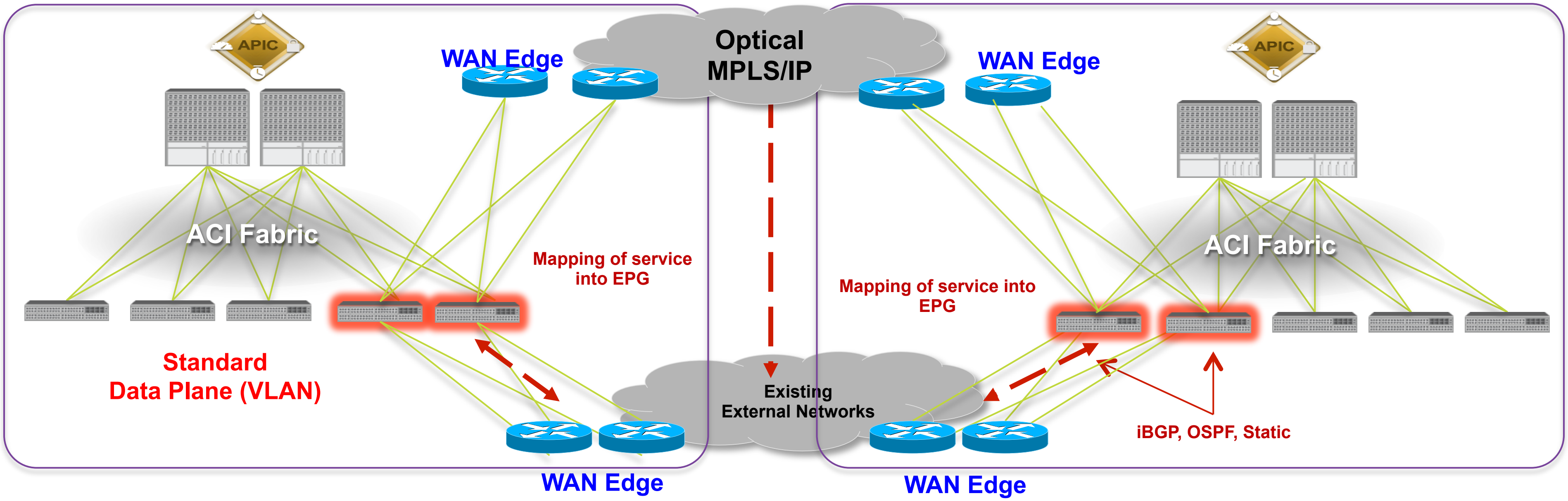
ACI Integration to Outside Network

Agenda

- ACI External Connectivity Use Cases
- ACI L2 Connection to Outside Network
- ACI L3 Connection to Outside Network
- Q&A

ACI Connection to Outside Network

Use Cases


APIC
WAN Edge
Optical MPLS/IP
WAN Edge
APIC
ACI Fabric
Mapping of service into EPG
Mapping of service into EPG
ACI Fabric
Standard Data Plane (VLAN)
Existing External Networks
iBGP, OSPF, Static
WAN Edge
WAN Edge

WAN and DCI Connection

- Targets IP and Ethernet (DCI) connectivity in/out of fabric services
- Leverage standard routing protocols with fabric/standalone routing domain
- Mapping external network entities (IP address, subnet, .1Q) to fabric (EPG)
- WAN Edge focus: ASR 9000, Nexus 7000, ASR 1000
- Existing principles of Inbound, Outbound traffic flows, security, DNS/GSS still apply

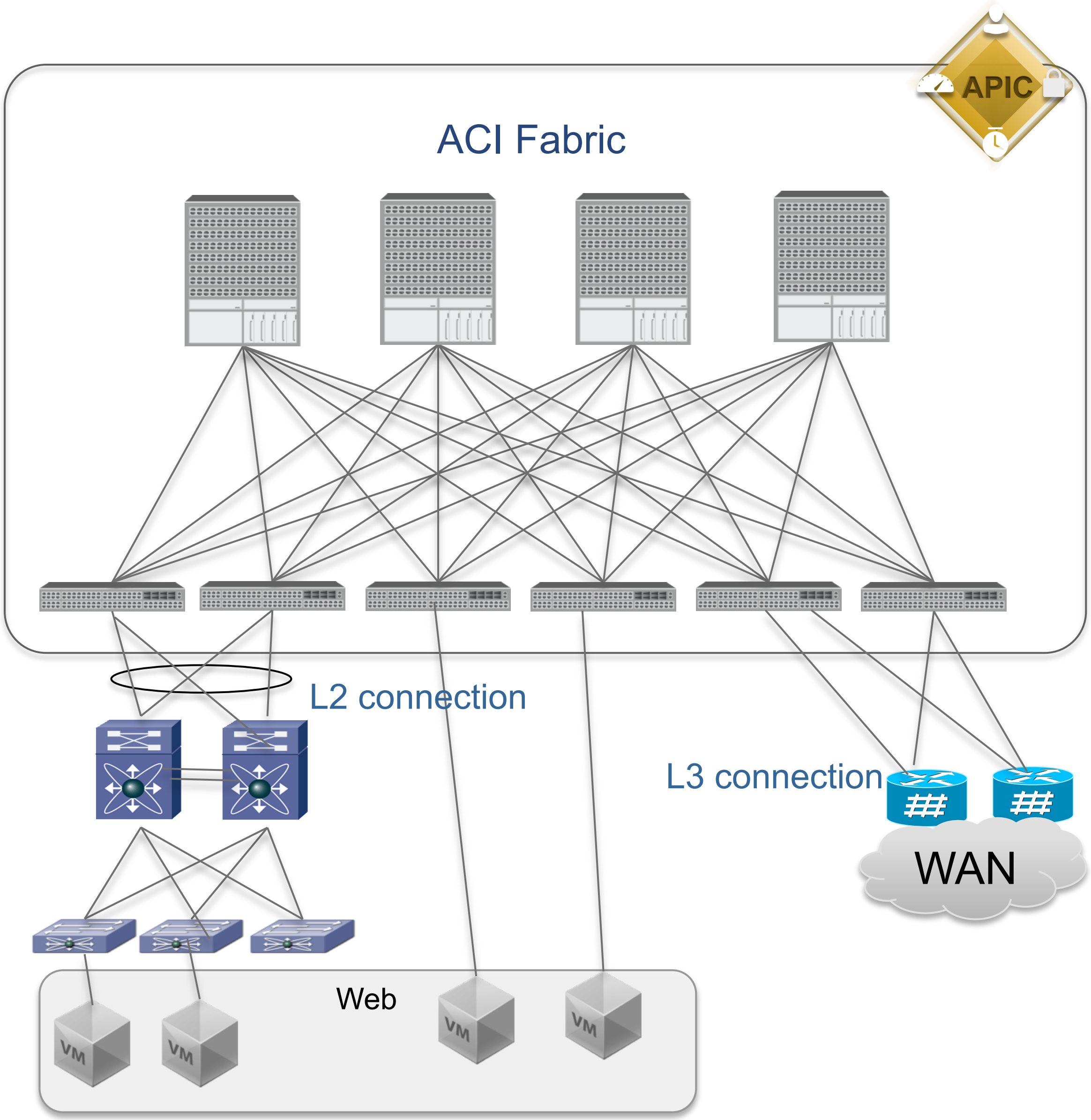
Brownfield connection/migration

- Connecting to existing DC network/server via standard L2 and L3 technology

ACI Connection to Outside Network

Options

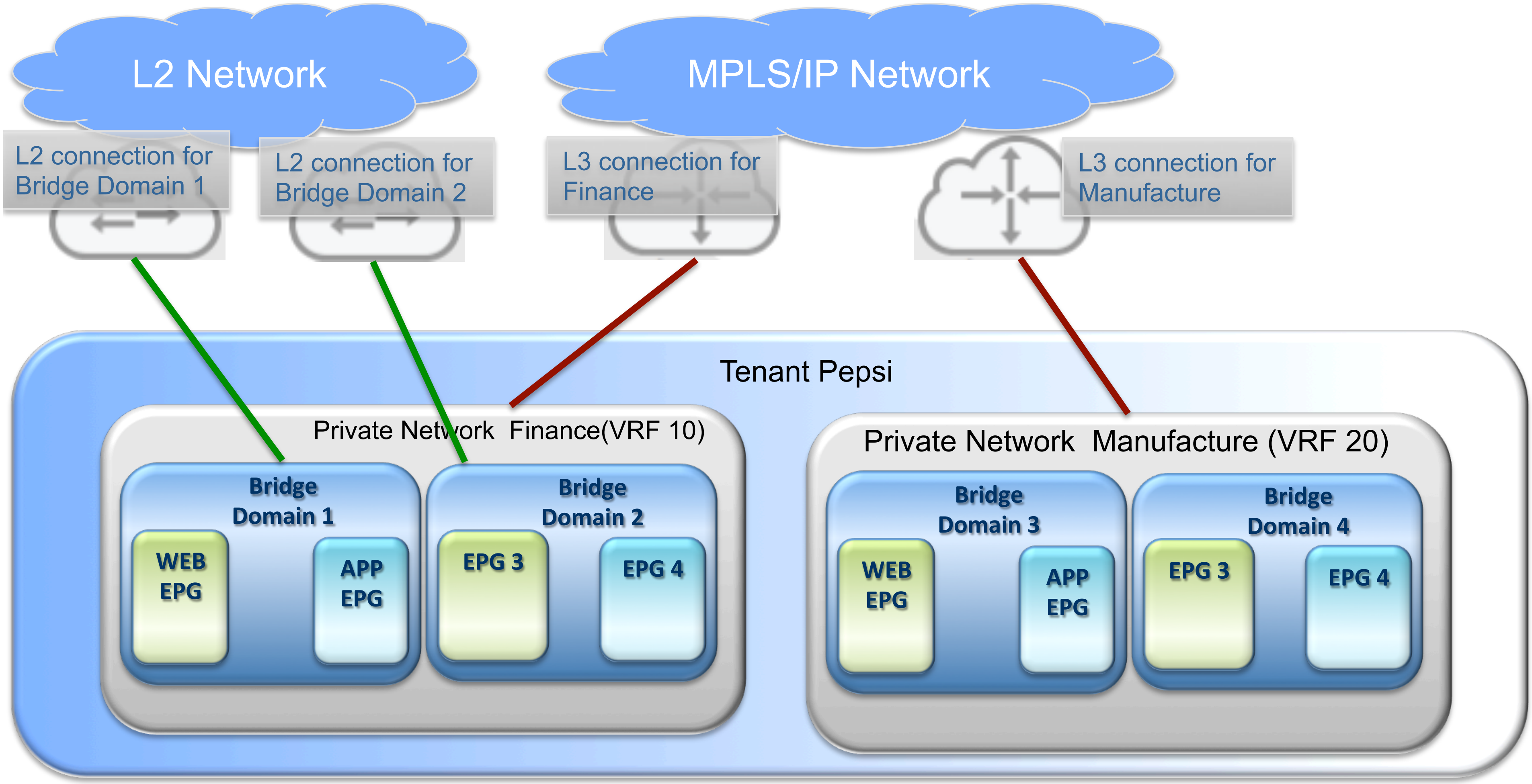
- L3 connection to outside
  - Provide L3 connection for tenants
  - Connecting to existing DC network
  - VRF-lite for tenant isolation
  - OSPFv2 ,iBGP and static route at FCS
- L2 connection to outside
  - Extend L2 domain outside of ACI fabric
  - Brownfield migration
  - L2 extend across POD/site
  - Support VLAN and VXLAN for tagging
  - vPC and STP connection at FCS



ACI Connection to Outside Network

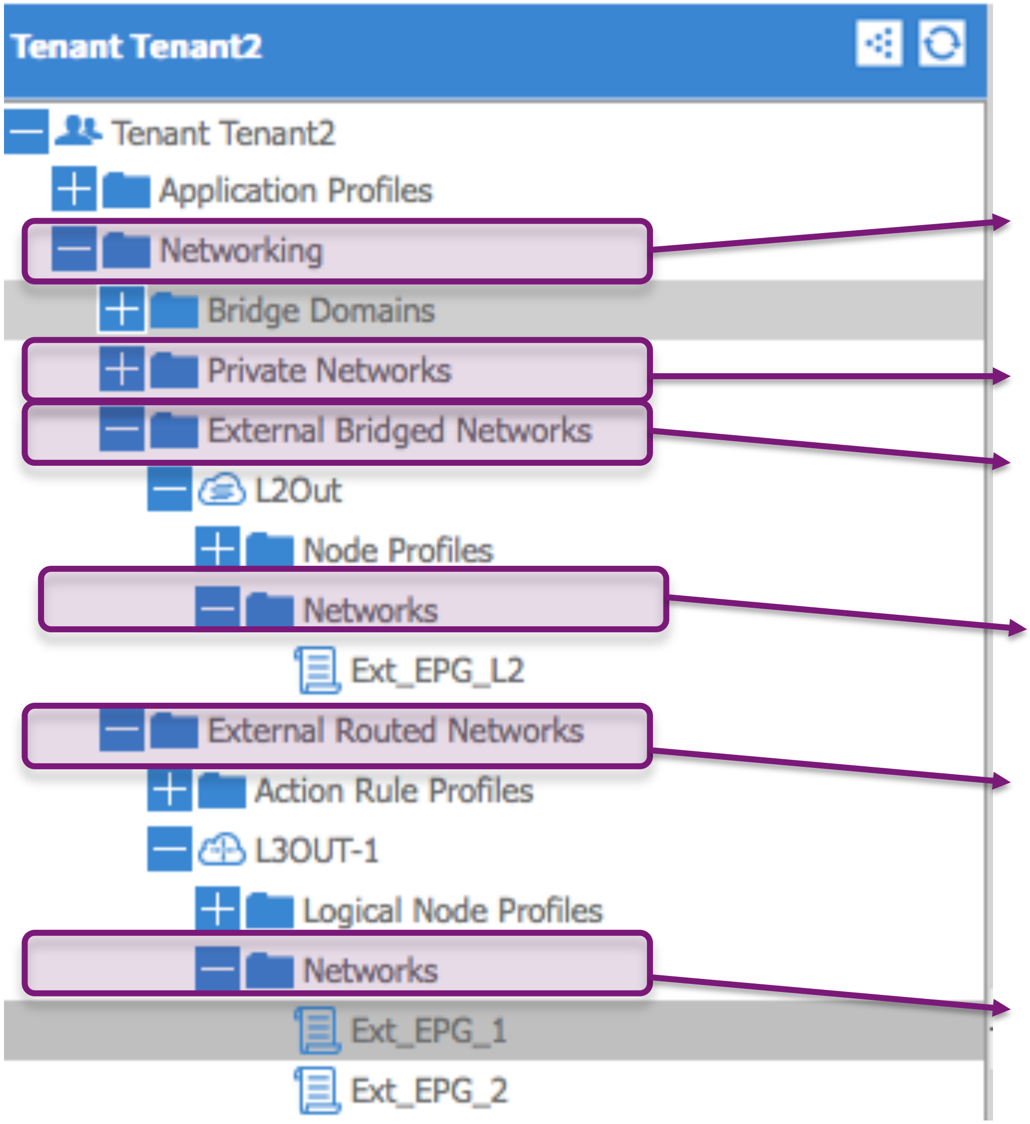
Relationship to rest of components(Connectivity view)

- Conceptual representation. Some components are not included. Some scenarios are not represented.


L2 Network
MPLS/IP Network
L2 connection for Bridge Domain 1
L2 connection for Bridge Domain 2
L3 connection for Finance
L3 connection for Manufacture
Tenant Pepsi
Private Network Finance(VRF 10)
Bridge Domain 1
WEB EPG
APP EPG
Bridge Domain 2
EPG 3
EPG 4
Private Network Manufacture (VRF 20)
Bridge Domain 3
WEB EPG
APP EPG
Bridge Domain 4
EPG 3
EPG 4

What is a Network on APIC

- The keyword “Network” is overloaded on APIC



Tenant Tenant2

- Tenant Tenant2
  - Application Profiles
  - Networking
    - Bridge Domains
    - Private Networks
    - External Bridged Networks
      - L2Out
        - Node Profiles
        - Networks
          - Ext_EPG_L2
    - External Routed Networks
      - Action Rule Profiles
      - L3OUT-1
        - Logical Node Profiles
        - Networks
          - Ext_EPG_1
          - Ext_EPG_2

Folder for all network related configuration for the tenant

VRF or L3 context

L2 connection to external network

External EPG reachable via this L2 connection. One per L2 connection

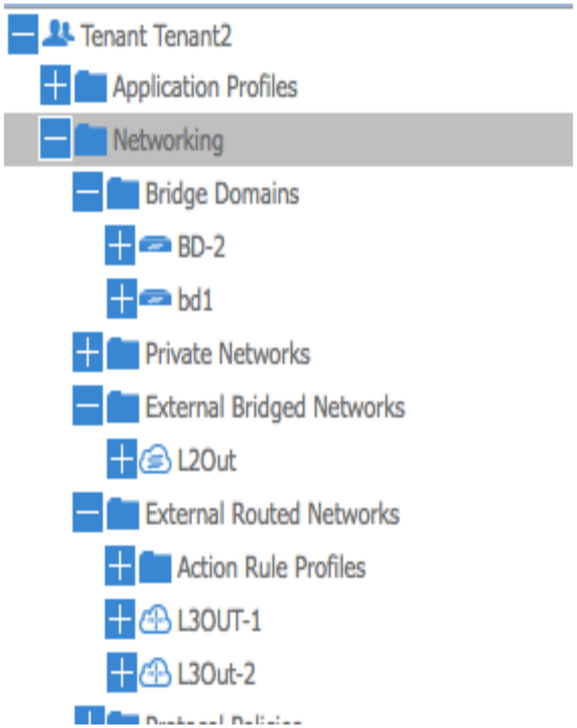
L3 connection to external network

External EPG reachable via this L3 connection.

ACI Connection to Outside Network

Relationship to rest of components(Connectivity view)

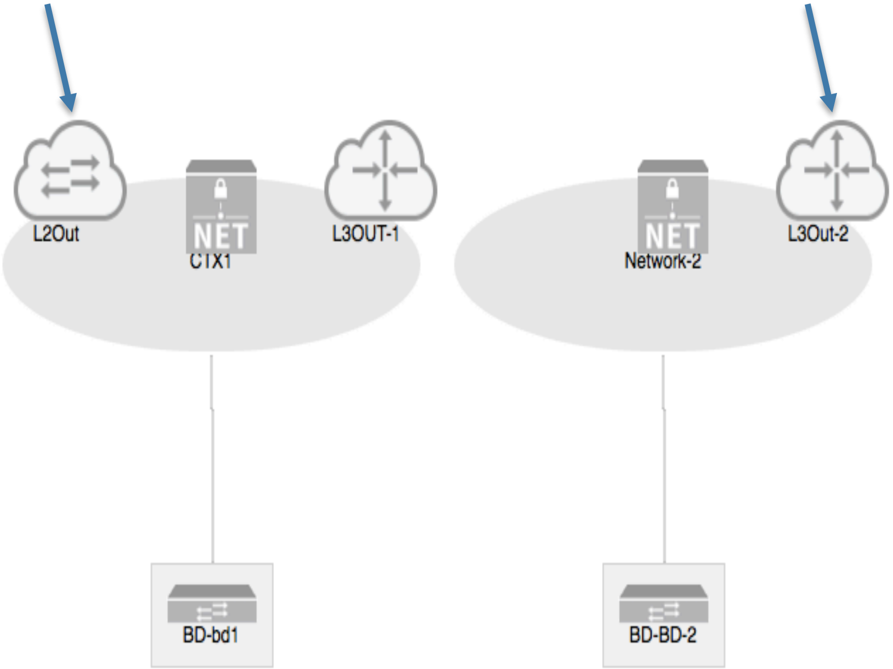
- The Network view on the APIC GUI



- Tenant Tenant2
  - Application Profiles
  - Networking
    - Bridge Domains
      - BD-2
      - bd1
    - Private Networks
    - External Bridged Networks
      - L2Out
    - External Routed Networks
      - Action Rule Profiles
      - L3OUT-1
      - L3Out-2
    - Protocol Policies

L2 connection to external network for bridge domain “bd1”

L3 connection to external network for private network(VRF) “Network-2”

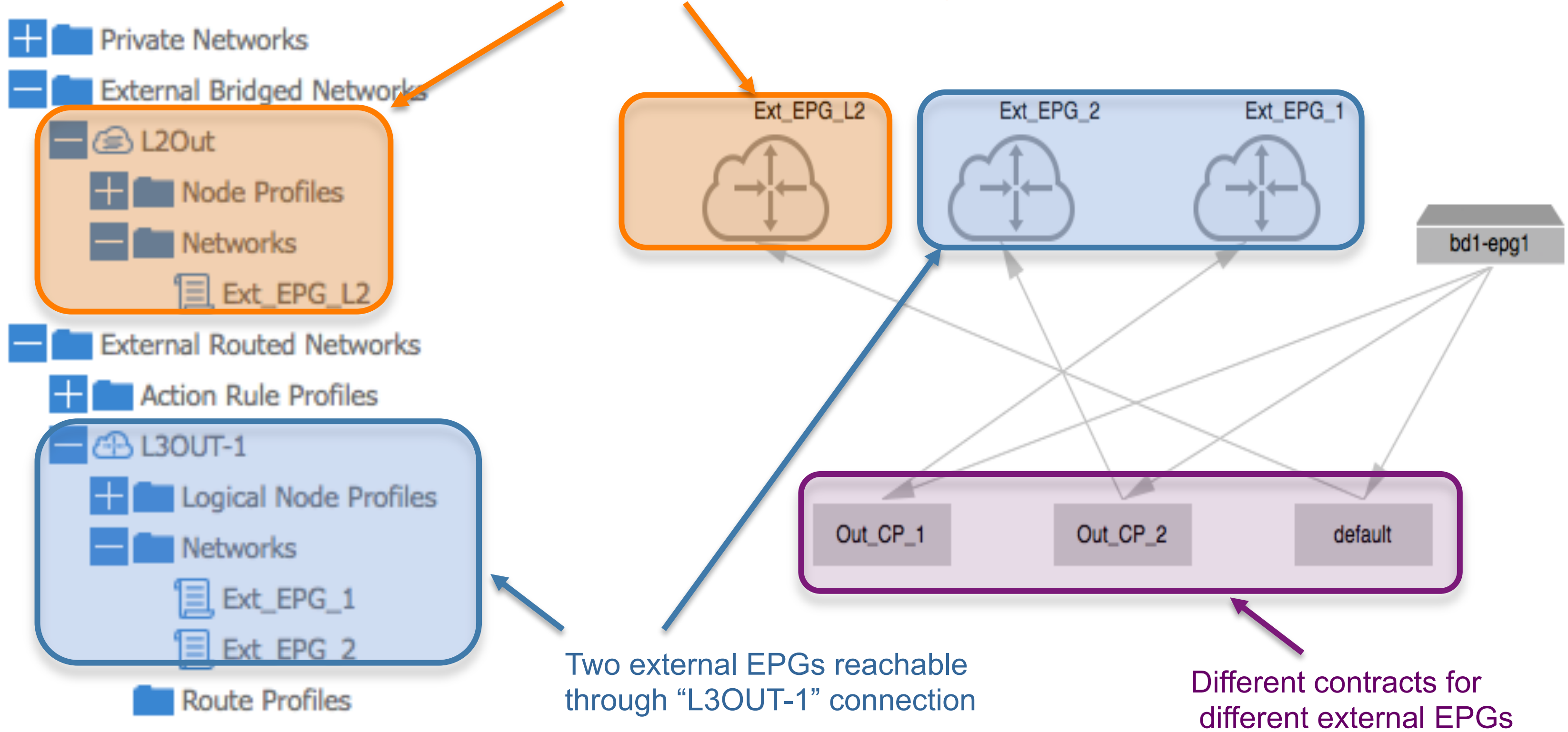


ACI Connection to Outside Network

Policy View

- The Policy view on the APIC GUI with respect to external connection

External EPG reachable through “L2Out” connection



- Private Networks
- External Bridged Networks
  - L2Out
    - Node Profiles
    - Networks
      - Ext_EPG_L2
- External Routed Networks
  - Action Rule Profiles
  - L3OUT-1
    - Logical Node Profiles
    - Networks
      - Ext_EPG_1
      - Ext_EPG_2
    - Route Profiles

Two external EPGs reachable through “L3OUT-1” connection

Different contracts for different external EPGs

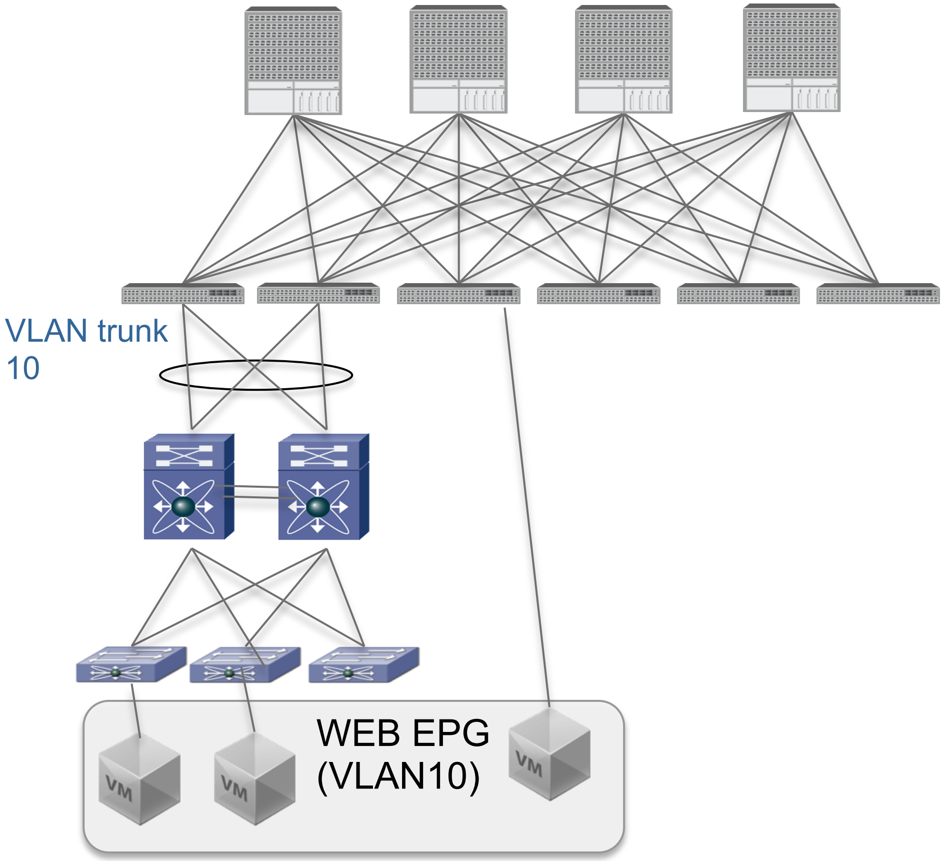
Extend L2 Domain Out of ACI

- **Three ways of extend L2 domain beyond ACI fabric**
  - Manually assign a port to a VLAN which in turn mapped to an EPG. This **extend EPG beyond ACI fabric**
  - Create a L2 connection to outside network. **Extend bridge domain** beyond ACI fabric. Allow contract between EPG inside ACI and EPG outside of ACI
  - Remote VTEP

Extend L2 Domain Out of ACI

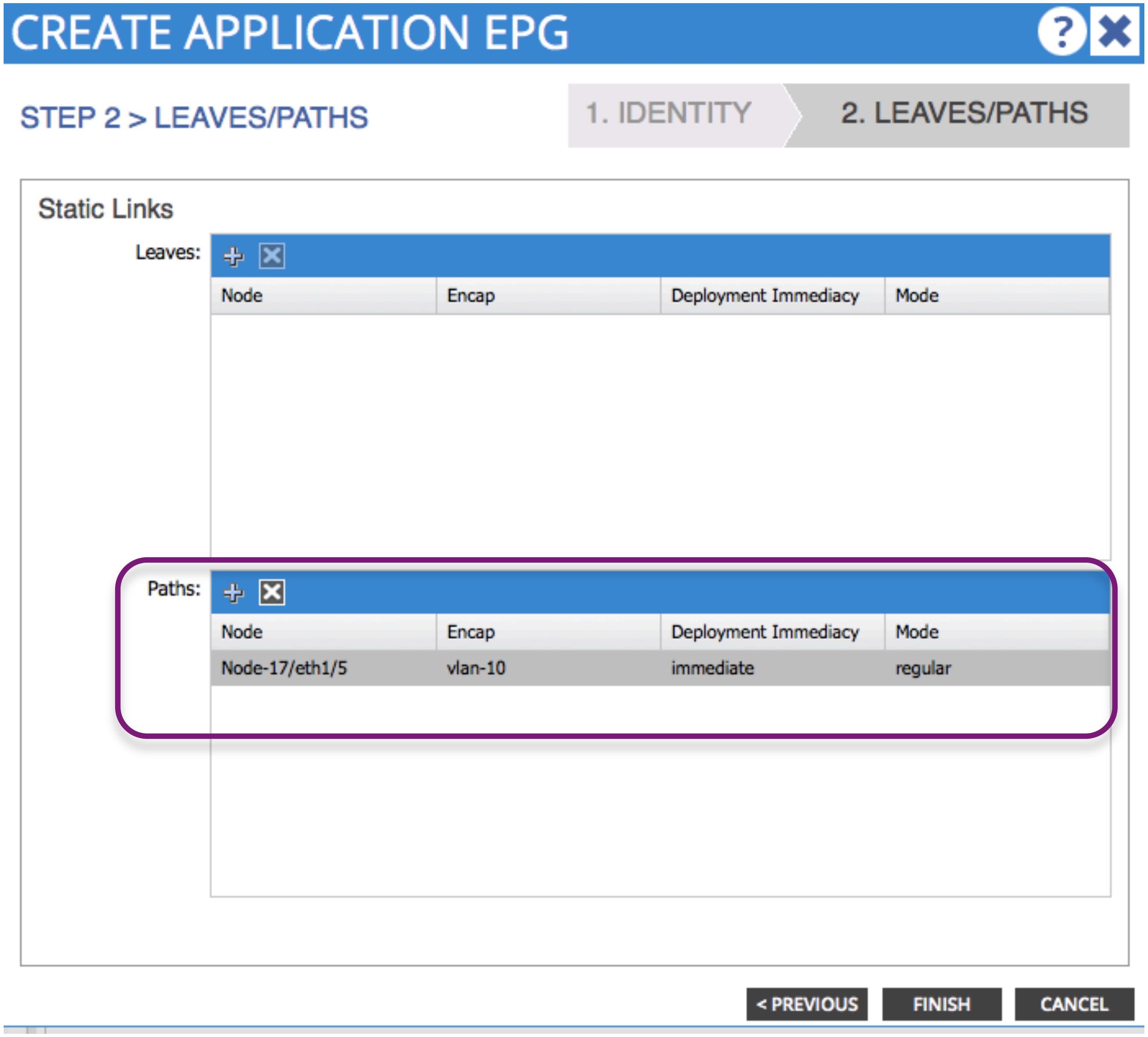
Assign Port to an EPG

- Manually assign a port to a VLAN which in turn mapped to an EPG. This **extend EPG beyond ACI fabric.**
- No contract within EPG
- BPDU is always flooded within EPG.



Assign Port to an EPG

- Traffic received on leaf node 17 interface eth1/5 with VLAN tag 10 will be assigned to the EPG
- Contract associated with the EPG applies in the normal way
- Note there is No contract within EPG.



CREATE APPLICATION EPG

STEP 2 > LEAVES/PATHS

Static Links

Leaves:

Node	Encap	Deployment Immediacy	Mode

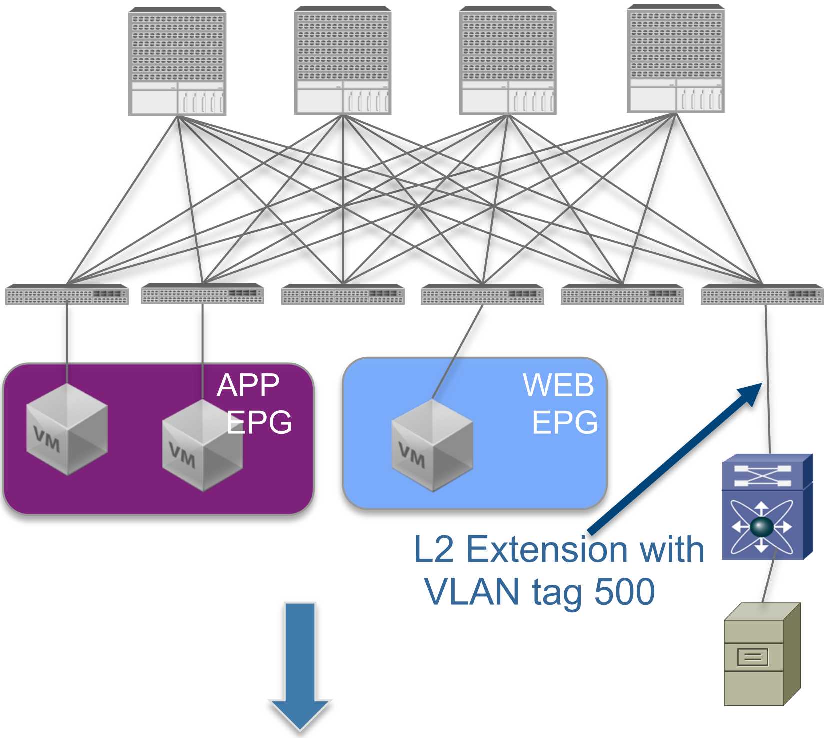
Paths:

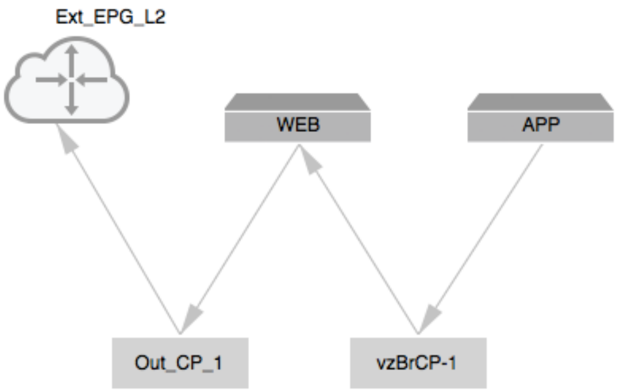
Node	Encap	Deployment Immediacy	Mode
Node-17/eth1/5	vlan-10	immediate	regular

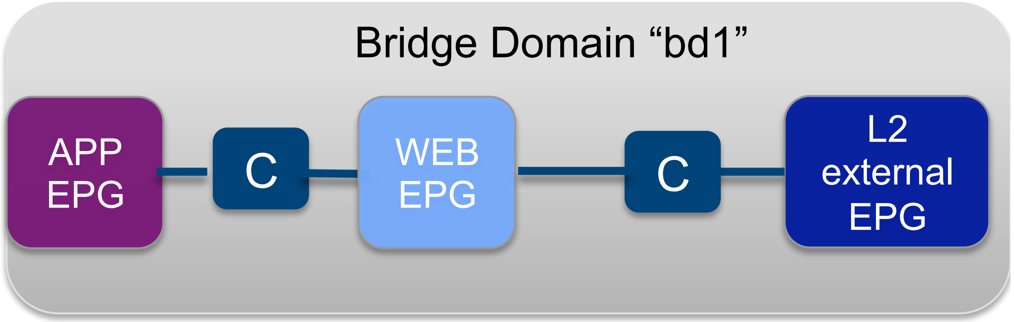
Extend L2 Domain Out of ACI

L2 external connection for a BD

- Extend bridge domain to an external VLAN or VNID
- Packet forwarding between EP in bridge domain “bd1” and external hosts in VLAN 500 is L2 bridge
- One external EPG for each L2 external connection
- Contract can be deployed between L2 external EPG and EPG inside ACI fabric
- APIC GUI Contract view. “Ext_EPG_L2” is the L2 external EPG

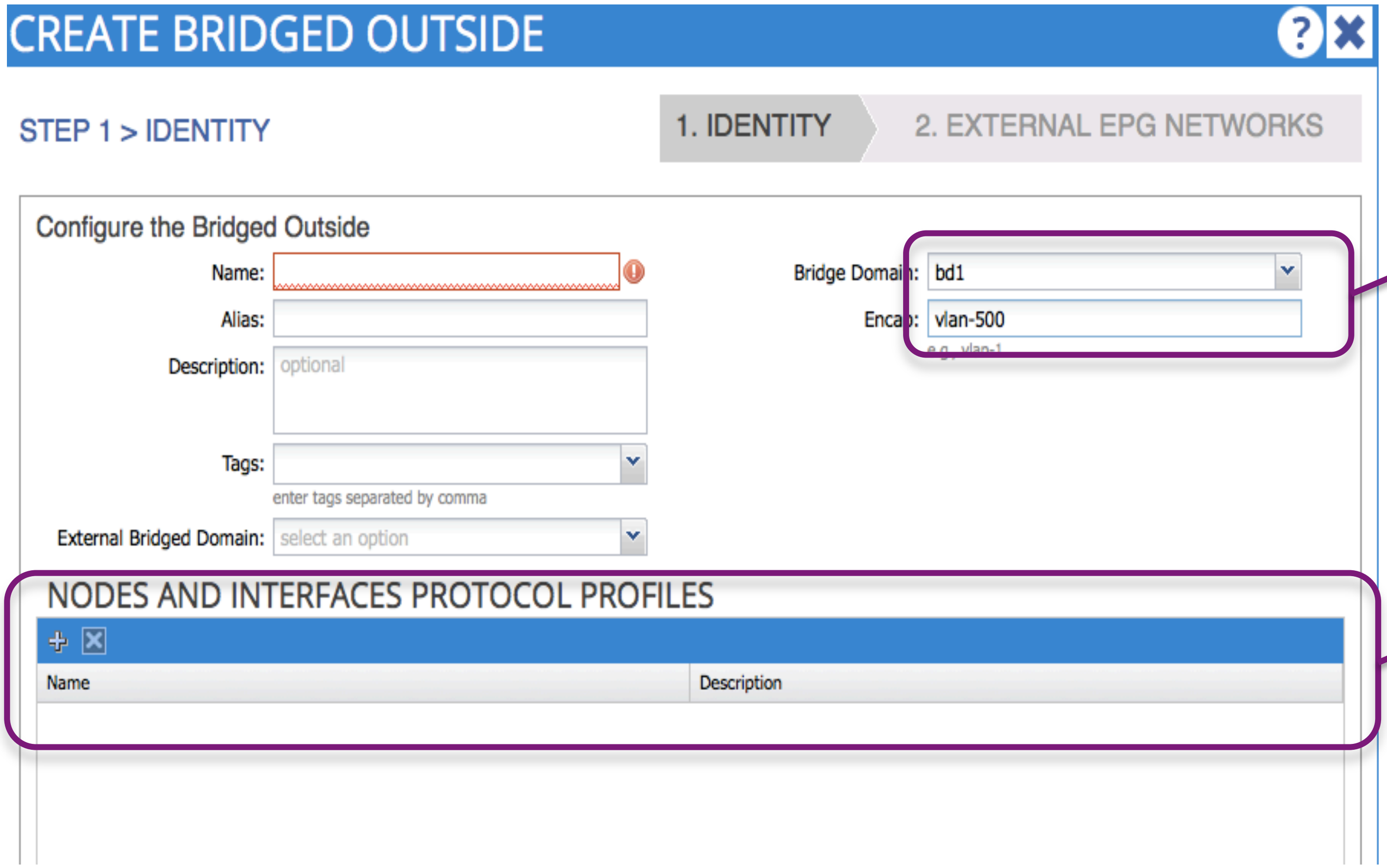






Extend L2 Domain Out of ACI

L2 external connection for a BD



CREATE BRIDGED OUTSIDE

STEP 1 > IDENTITY

Configure the Bridged Outside

Name:

Alias:

Description: optional

Tags:

enter tags separated by comma

External Bridged Domain: select an option

Bridge Domain: bd1

Encap: vlan-500

e.g., vlan-1

NODES AND INTERFACES PROTOCOL PROFILES

Name	Description

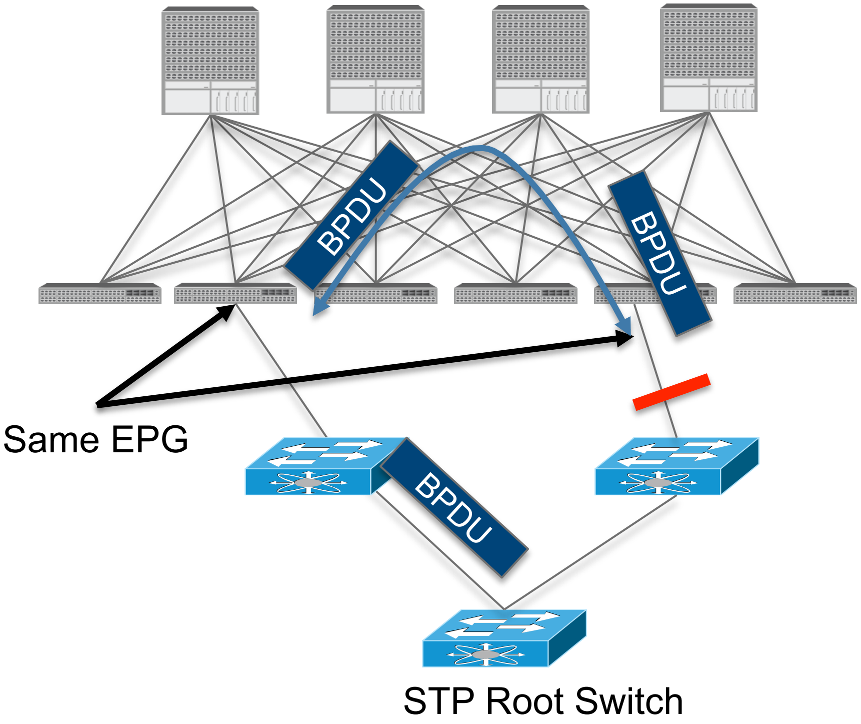
Extend bridge domain “bd1” to outside VLAN ID or VNID

Interface connecting to external L2 network

ACI L2 External Connection

STP Interaction

- No STP running on ACI fabric
- BPDU frame is flooded **within EPG**. No configuration required
- External switches break any potential loop upon receiving the flooded BPDU from ACI fabric



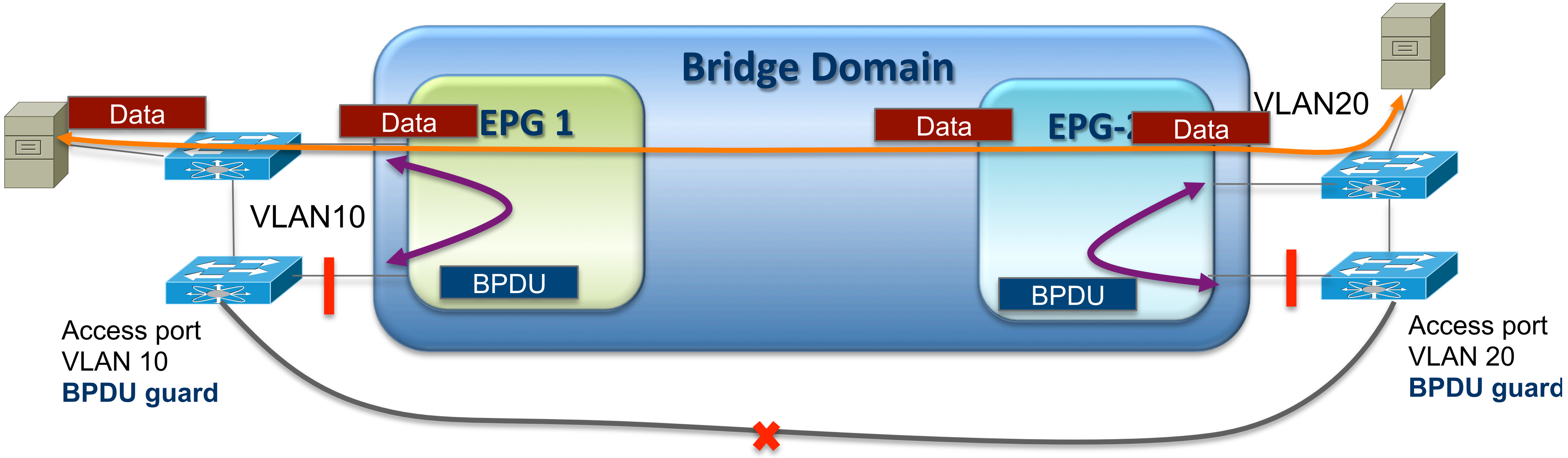
ACI L2 Outside Connection

BPDU Flooding

- **BPDU is always flooded within EPG**. BPDU frame is encapsulated in iVXLAN packet and carries VNID allocated for the EPG

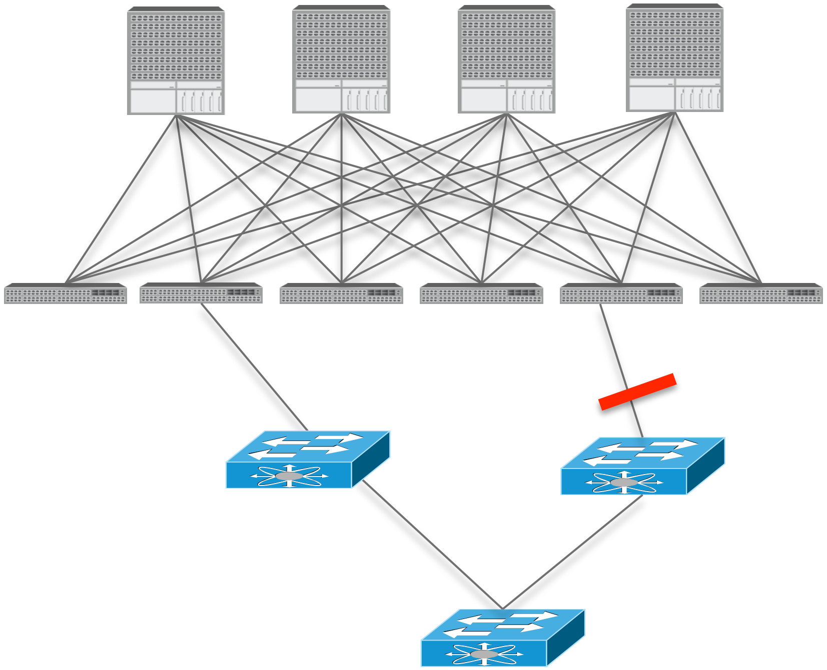
Flags	Flags/ DRE	Source Group	VNID for EPG	M/LB/SP

- Access policy can be created to enable BPDU filter and BPDU guard on selected ports
- Data traffic flooding can be turn on/off on the per **Bridge Domain** level
- Important to turn on BPDU guard on edge ports



ACI L2 External Connection
STP TCN Snooping

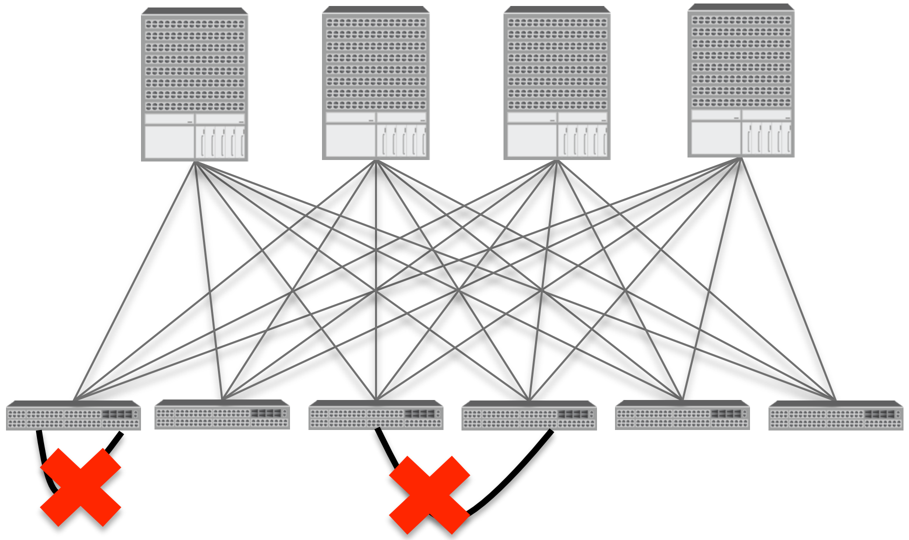
- Fabric intercept the BPDU TCN frame
- APIC flushes the MAC address for the corresponding EPG that has the STP topology change
- Bridge domain flooding vs. Convergence time with TCN.
- With MSTP user need to configure instance to VLAN mapping so APIC knows for what EPGs it need to flush the MAC
- **Recommend to have vPC connection to legacy switches to minimize the TCN**



ACI L2 External Connection

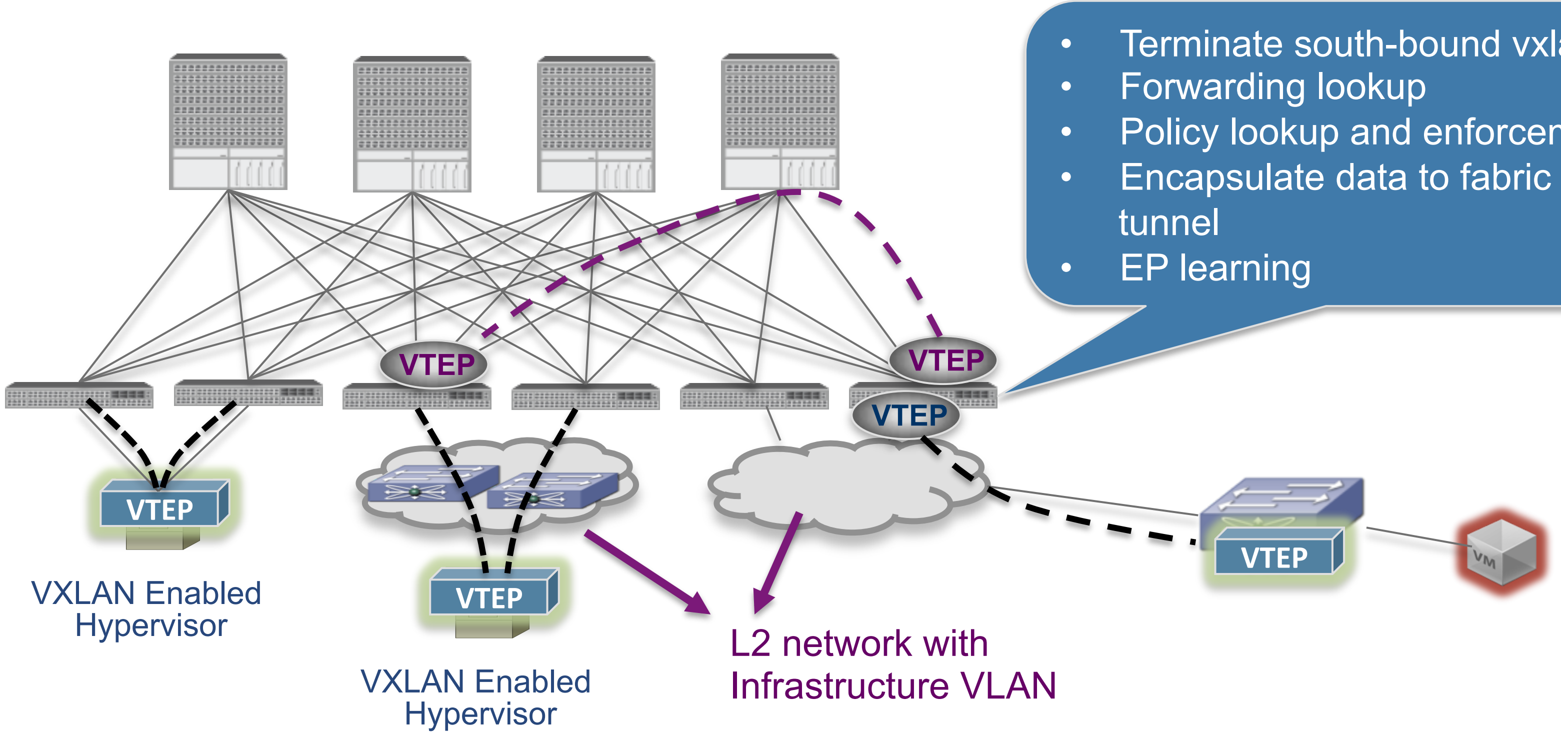
Local Loop Detection

- ACI Fabric doesn't generate STP BPDU frame
- Loop between two leaf switch ports are blocked by cable plant verification
    - Non-fabric ports are not allowed to connect to each other on leaf



ACI L2 External Connection

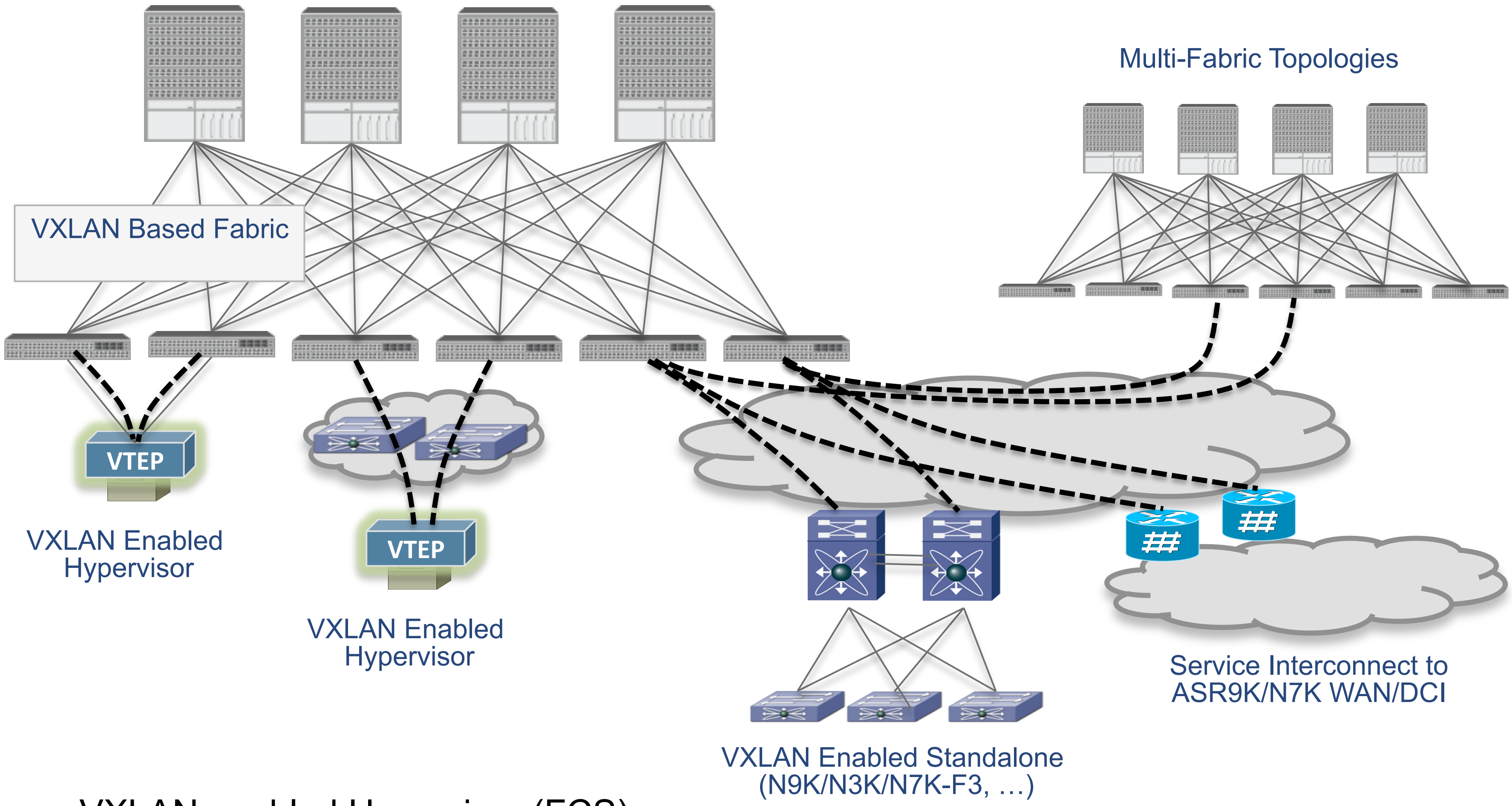
VXLAN Integration at FCS



- Terminate south-bound vxlan tunnel
- Forwarding lookup
- Policy lookup and enforcement
- Encapsulate data to fabric vxlan tunnel
- EP learning

- VXLAN VTEP in ESXi. Integration with vShield Manager. Exchange the VTEP end point, VNID and multicast group with vShield Manager
- Manual configuration for remote VTEP when there is no control plan integration
- **At FCS the remote VTEP need to be L2 adjacent to leaf**. Extend the Infrastructure VRF and VLAN to external VTEP
- Data plan learning for EP behind remote VTEP

ACI L2 External Connection
VXLAN

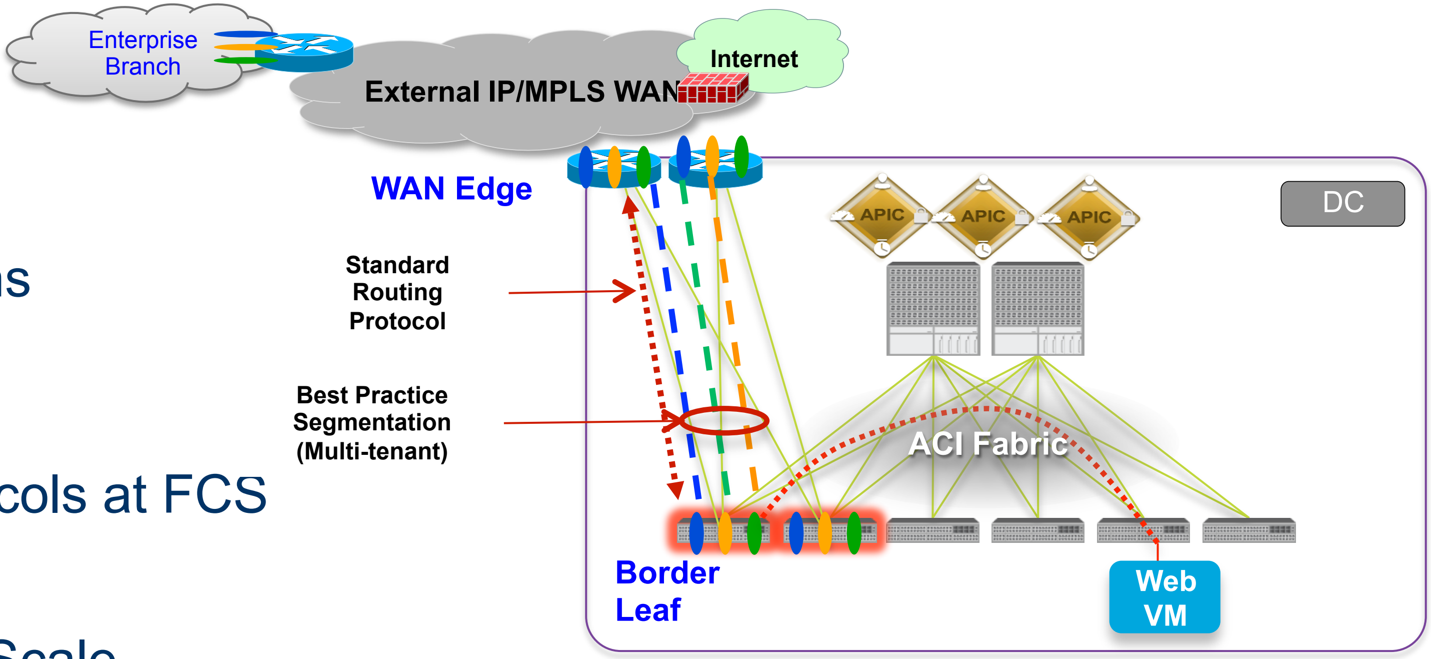


- VXLAN enabled Hypervisor (FCS)
- VXLAN Hardware VTEP (Nexus 9000 standalone, Nexus 3100/7000-F3, ASR9K, …)
- MP-BGP EVPN based control plane for external VTEP connectivity (post FCS)

L3 Outside

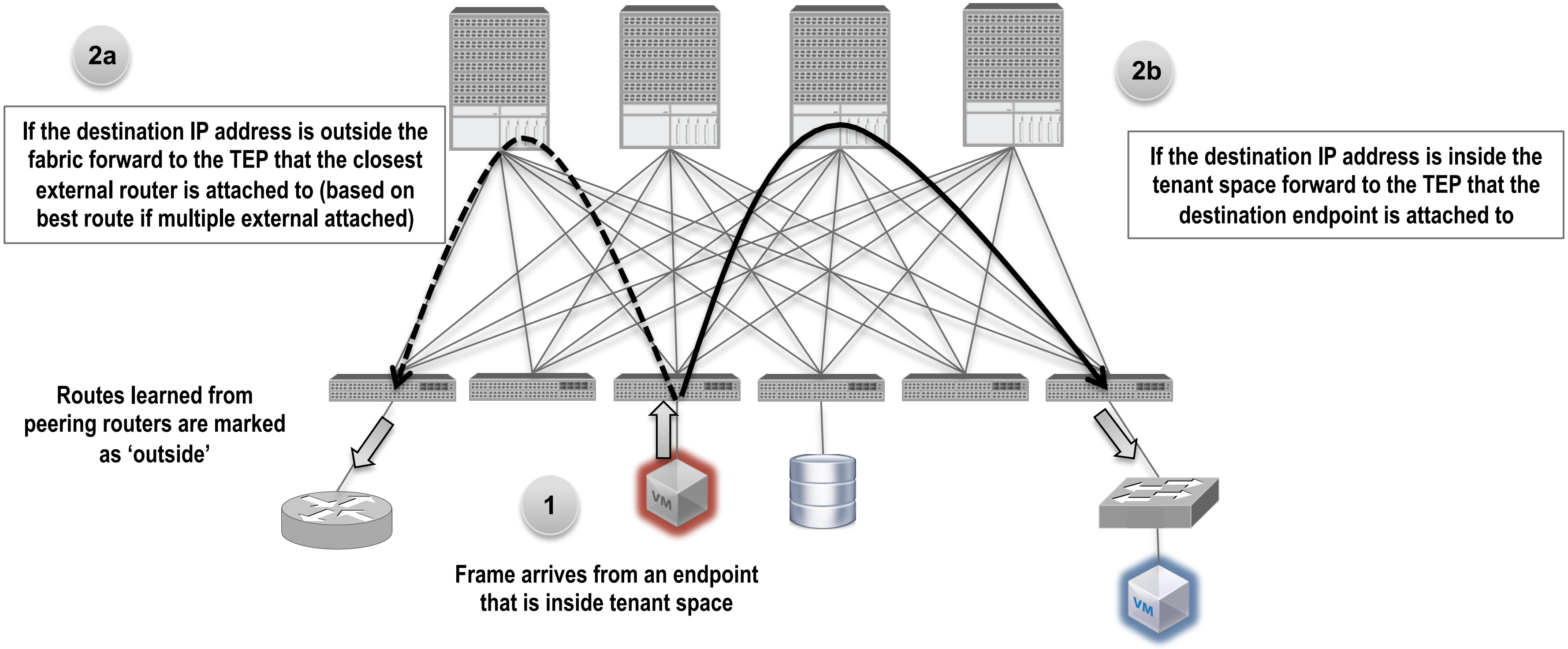
ACI L3 Connection to External Network

- Interface Options
  - L3 port
  - Sub-interface
  - SVI
- Choice of Protocols at FCS
  - IBGP
  - OSPFv2
- L3 Connection Scale
  - 1K VRF per leaf
  - 4K external summary routes(more in the future)
  - 1K LPM entries to derive EPG for external subnets



ACI L3 Packet Forwarding

Important Concepts – Inside and Outside



If the destination IP address is outside the fabric forward to the TEP that the closest external router is attached to (based on best route if multiple external attached)

If the destination IP address is inside the tenant space forward to the TEP that the destination endpoint is attached to

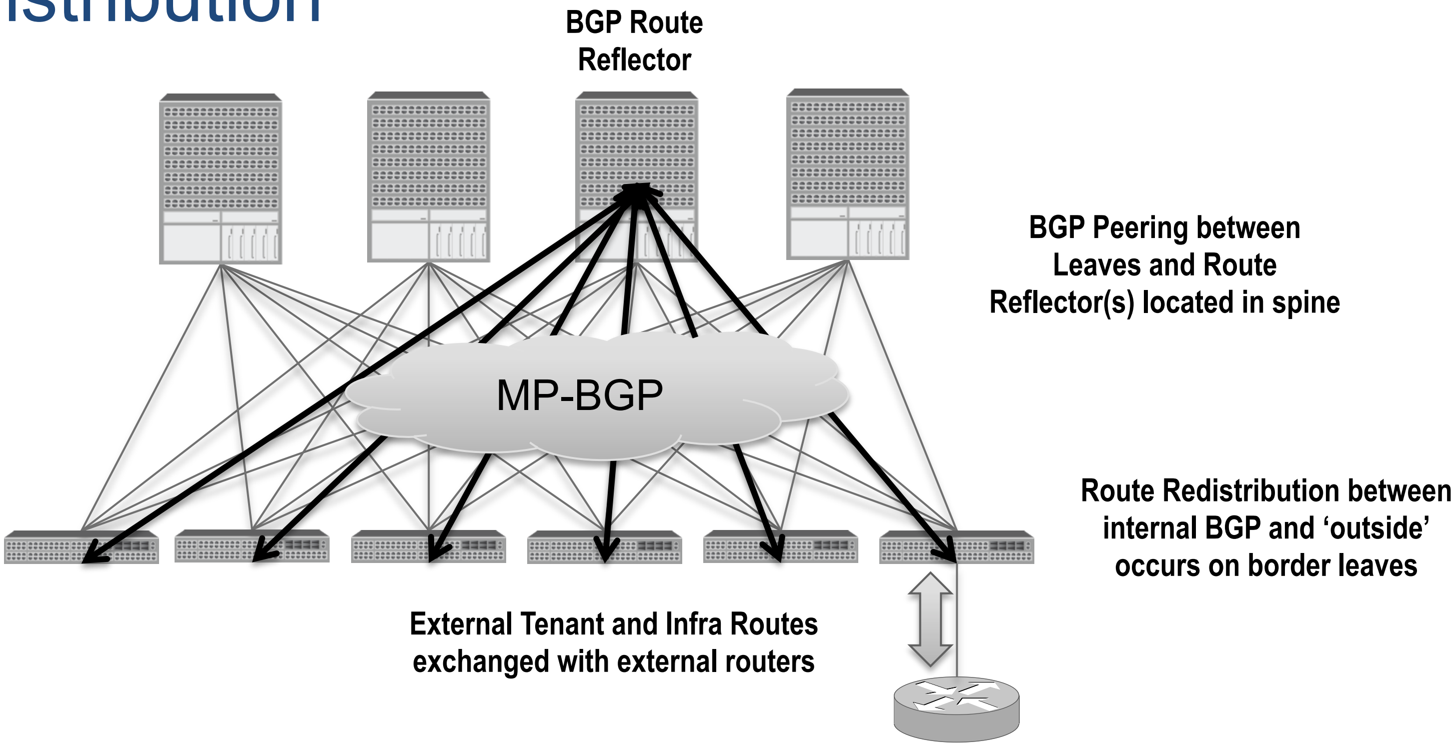
Routes learned from peering routers are marked as ‘outside’

Frame arrives from an endpoint that is inside tenant space

- Single Data Plane with Two Control Planes
- Which ‘forwarding space’ is used to forward a packet is determined by which IP network it is in and where is it going
  - Inside networks are those associated with tenants and their bridge domains (BD’s)
  - Outside networks are those associated with the outside routes for each of those tenants

ACI Layer 3 Connection to External Network

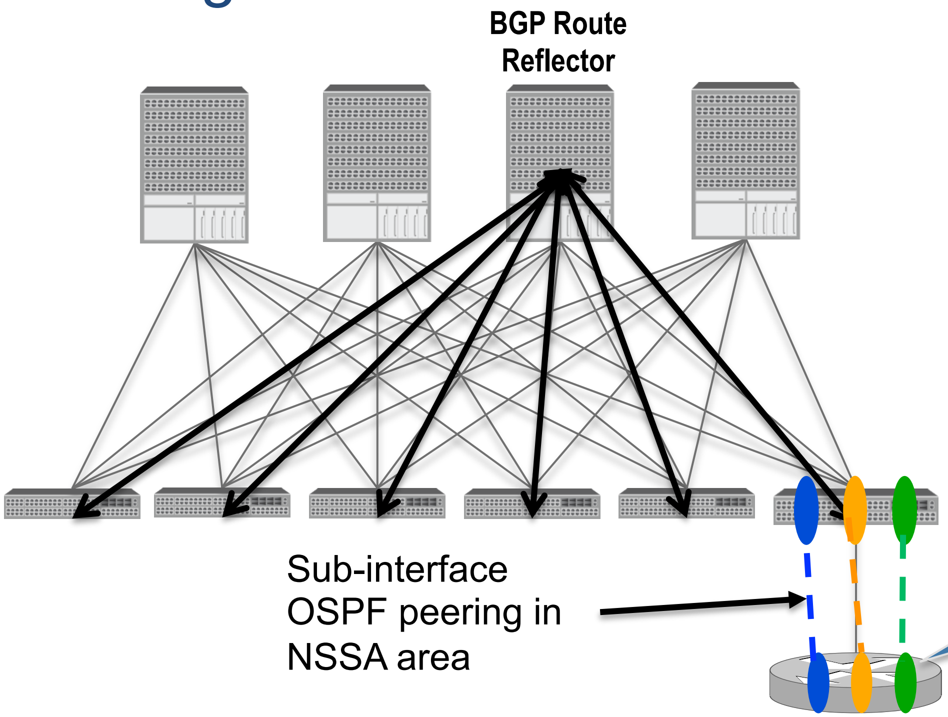
Route Distribution


BGP Route Reflector
BGP Peering between Leaves and Route Reflector(s) located in spine
MP-BGP
Route Redistribution between internal BGP and 'outside' occurs on border leaves
External Tenant and Infra Routes exchanged with external routers

- Fabric leverages MP-BGP for distributing external routes, "outside EPG's" to the leaf switches
- The border leaf switch can peer with external networks and redistribute routing information about external networks into the internal MP-BGP
  - OSPF, Static, iBGP (FCS)
  - MP-BGP w EVPN AF, EIGRP, IS-IS, OSPFv3 (Post FCS)
- Only "Public Subnet"(under Bridge Domain configuration) are announced to external network

ACI Layer 3 Connection to External Network

OSPFv2 Peering Consideration



BGP Route Reflector

Sub-interface OSPF peering in NSSA area

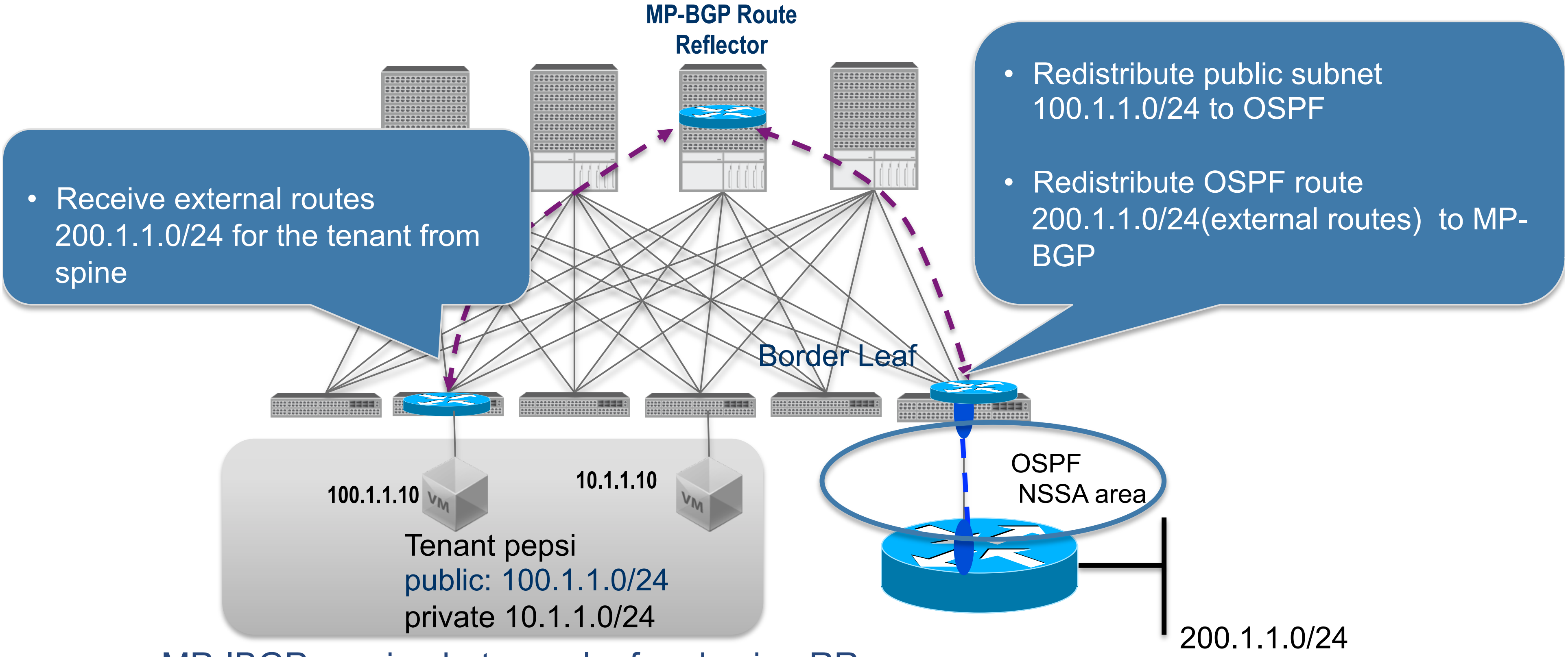
```
interface Ethernet1/1.1000
  encapsulation dot1q 1000
  vrf member Tenant2
  ip address 200.200.200.2/30
  ip ospf network point-to-point
  ip router ospf 1 area 0.0.0.1

router ospf 1
  vrf Tenant2
    area 0.0.0.1 nssa
    default-information originate always
  vrf Tenant3
    area 0.0.0.1 nssa
    default-information
```

- ACI fabric is considered as *stub network* and is not intended to be an transit network
- Must use non-backbone OSPF area and must use NSSA area
- VRF-lite for tenant routes separation. One OSPFv2 adjacency per tenant or use static routes. OSPF or static routes may required for iBGP peer address reachability
- Inside ACI, routes learnt via OSPF is redistributed to BGP and distributed to leaf nodes
- Tenant **public subnet** is redistributed to OSPF NSSA area in border leaf

ACI Layer 3 Connection to External Network

Route Distribution with OSPFv2



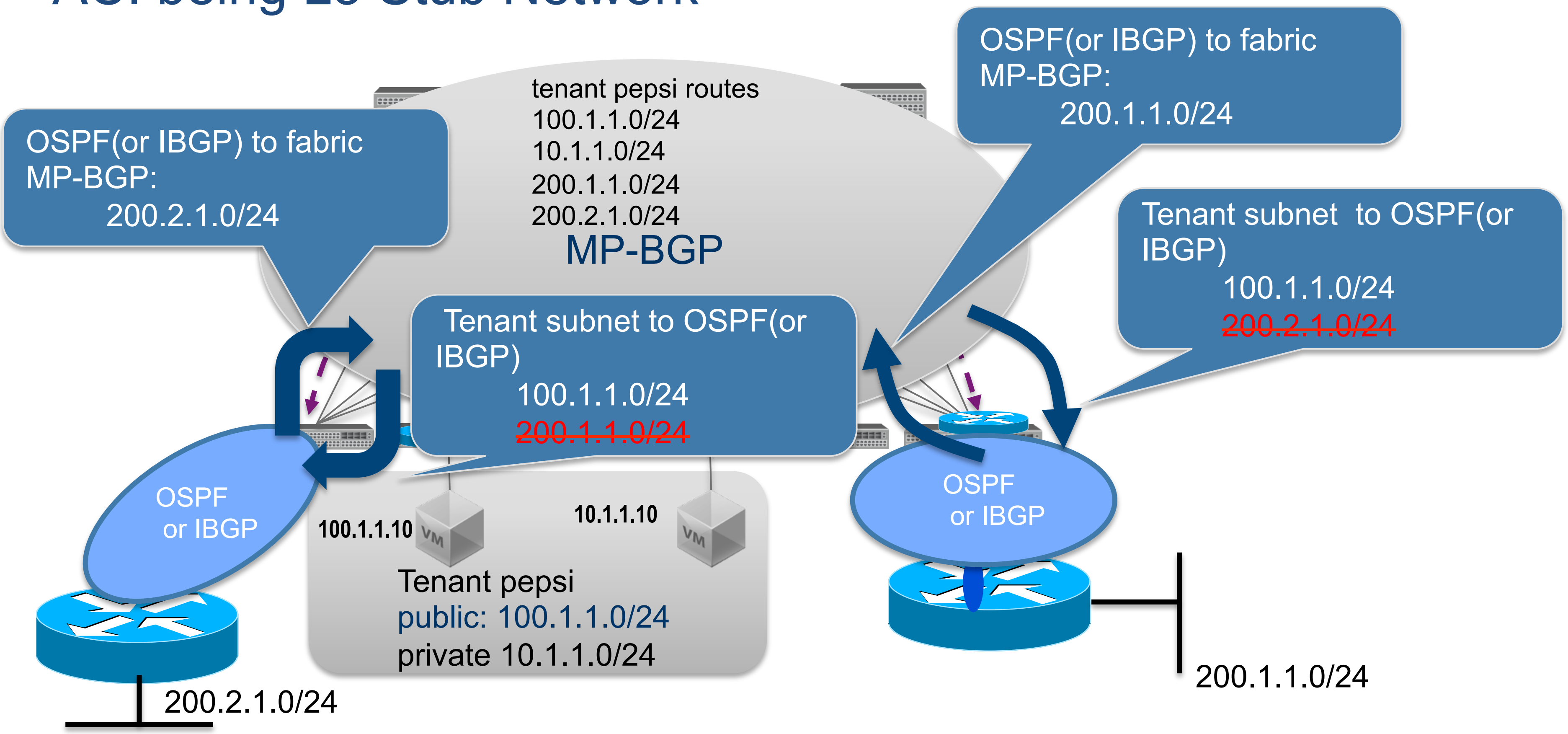
- Receive external routes 200.1.1.0/24 for the tenant from spine

- Redistribute public subnet 100.1.1.0/24 to OSPF
- Redistribute OSPF route 200.1.1.0/24(external routes) to MP-BGP

- MP-IBGP peering between leaf and spine RRs.
- Border leaf redistribute **tenant public subnets** to OSPF NSSA area. When both OSPF and BGP peering are enable on border leaf then the tenant routes will be announce to external router via IBGP only.
- Border leaf redistribute external routes to MP-BGP
- MP-BGP propagate external routes to all leafs where the VRF is instantiated.

ACI Layer 3 Connection to External Network

ACI being L3 Stub Network



OSPF(or IBGP) to fabric
MP-BGP:
200.2.1.0/24

tenant pepsi routes
100.1.1.0/24
10.1.1.0/24
200.1.1.0/24
200.2.1.0/24
MP-BGP

OSPF(or IBGP) to fabric
MP-BGP:
200.1.1.0/24

Tenant subnet to OSPF(or IBGP)
100.1.1.0/24
~~200.2.1.0/24~~

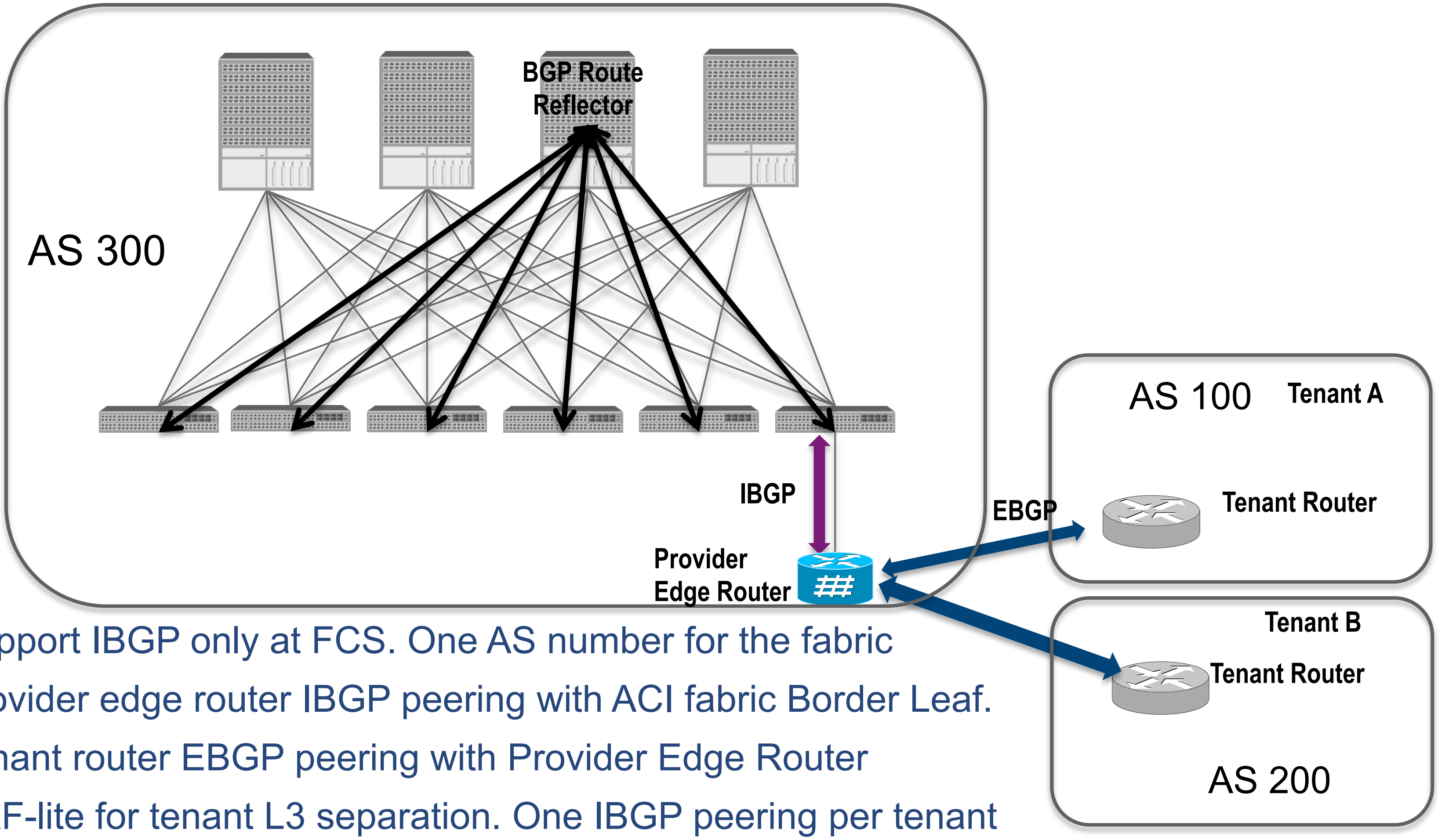
Tenant subnet to OSPF(or IBGP)
100.1.1.0/24
~~200.1.1.0/24~~

Tenant pepsi
public: 100.1.1.0/24
private 10.1.1.0/24

- ACI is not designed to be used as transit node or carrying transit traffic.
  - Routing table scale, full protocol policy and automation
- Border leaf only announce **tenant public subnet within ACI fabric** to external routes. **Border leaf DOES NOT announce transit routes to external routers.**
- Route redistribute policy is created automatically by APIC
- Additional development required to support policy enforcement for transit traffic

ACI Layer 3 Outside

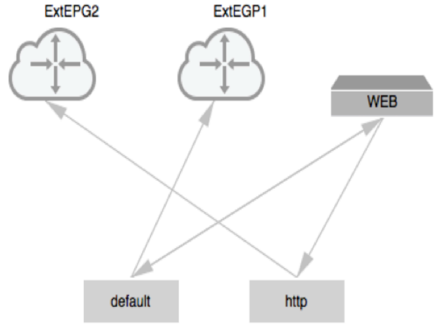
IBGP Peering Consideration

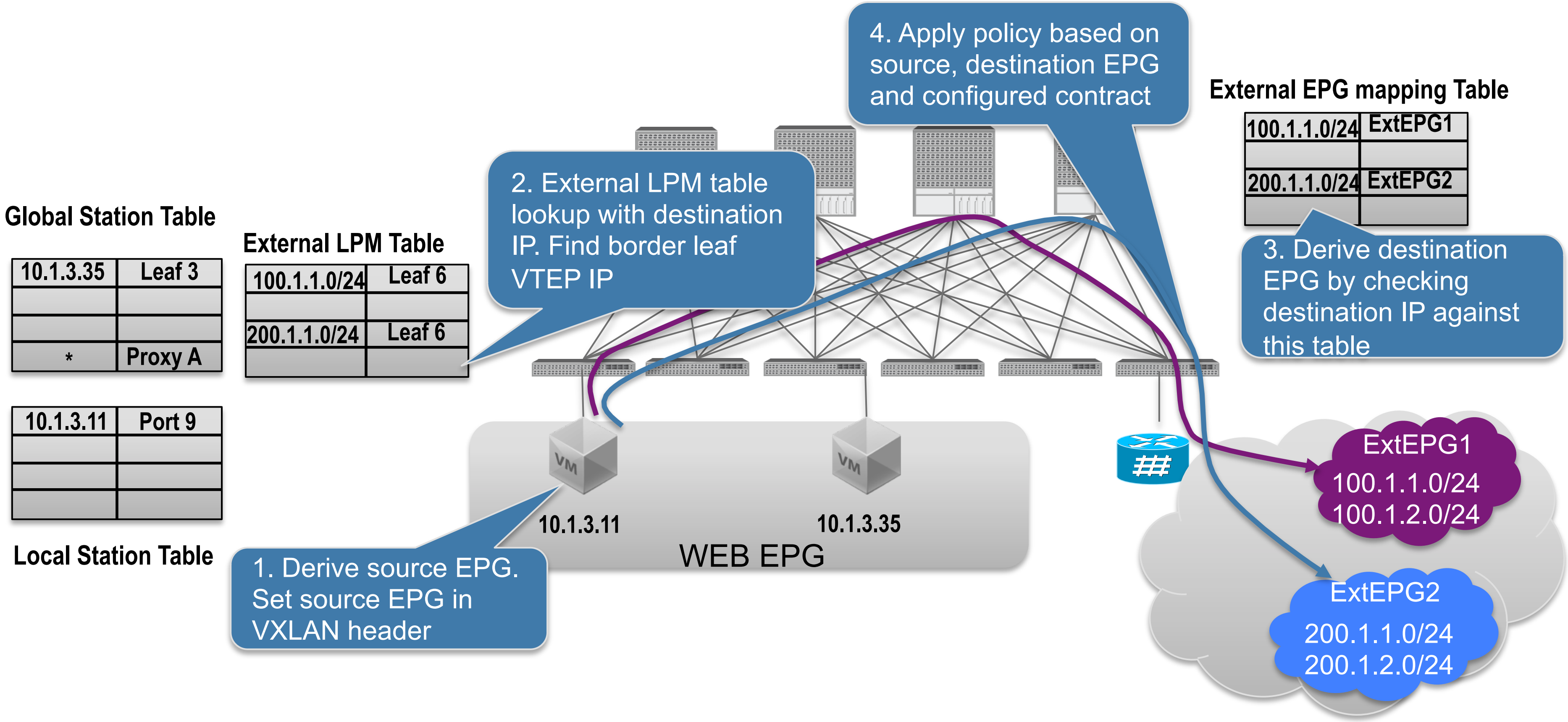


- Support IBGP only at FCS. One AS number for the fabric
- Provider edge router IBGP peering with ACI fabric Border Leaf.
- Tenant router EBGP peering with Provider Edge Router
- VRF-lite for tenant L3 separation. One IBGP peering per tenant
- Provider Edge Router is needed for
    - Large routing table
    - More VRF support
    - WAN features and more sophisticated BGP policy

Internal EPG to External EPG
Forwarding and Policy Lookup

- For L3 outside connection, external EPG is derived from subnet
- Support multiple external EPGs. External EPG1 could be remote branch or another DC. External EPG2 could be Internet
- Different policy for different external EPGs





1. Derive source EPG. Set source EPG in VXLAN header

2. External LPM table lookup with destination IP. Find border leaf VTEP IP

3. Derive destination EPG by checking destination IP against this table

4. Apply policy based on source, destination EPG and configured contract

Global Station Table

10.1.3.35	Leaf 3
*	Proxy A

Local Station Table

10.1.3.11	Port 9

External LPM Table

100.1.1.0/24	Leaf 6
200.1.1.0/24	Leaf 6

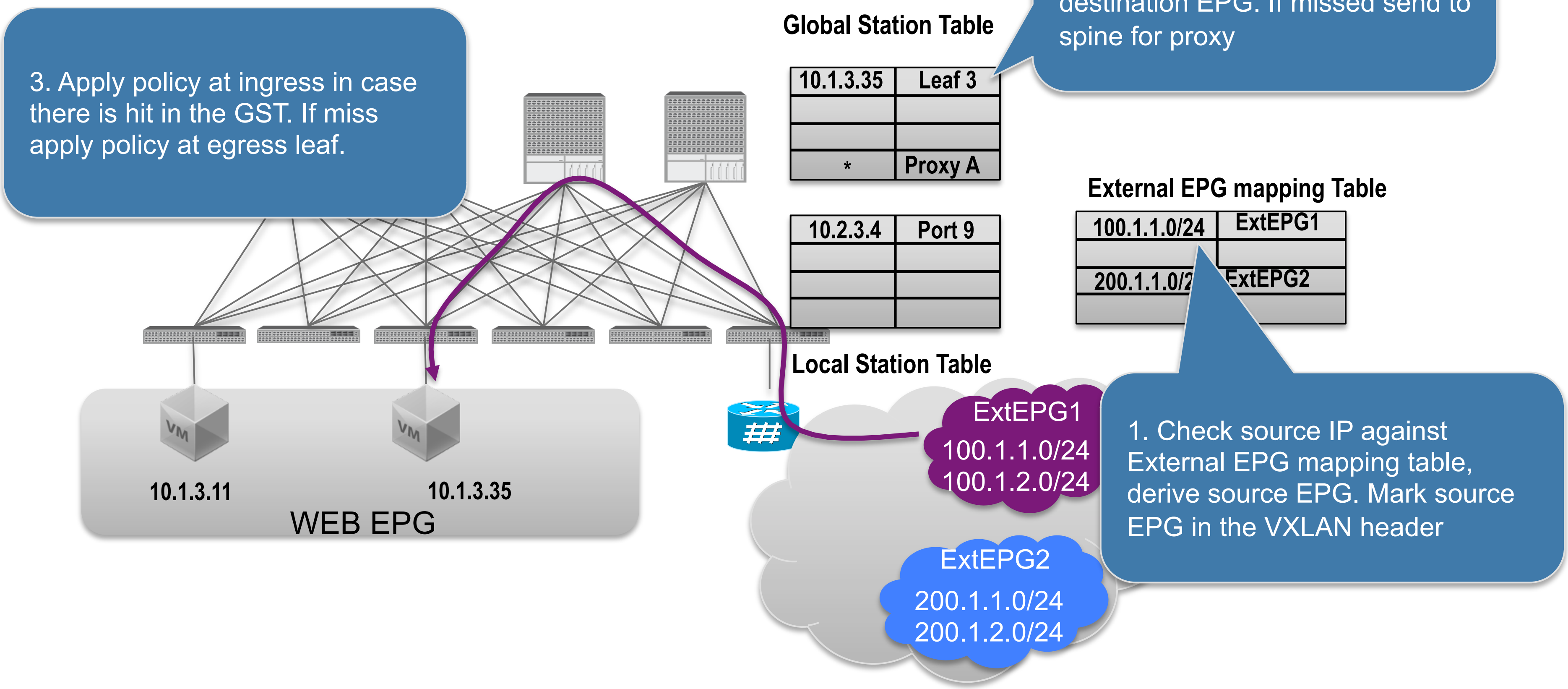
External EPG mapping Table

100.1.1.0/24	ExtEPG1
200.1.1.0/24	ExtEPG2

External EPG to Internal EPG

Forwarding and Policy Lookup

- For L3 outside connection, external EPG is derived from subnet
- Support multiple external EPGs. External EPG1 could be remote branch or another DC. External EPG2 could be Internet
- Different policy for different external EPGs



3. Apply policy at ingress in case there is hit in the GST. If miss apply policy at egress leaf.

2. Check destination IP against GST. Send to egress leaf. Derive destination EPG. If missed send to spine for proxy

1. Check source IP against External EPG mapping table, derive source EPG. Mark source EPG in the VXLAN header

Global Station Table

10.1.3.35	Leaf 3
*	Proxy A

Local Station Table

10.2.3.4	Port 9

External EPG mapping Table

100.1.1.0/24	ExtEPG1
200.1.1.0/2	ExtEPG2

ACI L3 Outside

Scaling

Global Station Table

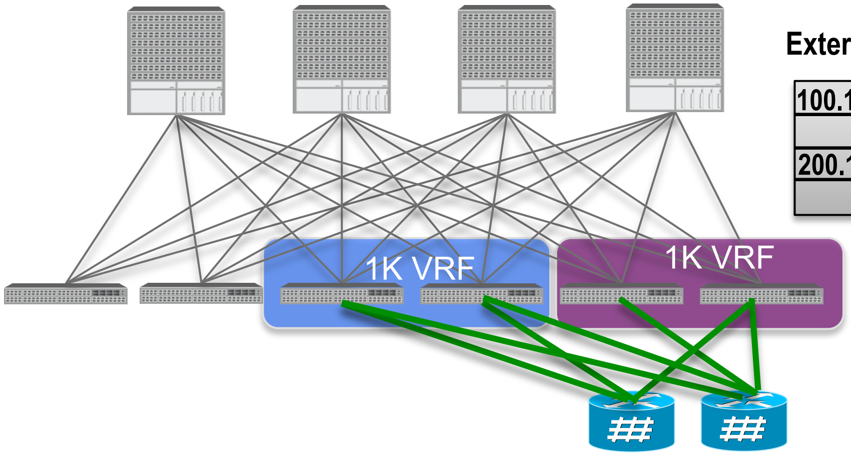
10.1.3.35	Leaf 3
*	Proxy A

10.1.3.11	Port 9

Local Station Table

External LPM Table

100.1.1.0/24	Leaf 6
200.1.1.0/24	Leaf 6



External EPG mapping Table

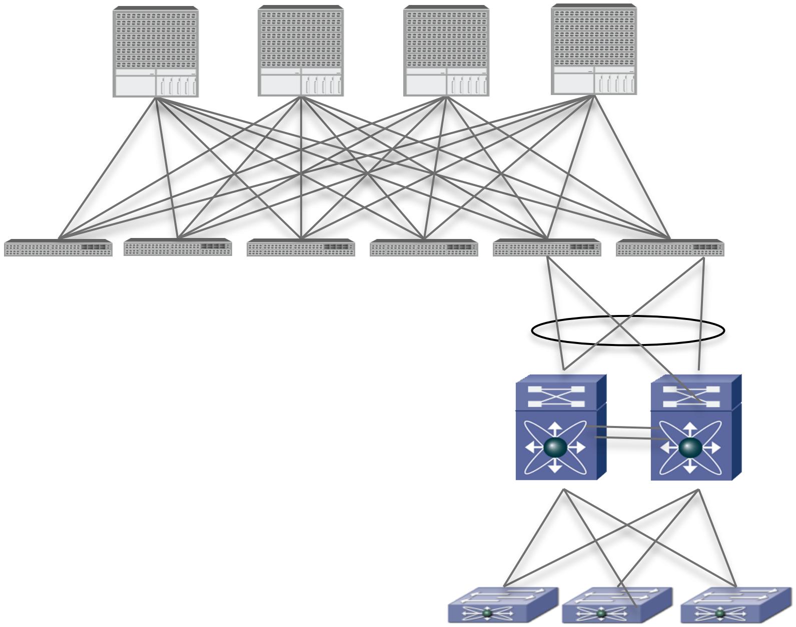
100.1.1.0/24	ExtEPG1
200.1.1.0/24	ExtEPG2

- External LPM has the external routes. 4K at FCS. HW support more(leverage T2 LPM table)
- 1K VRF supported per leaf. Scale the border leaf horizontally for more VRFs in ACI fabric.
- External EPG mapping table 1K entries.
  - IP prefix based EPG
  - Prefix and mask can be different than the external LPM table
  - Support multiple external EPG to have different policies with external devices
- Routing protocol peering scaling pending testing

ACI L3 Outside

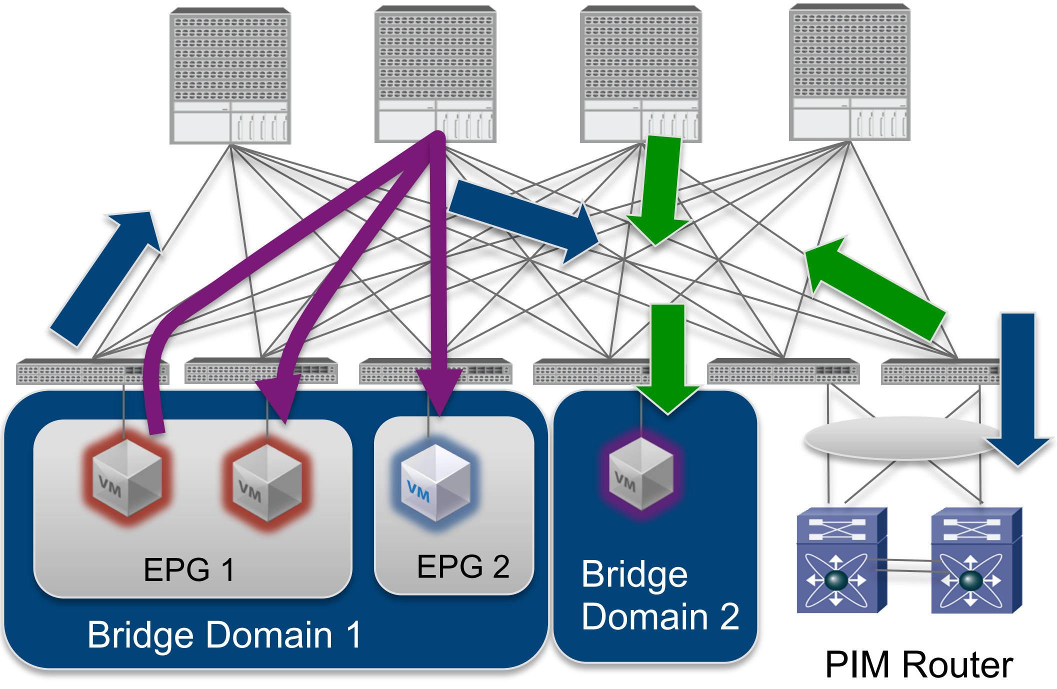
SVI Connection

- L3 connection to outside network with
  - L3 port
  - Sub-interface for multi-tenant
  - SVI
- SVI is needed when the same set of interfaces are used for L2 and L3 connection to outside network



ACI L3 Outside Connection
IP Multicast Traffic

- ACI supports IGMP snooping and L2 bridging for IP multicast traffic
- L2 multicast bridging within Bridge Domain based on IGMP snooping entries.
- Need external PIM router for L3 routing across Bridge Domain boundary
- L2 outside connection to the external PIM router for source and receiver bridge domain



Thank you.