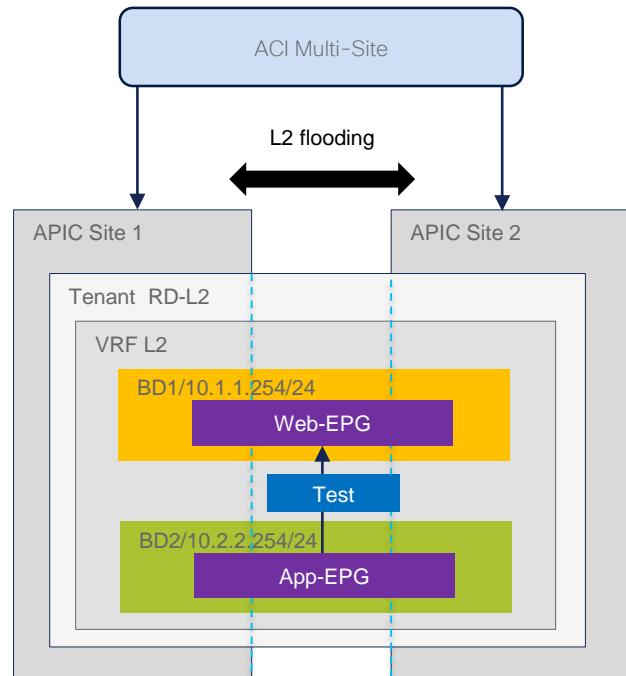
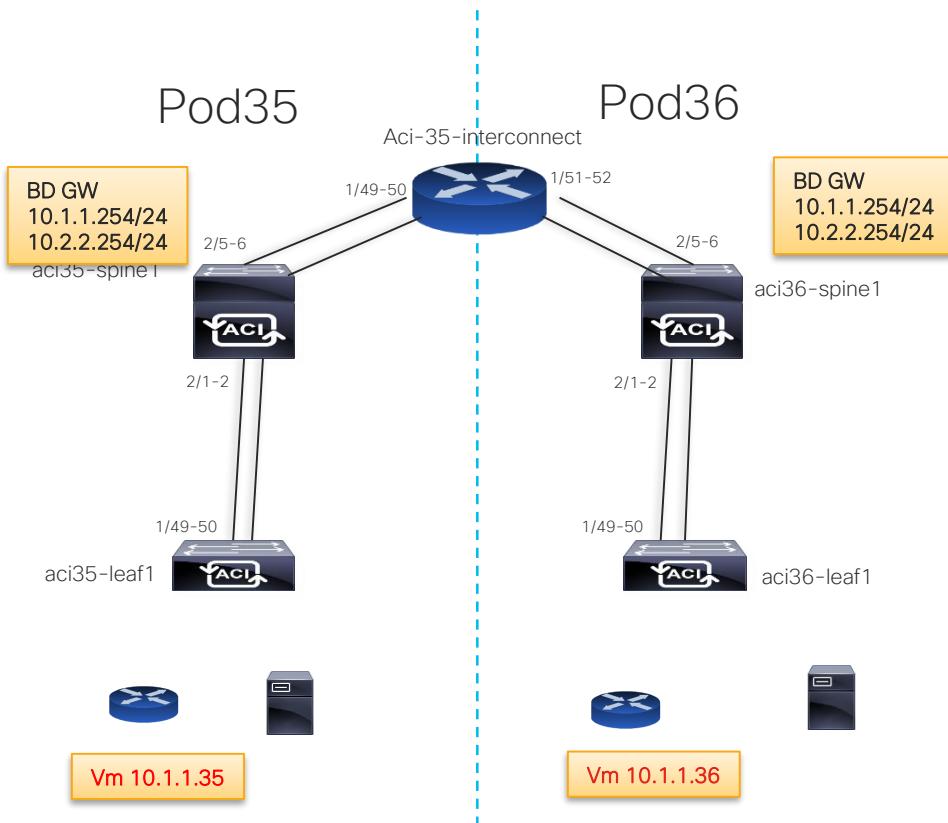
Stretched BD with L2 Broadcast Extension



Use Case Properties

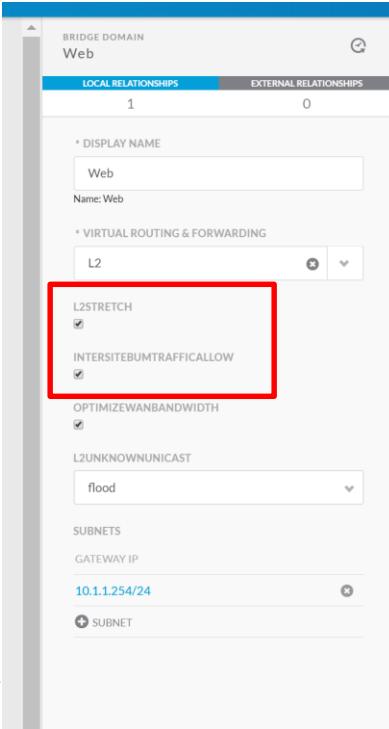
- Active/Active deployment with inter-site Layer 2 extension
- Objects stretched across sites:
 - Tenant ID
 - VRF context
 - BD/Subnet
 - Provider and Consumer EPGs
 - Policy between EPGs
- **L2 flooding enabled at the BD level**
 - L2 BUM traffic forwarded over head-end replicated VXLAN tunnels
- L2 application clustering and 'live' VM migration

Use case - lab VRF RD-L2:L2



Config Check

- BD must be set with intersite BUM allow flag



```
admin@bdsol-aci35-apic1:~> moquery -d uni/tn-RD-L2/BD-Web  
Total Objects shown: 1
```

```
# fv.BD  
name : Web  
OptimizeWanBandwidth : yes  
arpFlood : yes  
bcastP : 225.0.216.80  
childAction :  
configIssues :  
descr :  
dn : uni/tn-RD-L2/BD-Web  
epClear : no  
epMoveDetectMode :  
extMngdBy : msc  
intersiteBumTrafficAllow : yes  
intersiteL2Stretch : yes  
ipLearning : yes  
lcOwn : local  
limitIpLearnToSubnets : yes  
llAddr : ::  
mac : 00:22:BD:F8:19:FF  
mcastAllow : no  
modTs : 2018-05-03T03:14:39.650+00:00  
monPolDn : uni/tn-common/monepg-default  
mtu : inherit  
multiDstPktAct : bd-flood  
nameAlias :  
ownerKey :  
ownerTag :  
pcTag : 32770  
rn : BD-Web  
scope : 2457600  
seg : 15204288  
status :  
type : regular  
uid : 15374  
unicastRoute : yes  
unkMacUcastAct : flood  
unkMcastAct : flood  
vmac : not-applicable
```

Config check

- Multicast HREP TEP IP per Site
- Tunnel to each Remote site's Multicast HREP TEP

```
pod35-spine1# show ip interface vrf overlay-1 | egrep -A 1 mcast-hrep
loopback14, Interface status: protocol-up/link-up/admin-up, iod: 120, mode: mcast-hrep, vrf_vnid: 16777199
IP address: 10.10.35.121, IP subnet: 10.10.35.121/32

pod35-spine1# show interface tunnel 5
Tunnel5 is up
    MTU 9000 bytes, BW 9 Kbit
    Transport protocol is in VRF "overlay-1"
    Tunnel protocol/transport ivxlan
    Tunnel source 10.0.112.65, destination 10.10.35.122
```

Control Plane interaction

- ISIS

- For the stretched BDs (with intersiteBUMTrafficAllow), based on HREP-TEP configuration, ISIS adds the Remote site's HREP Tunnel If to the BD-GIPO of the Stretched BD.
- BD-GIPOs are striped across the Multisite-capable Spines – meaning HREP Tunnel If is added to BD-GIPO only on one of the Multi-site capable Spines in a site
- Unlike Multi-pod, no IGMP joins are sent out towards IPN, since native multicast is not used for forwarding BUM traffic across the sites

```
pod35-spine1# show isis internal mcast routes gipo | egrep -A 6 "225.0.216.80"
GIPO: 225.0.216.80 [LOCAL]
OIF List:
  Ethernet2/1.35
  Ethernet2/2.36
  Tunnel15
```

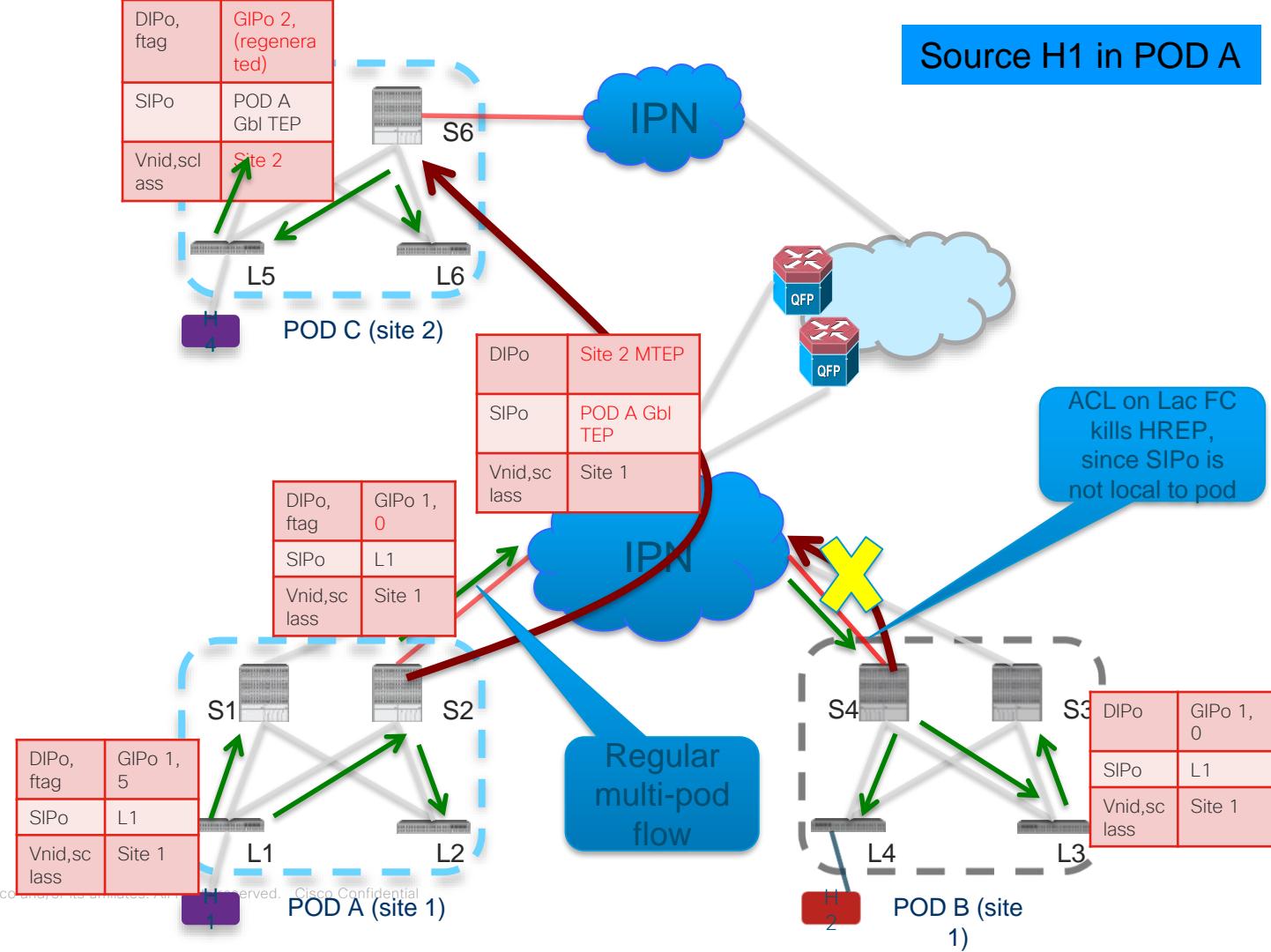
One spine per site
Should have Tunnel Interface as BD GIPO OIL

Forwarding path on TX Site (Multisite-capable Spine)

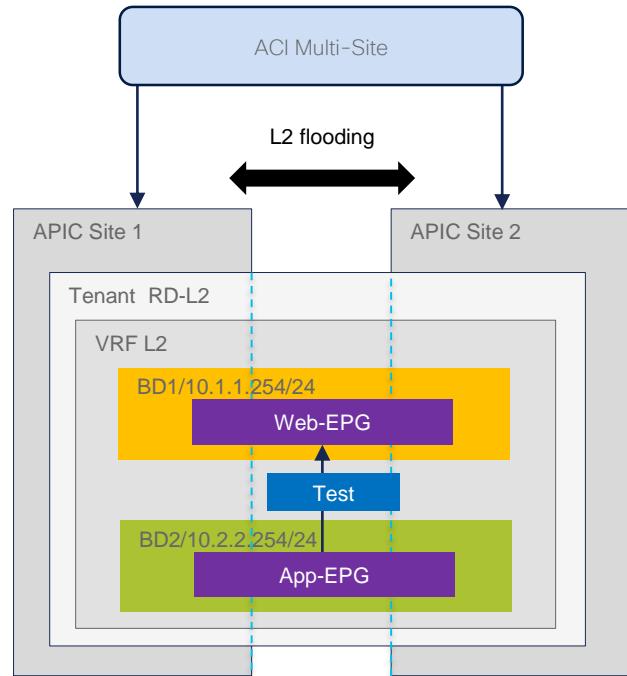
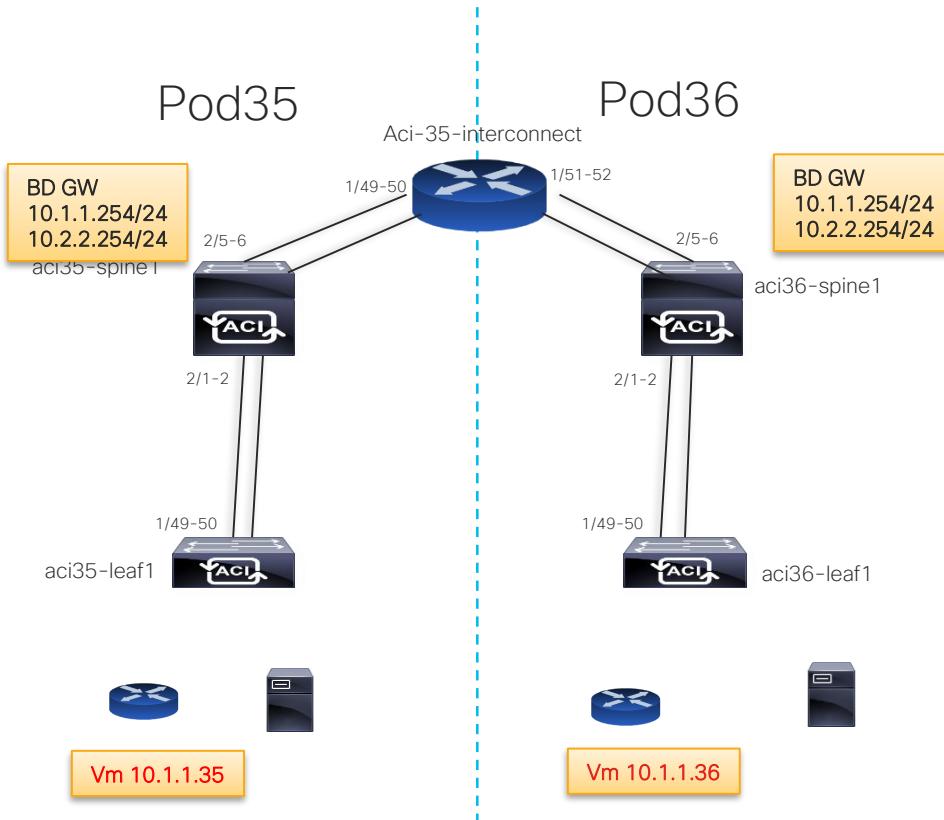
- Ingress LC/FC
 - ToR sends the BUM packet towards the spine
 - Ingress LC receives the packet with BD-GIPo as the outer dest IP
 - GIPo route lookup on FC results in a Replication List. Replication List contains 1 replication entry each for each of the Remote sites – HREP copy.
 - For the HREP copy, the packet is sent to Egress LC with some reserved internal vlan representing the corresponding Remote site.
- Egress LC
 - Egress LC interprets the internal vlan, and re-writes the outer DIPo with the remote site's Multicast HREP Tep IP. Outer Dmac is re-written with that of the selected NH
 - Also, SIPo gets re-written with Unicast ETEP IP of local site.
 - HREP copy of the packet is sent towards the IPN

Forwarding path on RX Site (Multisite capable Spine)

- Ingress LC
 - HREP packet is received from the IPN with SIPo = Remote site's unicast ETEP, and DIPo = local site's Mcast Hrep TEP IP. This will drive the pkt to DCI Multicast RX path.
 - Incoming vnid in the pkt is translated to local vnid/BD. BD lookup gives a gipo_idx
 - Gipo_idx is used to derive the local site's BD-GIPo. New Ftag is generated for the packet.
 - GIPo packet is then sent to FC
- FC
 - Packet follows the regular GIPo lookup path replicating it within the local site
 - DCI to DCI pruning on Egress path on LC ensures that packet is not sent back as a HREP packet.

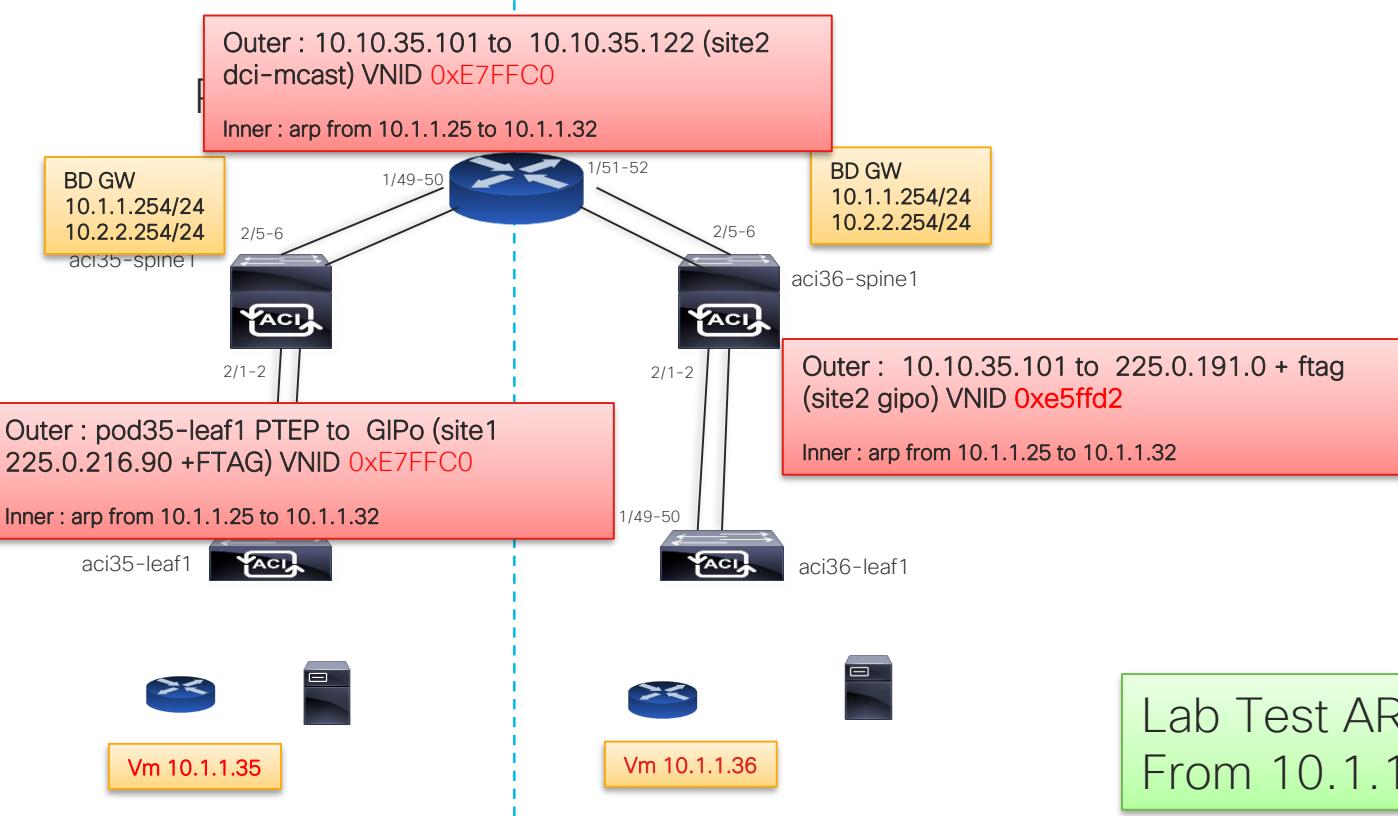


Use case - lab VRF RD-L2:L2



Lab Test ARP broadcast
From 10.1.1.35 to 10.1.1.32

Use case - lab VRF RD-L2:L2



Debugging on Spine TX site – hrep reach (vsh)

```
pod35-spine1# show forwarding distribution multicast hrep
MFDM HREP NODE TABLE
-----
IP Address: 0xa0a237a      - 10.10.35.122 (remote mcast-hrep address)
Table Id: 4
Flags: 0x0  Type: 1
IfIndex: 0x18010005
Internal BD 0x1001 bd_label 0x0 (hw_label 0x0)
Internal encap 0xb54
Nexthop Information: (num: 2)
Address          Ifindex          Dvif
0xa0a2302        0x1a084025      0x3 (Selected)    - Selected next-hop to reach hrep 10.10.35.2
0xa0a2306        0x1a085026      0x0

pod35-spine1# show ip route vrf overlay-1 10.10.35.122
IP Route Table for VRF "overlay-1"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.10.35.122/32, ubest/mbest: 2/0
  *via 10.10.35.2, Eth2/5.37, [110/3], 5w2d, ospf-default, intra
  *via 10.10.35.6, Eth2/6.38, [110/3], 5w2d, ospf-default, intra
```

Spine Tx – vsh_lc HREP reachability (mfdm and HAL)

```
module-2# show forwarding multicast hrep tep_routes
```

```
****HREP TEP ROUTES****
```

Tep Ip	Tep If	NH Ip	NH If	NH dmac	NH dvif	Vlan Id	Bd Id
10.10.35.122	0x18010005	10.10.35.2	0x1a084025	00a6.ca34.101f	3	2900	4097

```
module-2# show platform internal hal objects mcast hreptep extensions  
## Get Extended Objects for mcast hreptep for Asic 0
```

```
OBJECT 0:  
Handle : 38119  
tepifindex : 0x18010005  
tepipaddr : 10.10.35.122/0  
intbdid : 0x1001  
intvlanid : 0xb54  
nexthopipaddr : 10.10.35.2/0  
nexthopifindex : 0x1a084025  
nexthopmacaddr : 00:a6:ca:34:10:1f
```

GIPo route on line card

```
module-2# show forwarding multicast route group 225.0.216.80 vrf all  
(*, 225.0.216.80/32), RPF Interface: NULL, flags: Dc  
Received Packets: 0 Bytes: 0  
Number of Outgoing Interfaces: 3  
Outgoing Interface List Index: 15  
Ethernet2/1.35 Outgoing Packets:0 Bytes:0  
Ethernet2/2.36 Outgoing Packets:0 Bytes:0  
Tunnel15 Outgoing Packets:0 Bytes:0
```

HAL mcast route on LC and on FC

```
module-2# show platform internal hal objects mcast 13mcastroute groupaddr  
225.0.216.80/32 extensions  
## Get Extended Objects for mcast 13mcastroute for Asic 0  
  
OBJECT 0:  
Handle : 331521  
groupaddr : 225.0.216.80/32  
grpPrefixlen : 0x20  
sourceaddr : 0.0.0.0/32  
ispimbidir : Enabled  
ctrlflags :  
rtflags : OverlayGroup,  
OverlayDecapOnlyGroup,  
acirtpolicy : none  
aciepgid : 0x0  
aciclass : 0x0  
aciageinterval : 0x0  
minmtu : 0x5dc  
l3iif : 0x0  
rpfbdbd : 0x0  
id : 0x4  
  
    GPD OBJECT MCAST GPDL3VRFGROUP  
        APD OBJECT MCAST HOMASICPDL3VRFGROUP  
rtidx : id -1 type 0  
l2ptridx : 0x0  
overlaygrptblindex : id 5012 type 2  
Executing Custom Apd Private Handler function  
  
    VRF HWId : 0x4  
  
    Overlay Phytype : 1          HitIndex : 0x1ee  
                      dleft-494
```

```
module-24# show platform internal hal objects mcast 13mcastroute groupaddr  
225.0.216.80/32 extensions  
## Get Extended Objects for mcast 13mcastroute for Asic 0  
  
OBJECT 0:  
Handle : 438076  
groupaddr : 225.0.216.80/32  
grpPrefixlen : 0x20  
sourceaddr : 0.0.0.0/32  
ispimbidir : Enabled  
ctrlflags : UseMetFlag,  
rtflags : none,  
UseMetEntry,  
acirtpolicy : none  
aciepgid : 0x0  
aciclass : 0x0  
aciageinterval : 0x0  
minmtu : 0x5dc  
l3iif : 0x0  
rpfbdbd : 0x0  
id : 0x4  
  
Relation Object repllistnextobj :  
rel-repllistnextobj-mcast-mcast_mcast_repl_list-handle : 438073  
rel-repllistnextobj-mcast-mcast_mcast_repl_list-id : 0x6000000f  
  
    GPD OBJECT MCAST GPDL3VRFGROUP  
        APD OBJECT MCAST LACASICPDL3VRFGROUP  
rtidx : id 613 type 1  
l2ptridx : 0xd  
overlaygrptblindex : id -1 type 0  
Executing Custom Apd Private Handler function  
  
    VRF HWId : 0x4  
    Route Phytype : 5          HitIndex : 0xa20  
                      tile-entry-2592-tile-0-subtile-0-fp-id-0  
  
    L2Ptr Phytype : 5          HitIndex : 0xd  
                      tile-entry-13-tile-2-subtile-0-fp-id-4
```

HAL replist on Fabric card

```
module-24# show platform internal hal objects mcast mcastrepllist id 0x6000000f extensions
## Get Extended Objects for mcast mcastrepllist for Asic 0

    OBJECT 0:
Handle                      : 438073
id                          : 0x6000000f
rsvdmetptr                 : 0x0
ctrlflags                   :
Relation Object mcastreplentry :
    rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle : 438072
    rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id   : 0xe
Relation Object mcastreplentry :
    rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle : 438108
    rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id   : 0x10

    GPD OBJECT MCAST GPDMCASTREPLLIST

        APD OBJECT MCAST ASICPDMCASTREPLLIST
replindex                   : 0x6000000f
Executing Custom Apd Private Handler function

Repl-List Asicpd Debug :
Entry-Num 0
Repl Entry Id:      0xe Hw Epg Id:  4097 Hw Bd Id:  4097
Mc Id:              3 Met Id:     21 Encap Id:    -1
Sh Grp:             0 Next Met Id: 23
Entry-Num 1
Repl Entry Id:      0x10 Hw Epg Id:  4050 Hw Bd Id:  4050
Mc Id:              15 Met Id:    23 Encap Id:    -1
Sh Grp:             0 Next Met Id: 0
```

Elam ingress spine – ingress line card

```
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv4: 0x2801C103840361444480002
..
    hom_elam_in_13v_ipv4.da: 0xE100D851      225.0.216.81
    hom_elam_in_13v_ipv4.sa: 0xA007040      10.0.112.64
*** Parsed Outer 14 vector
hom_elam_in_14v_tn: 0x39FFF000002000F2
    hom_elam_in_14v_tn.14_type: 0x2
    hom_elam_in_14v_tn.tn_nonce: 0x1
    hom_elam_in_14v_tn.tn_lsb: 0x1
    hom_elam_in_14v_tn.tn_nonce_info: 0x8003
    hom_elam_in_14v_tn.tn_lsb_info: 0x0
    hom_elam_in_14v_tn.tn_seg_id: 0xE7FFC0
*** Parsed Inner 13 ARP vector
hom_elam_in_13v_arp:
0x2804048000000000002804048C000410182000201804
    hom_elam_in_13v_arp.13_type: 0x4
    hom_elam_in_13v_arp.pyld_len: 0x0
    hom_elam_in_13v_arp.etype: 0x806
    hom_elam_in_13v_arp.pro: 0x800
    hom_elam_in_13v_arp.hln: 0x6
    hom_elam_in_13v_arp.pln: 0x4
    hom_elam_in_13v_arp.op: 0x1
    hom_elam_in_13v_arp.spa: 0XA010123
    hom_elam_in_13v_arp.tha: 0x0000000000
    hom_elam_in_13v_arp.tpa: 0XA010120
```

DCI RX – 1/ Find out GIPo in dest site

```
bdsol-aci36-apic1# moquery -d uni/tn-RD-L2/BD-Web
Total Objects shown: 1

# fv.BD
name          : Web
OptimizeWanBandwidth   : yes
arpFlood      : yes
bcastP       : 225.0.191.0
childAction    :
configIssues   :
descr         :
dn            : uni/tn-RD-L2/BD-Web
epClear        : no
epMoveDetectMode  :
extMngdBy     : msc
intersiteBumTrafficAllow : yes
intersiteL2Stretch  : yes
ipLearning     : yes
lcOwn         : local
limitIpLearnToSubnets : yes
llAddr         : ::
mac           : 00:22:BD:F8:19:FF
mcastAllow    : no
modTs         : 2018-05-03T03:14:31.253+00:00
monPolDn      : uni/tn-common/monepg-default
mtu           : inherit
multiDstPktAct  : bd-flood
nameAlias     :
ownerKey      :
ownerTag      :
pcTag         : 49153
rn            : BD-Web
scope         : 2162688
seg           : 15073234
status        :
```

DCI RX – BD vnid GIPo map

```
module-2# show platform internal hal objects dci vnidmap extensions | egrep -A 20 -B 5 "225.0.191.0"

OBJECT 7:
Handle : 337422
isbdvnid : Enabled
localvnid : 0xe5ffd2
localgipo : 225.0.191.0/32
remotevnid : 0xe7ffc0
remotevrfnid : 0x258000
islocalbdctrl : Enabled
siteid : 0x1

      GPD OBJECT DCI GLOBALPDVNIDMAP
gipoidx : 0x0
spineidx : 0x3c0e
bdstateidx : 0x3c0e

      APD OBJECT DCI HOMASICPDVNIDMAP
vlanxlateidx : id 5034 type 2
vlanxlateidx_hw : 0xba9
```

Packet BUM RX with VNID BD 0xe7ffc0 will have
DIP to GIPo 225.0.191.0
Vnid is translated to 0xe5ffd2

ELAM DCI RX spine

- Elam on rx spine.
- We see vnid translate (before and after)

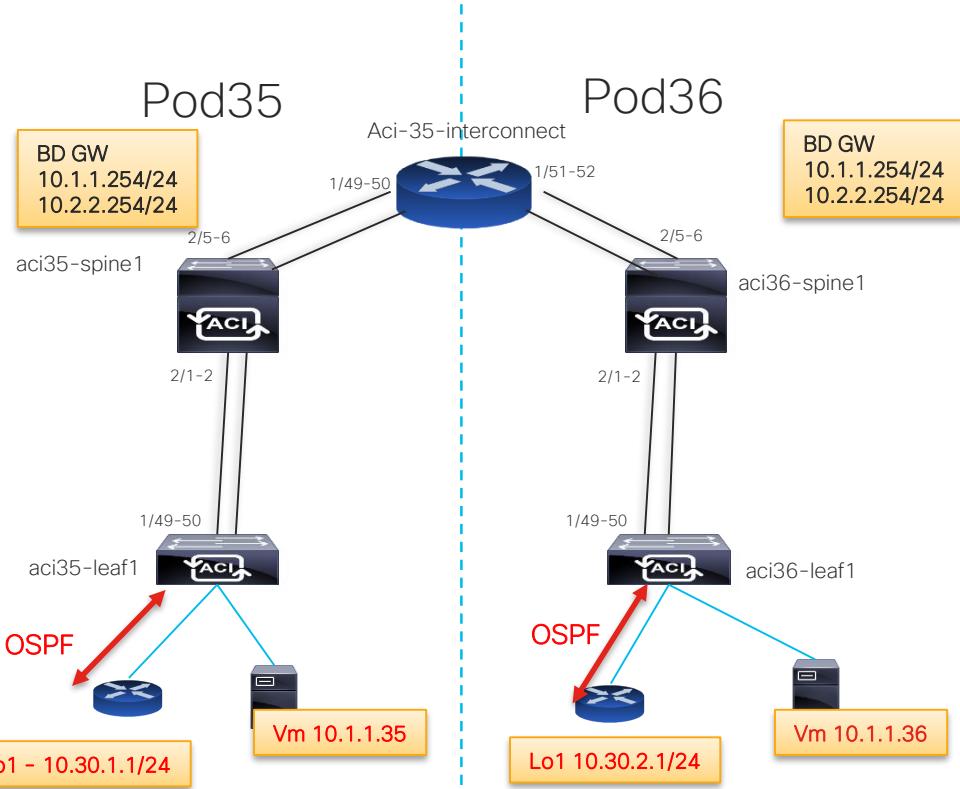
```
*** Parsed Outer 13 vector
hom_elam_in_13v_ipv4: 0x28288D9428288DE84478002
..
    hom_elam_in_13v_ipv4.da: 0xA0A237A 10.10.35.122
    hom_elam_in_13v_ipv4.sa: 0xA0A2365 10.10.35.101

*** Parsed Outer 14 vector
hom_elam_in_14v_tn: 0x39FFF000002000F2
    hom_elam_in_14v_tn.tn_nonce_info: 0x8003
    hom_elam_in_14v_tn.tn_lsb_info: 0x0
    hom_elam_in_14v_tn.tn_seg_id: 0xE7FFC0

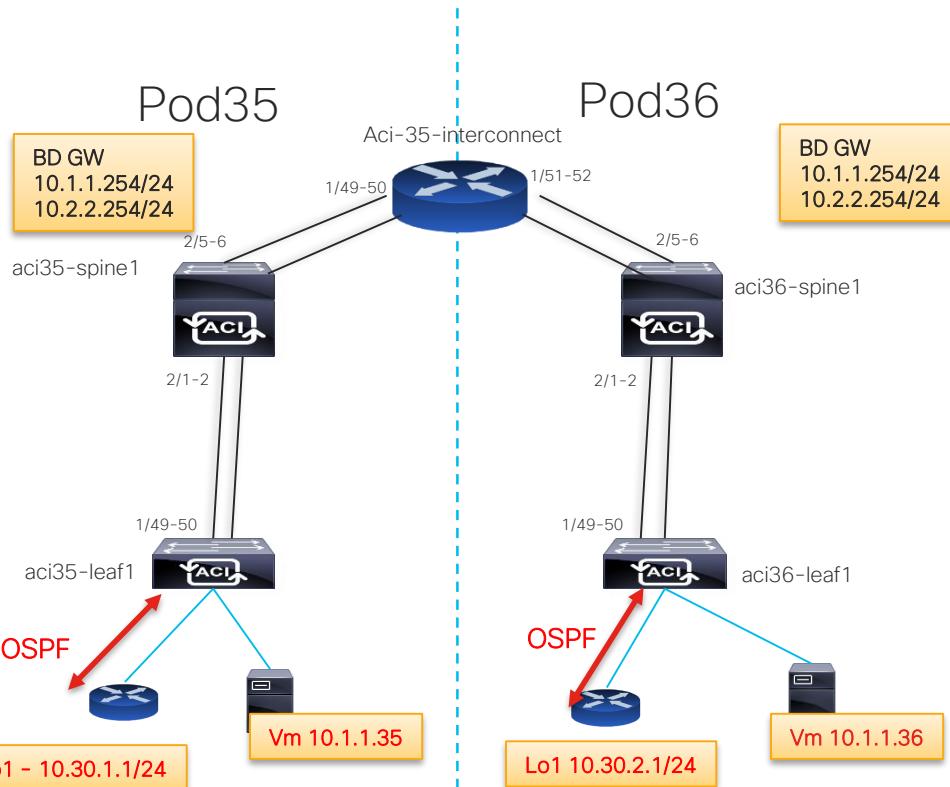
hom_lurw_vec.info.ifabric_spine.gipo_idx: 0xBF0
hom_lurw_vec.info.ifabric_spine.is_bd_vnid: 0x1
hom_lurw_vec.info.ifabric_spine.vnid: 0xE5FFD2
```

L3 out and Multisite

Use case 3 – lab VRF RD-L2:L2



L3 out - lab VRF RD-L2:L2 (tested with 3.1)



• Working Session :

- in EPG web : 10.1.1.35 to 10.1.1.36

• Local L3 out :

- 10.1.1.35 to 10.30.1.1
- 10.1.1.36 to 10.30.2.1

• Non working connection (expected)

- 10.1.1.35 to 10.30.2.1
- 10.1.1.36 to 10.30.1.1

• Might or not be Working Direction (return from L3 out):

- 10.30.2.1 can reach 10.1.1.35
- 10.30.1.1 can reach 10.1.1.36

• Non working direction (from VM to L3 remote):

- 10.1.1.35 to 10.30.2.1
- 10.1.1.36 to 10.30.1.1

Why EP to remote L3 out do not work

- No VPNv4 route exchange across multisite BGP session
- No l2vpn evpn type 5 neither
- Site 2 never got route from Site 1 – L3 out

Only l2vpn evpn capa nego with Peer on the intersite)

```
pod35-spine1# show bgp l2vpn evpn neigh 10.10.35.112 vrf overlay-1 | egrep -A 1  
"capabili"  
  
Additional Paths capability: advertised received  
Additional Paths Capability Parameters:  
Send capability advertised to Peer for AF:  
    L2VPN EVPN  
Receive capability advertised to Peer for AF:  
    L2VPN EVPN
```

No l2vpn evpn type5 for subnet neither is advert

```
pod35-spine1# show bgp l2vpn evpn neigh 10.10.35.112 adver vrf overlay-1 | egrep  
"10.10.30"  
pod35-spine1#
```

Traffic return from L3 out

- Will work if no subnet configured on L3 out epg (or 0.0.0.0/0)
- Will not work if any subnet is configured under the L3 out with ext subnet for external EPG
- See MSC release notes :
https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/aci_multi-site/sw/1x/release_notes/Cisco_ACI_Multi-Site_RN_112.html
- **NOTE:** The subnet in the L3extInstP must be the same for all inter-related sites (and variable length network masks are not supported).

Why does it work from L3 out to remote EP

Example from 10.30.1.1 to 10.1.1.36

PcTag ingress Site

```
admin@bdsol-aci35-apic1:~> moquery -d uni/tn-RD-L2/out-L3-35/instP-epg-35 | egrep "dn|pcTag|scope"
dn      : uni/tn-RD-L2/out-L3-35/instP-epg-35
pcTag   : 32772
scope   : 2457600

bdsol-aci35-apic1# moquery -d uni/tn-RD-L2/ctx-L2 | egrep "dn|pcTag"
dn      : uni/tn-RD-L2/ctx-L2
pcTag   : 16386
```

Elam ingress leaf – setting pcTag to vrf pcTag

```
hom_lurw_vec.info.ifabric_leaf.dclass: 0x1
    hom_lurw_vec.info.ifabric_leaf.sclass: 0x4002
module-1(DBG=elam-insel6)# dec 0x4002
16386
```

Egress spine translate 16386 to 32700

```
pod36-spine1# show dcimgr repo sclass-maps | egrep "2162688"
-----
          Remote           |           Local
site  Vrf     PcTag  |  Vrf     PcTag  Rel-state
-----
1     2457600  16386  |  2162688  32770  [formed]
```

Zoning rule on egress leaf

```
pod36-leaf1# show sys internal policy-mgr stat | egrep "4133"
Rule (4133) DN (sys/acctrl/scope-2162688/rule-2162688-s-32770-d-49155-f-default) Ingress: 0, Egress: 0, Pkts: 1475 RevPkts: 0
pod36-leaf1# show zoning-rule | egrep "4133"
4133      32770      49155      default      enabled      2162688      permit
src_dst_any(9)
```

Why would that flow break ? from L3 out to remote EP Example from 10.30.1.1 to 10.1.1.36

I added in site1 I3 out subnet 10.30.0.0/16 as ext to ext epg
→ Ingress leaf do not drive pcTag from vrf but from L3 out

```
module-1(DBG-elam-insel6) # show system internal aclqos prefix
Vrf-Vni VRF-Id Table-Id          Addr          Class Shared Remote Complete
2457600 8      0x8             10.30.0.0/16      32772  0      0      No
```

Elam ingress leaf – setting pcTag to vrf pcTag

```
hom_lurw_vec.info.ifabric_leaf.dclass: 0x1
    hom_lurw_vec.info.ifabric_leaf.sclass: 0x8004
module-1(DBG-elam-insel6) # dec 0x8004
32772
```

Egress spine translate 16386 to 32700

```
pod36-spine1# show dcimgr repo sclass-maps | egrep "2162688"
 1  2457600  16386  |  2162688  32770  [formed]
 1  2457600  32771  |  2162688  49155  [formed]
 1  2457600  49154  |  2162688  16386  [formed]
```

Summary

EP Control plane

- EP known locally to COOP
- Sent as BGP I2vpn evpn path in Global EVI 1
- Imported in EVI of the BD
- Send with EVI of the BD
- Rx on remote site imported to BD EVI per RT and imported to EVI 1
- EVI1 to COOP on remote Site
- BD subnet or I3 out routes are not send by BGP
- BD subnet is send to remote site per Object model pushed by MSC to both APIC SITE

Unicast Dataplane

- EP unknown in ingress leaf (proxy)
 - Ingress Leaf Site 1 – ingress Spine (SIPo, DIPo, Rwp) – Egress Spine – Egress spine Xlate VNID and SCLASS – egress leaf learns EPM to Spine ingress site
- EP known in ingress leaf
 - Ingress leaf direct to egress site spine – Egress spine make Xlate vnid, sclass and send to egress leaf per COOP
- EP unknown (silent) – Drop if L3 as no ARP glean support before 3.2

VNID and SCLASS translate is ALWAYS ON egress site SPINE
Ingress Border spine always rewrite Source IP to its own DP-ETEP

Multicast Dataplane

- Only for BUM inside a BD
- Ingress spine do Head End Rep to Egress site spine Mcast-ETEP

Translation on spine

- Coming from MSC pushing to each APIC site logical object for VRF/BD/EPG/L3 epg that needs to be extended
- Concrete MO resulting from that on spine
- Concrete MO triggers DCIMGR on sup
- DCIMGR client on lc
- HAL creates Translation to ASIC