



You make **possible**



Introduction to VXLAN

The future path of your datacenter

Rahul Parameswaran, Technical Marketing Engineer

BRKDCN-1645

CISCO *Live!*

Barcelona | January 27-31, 2020



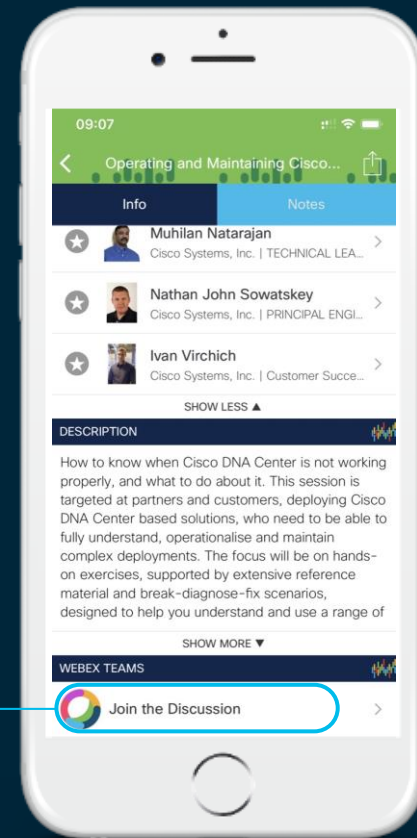
Cisco Webex Teams

Questions?

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

How

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



A few things..

- Prerequisites
 - Good understanding of Unicast Routing Protocols – OSPF/ISIS
 - Knowledge of Multi protocol BGP (MP-BGP)
 - Basics of Multicast forwarding and PIM
- Use Cisco WebEx Teams for Questions
- Watch out for the hidden slides 😊

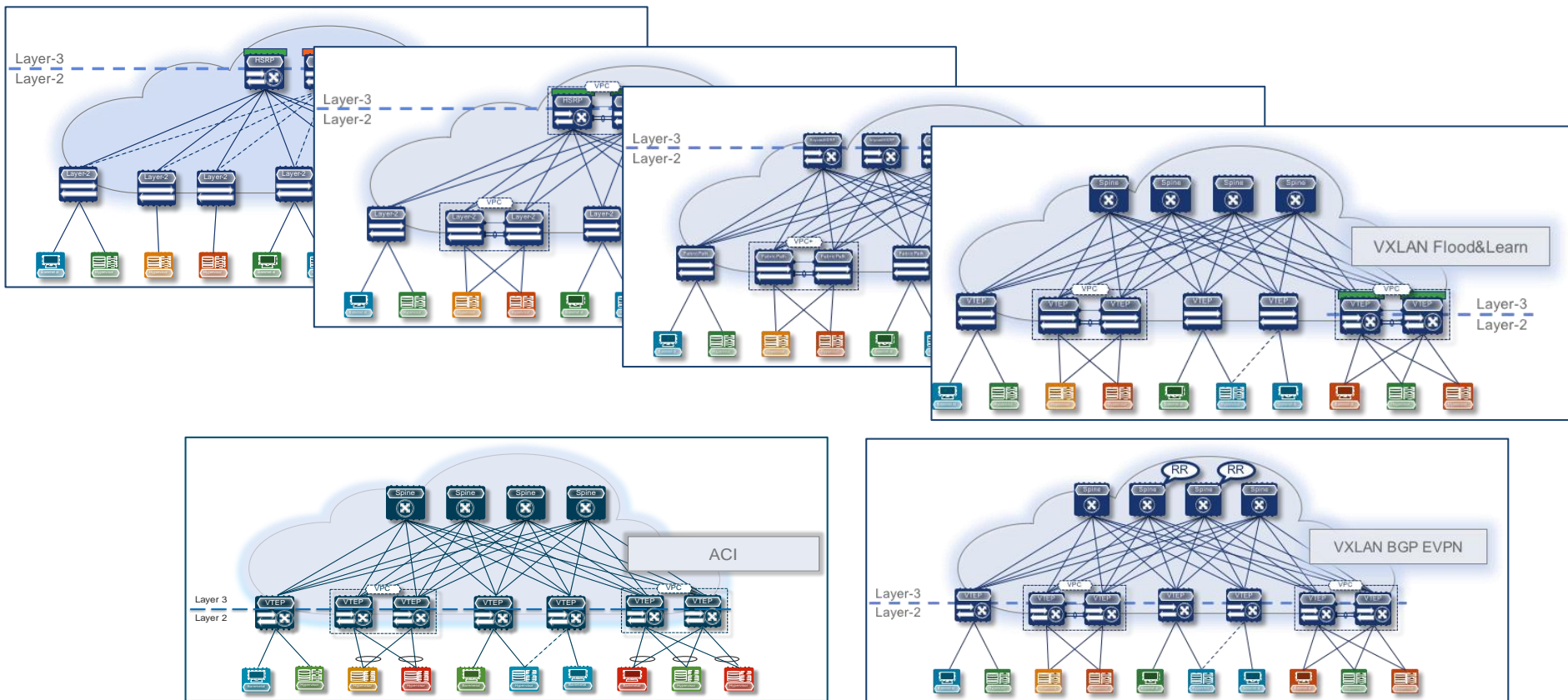
Session Objective

- A short overview on Data Center Evolution
- Introduction to Overlays and VXLAN
- Understanding how MP-BGP is used as a control plane
- Packet Walk with VXLAN
- Design options and additional use cases

Agenda

- Data Center evolution
- Overlay Taxonomy
- VXLAN with MP-BGP EVPN Control Plane
- Packet Walk
- VXLAN Design Options
- Use cases

Data Center “Fabric” Journey



Why VXLAN Overlay

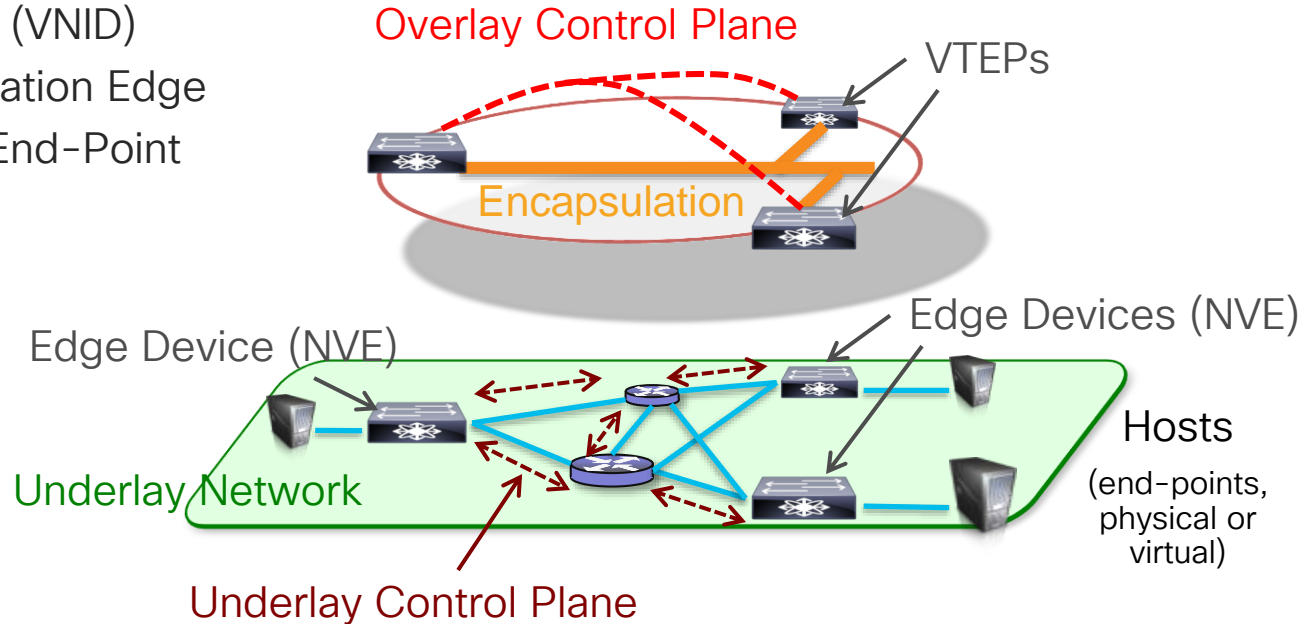
Customer Needs	VXLAN Delivered
Any workload anywhere – VLANs limited by L3 boundaries	Any Workload anywhere- across Layer 3 boundaries
VM Mobility	Seamless VM Mobility
Scale above 4k Segments (VLAN limitation)	Scale up to 16M segments
Efficient use of bandwidth	Leverages ECMP for optimal path usage over the transport network
Secure Multi-tenancy	Traffic & Address Isolation

Overlay Taxonomy

Identifier = VN Identifier (VNID)

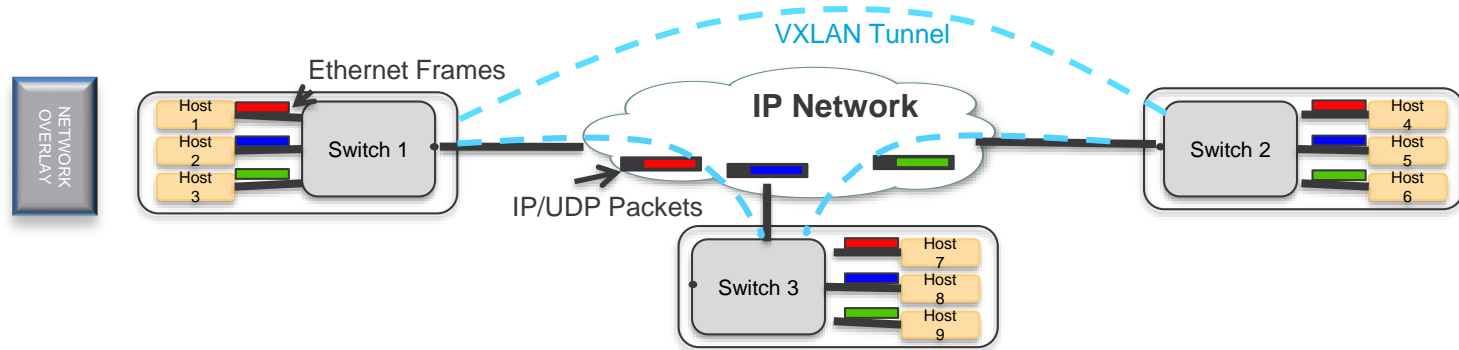
NVE = Network Virtualisation Edge

VTEP = VXLAN Tunnel End-Point

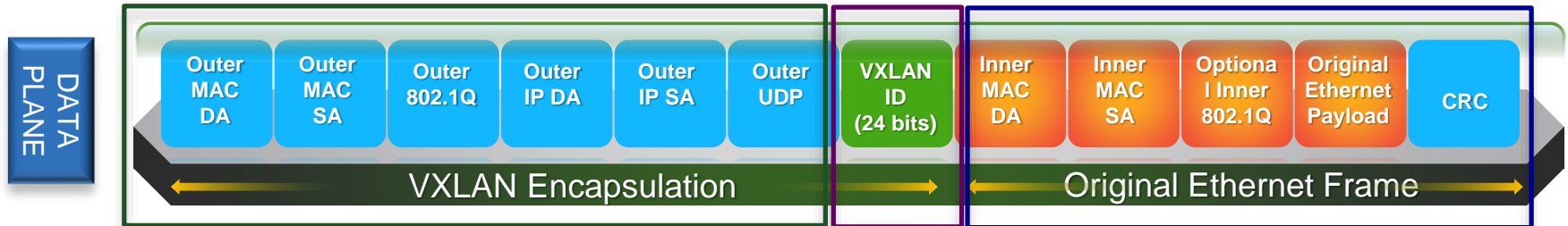


VXLAN Packet

- VXLAN is point to multi-point tunneling mechanism to extend Layer 2 networks over an IP network



- VXLAN uses MAC in UDP encapsulation (UDP destination port 4789)

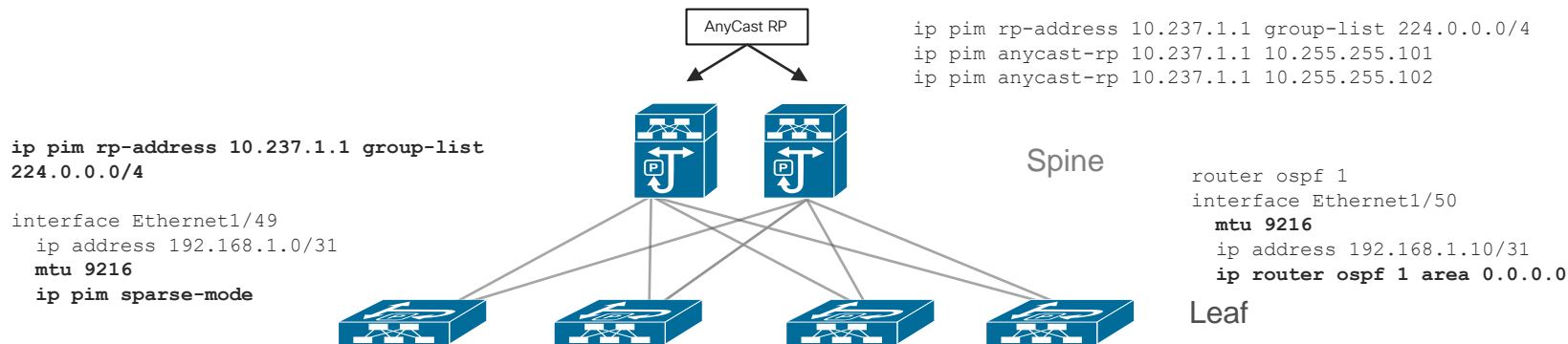


Lets Build a VXLAN Fabric

VXLAN Fabric – Creating the underlay network

IP routed Network

- Flexible topologies
- Recommend a network with redundant paths using ECMP for load sharing
- Support any routing protocols --- OSPF, IS-IS, BGP, etc.
- All proven best practices for IP routing network apply



Two Modes of VXLAN

Flood-and-Learn VXLAN:

- No control plane
- Data driven flood and learning
→ Ethernet in the overlay network

VXLAN EVPN:

- EVPN as control plane
- VTEPs exchange L2/L3 host and subnet reachability through EVPN control plane
→ Routing protocol for both L2 and L3 forwarding

- Limited scale
- Limited workload mobility
- Centralized Gateway
- Security Risk

- Increased scale and stability
- Optimized workload mobility
- Distributed Anycast Gateway
- Increased Security

VXLAN BUM Traffic Handling

- BUM Traffic --- Multi-destination traffic
 - Broadcast
 - Unknown Layer-2 Unicast
 - Multicast

BUM Traffic transport mechanisms

- Multicast replication
 - Requests the underlay network to run IP multicast
- Ingress unicast replication
 - One unicast replica per remote VTEP
 - Increase traffic load throughout the network

VXLAN with BGP EVPN Control Plane

EVPN Primer --- MP-BGP Review

Virtual Routing and Forwarding (VRF)

Layer-3 segmentation for tenants' routing space

Route Distinguisher (RD):

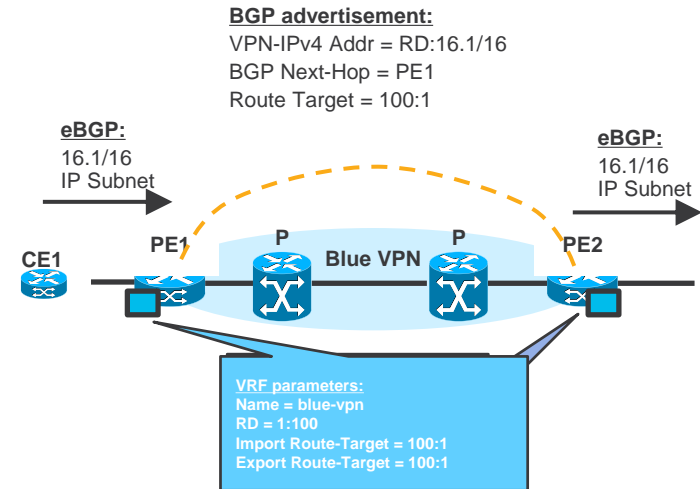
8-byte field, VRF parameters; unique value to make VPN IP routes unique: RD + VPN IP prefix

Selective distribute VPN routes:

Route Target (RT): 8-byte field, VRF parameter, unique value to define the import/export rules for VPNv4 routes

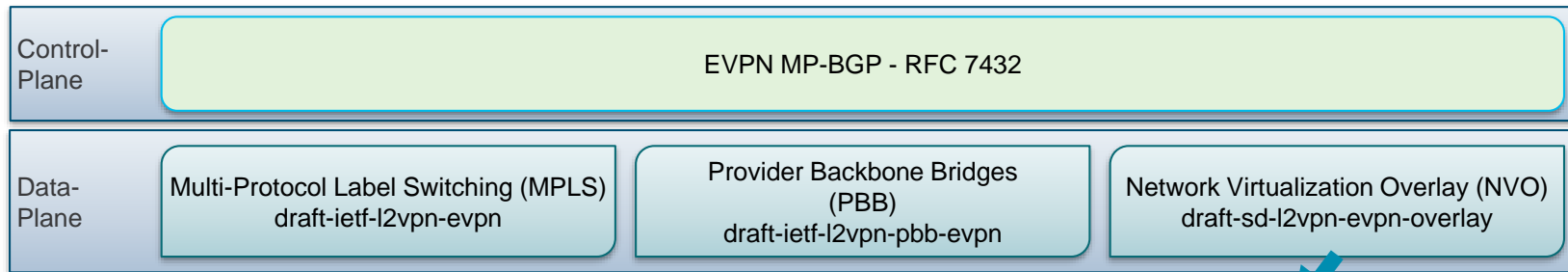
VPN Address-Family:

Distribute the MP-BGP VPN routes



What is VXLAN/EVPN?

- Standards based Overlay (VXLAN) with Standards based Control-Plane (BGP)
- Layer-2 MAC and Layer-3 IP information distribution by Control-Plane (BGP)
- Forwarding decision based on Control-Plane (minimizes flooding)
- Integrated Routing/Bridging (IRB) for Optimized Forwarding in the Overlay

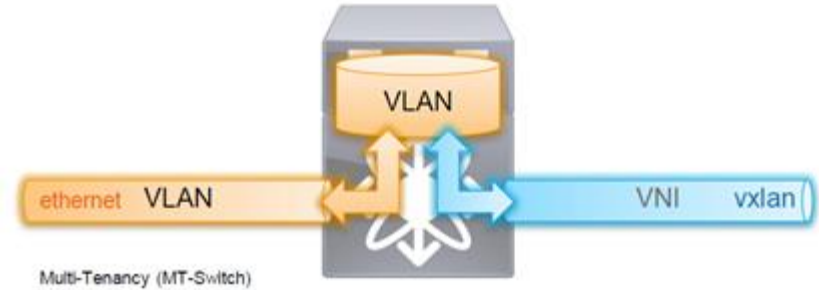


- EVPN over NVO Tunnels (VXLAN, NVGRE, MPLSoE) for Data Center Fabric encapsulations
- Provides Layer-2 and Layer-3 Overlays over simple IP Networks

Layer 2 Multi-tenancy

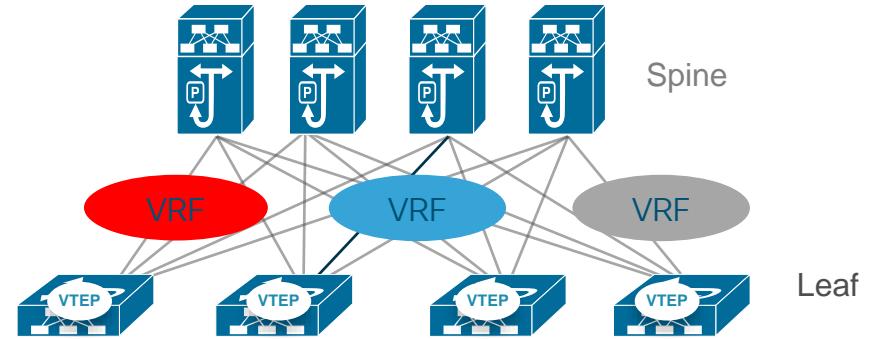
Switch level multi-tenancy

- VLAN to Segment ID mapping (4K vlans per switch)
- With VLAN we can achieve per port significance



Layer 3 Multi-tenancy

Tenants or VRF for L3 logical separation



EVPN based VXLAN Fabric



EVPN Route Reflector

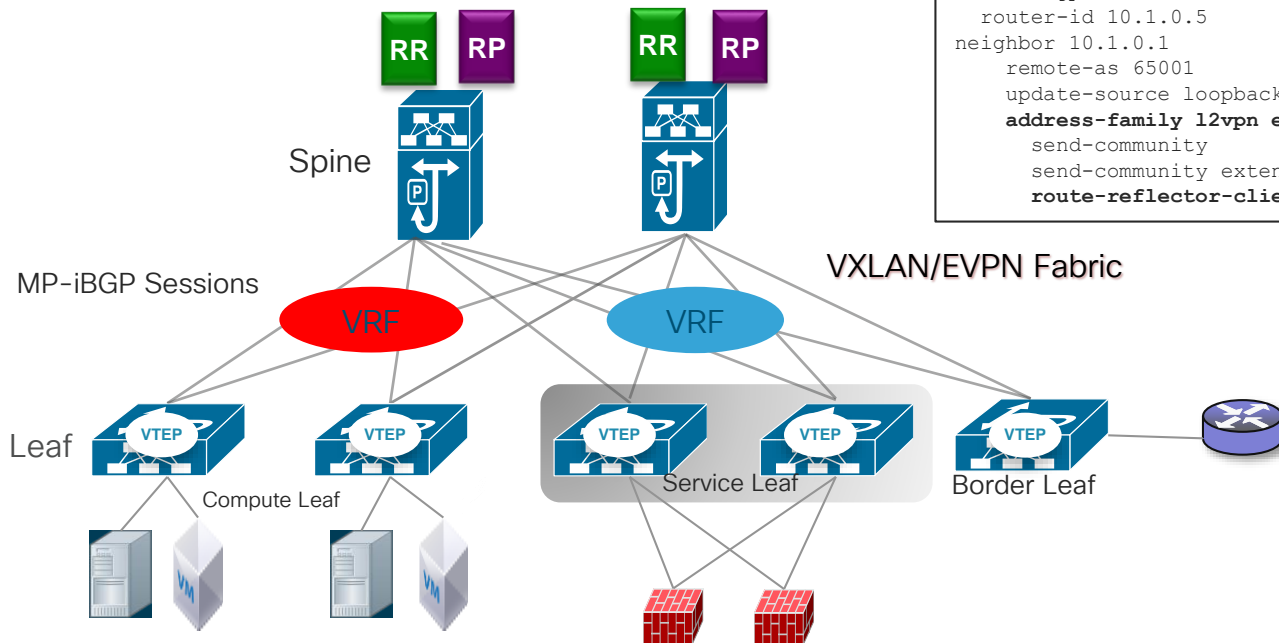


Rendezvous Point (Underlay)

```
! leaf bgp config
router bgp 65001
  router-id 10.1.0.4
  neighbor 10.1.0.5
    remote-as 65001
  update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
  vrf VRF-RED
    address-family ipv4 unicast
    advertise l2vpn evpn
  address-family ipv6 unicast
  advertise l2vpn evpn
  vrf VRF-BLUE
    address-family ipv4 unicast
    advertise l2vpn evpn
  address-family ipv6 unicast
  advertise l2vpn evpn
```

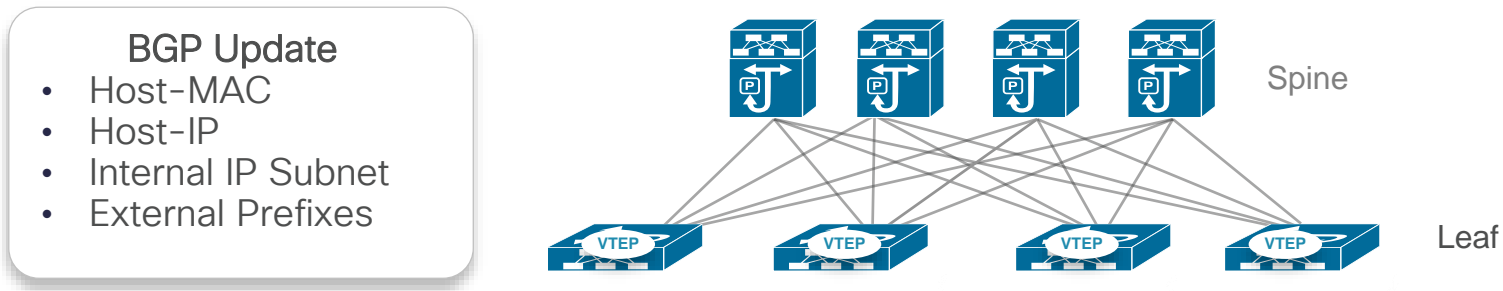
```
! spine bgp config
```

```
router bgp 65001
  router-id 10.1.0.5
  neighbor 10.1.0.1
    remote-as 65001
  update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
  route-reflector-client
```



EVPN Control Plane – Reachability Distribution

- EVPN Control Plane -- Host and Subnet Route Distribution



- Use MP-BGP with EVPN Address Family on leaf nodes to distribute internal host MAC/IP addresses, subnet routes and external reachability information
- MP-BGP enhancements to carry up to 100s of thousands of routes with reduced convergence time

Configuration Snippet

```
Vlan 10
  vn-segment 5010
Vlan 20
  vn-segment 5020
```

Layer 2 VNI

```
Vlan 1000
!Layer 3 VNI
  vn-segment 9999
Vlan 2000
!Layer 3 VNI
  vn-segment 9998
```

Layer 3 VNI

```
interface Vlan10
  no shutdown
  vrf member VRF-RED
  ip address 192.168.10.254/24 tag 12345
  ipv6 address 2001::1/64 tag 12345
  fabric forwarding mode anycast-gateway
```

```
interface Vlan20
  no shutdown
  vrf member VRF-BLUE
  ip address 192.168.20.254/24 tag 12345
  ipv6 address 2002::1/64 tag 12345
  fabric forwarding mode anycast-gateway
```

Layer 3 VNI

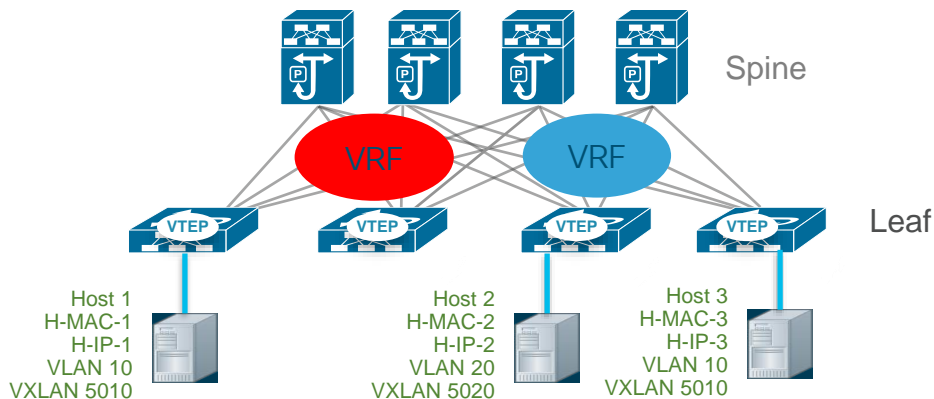
```
interface nve1
  source-interface loopback0
  host-reachability protocol bgp
  member vni 5010
    mcast-group 239.1.1.1
  member vni 5020
    mcast-group 239.1.1.1
  member vni 9999 associate-vrf
  member vni 9998 associate-vrf
```

Map L2VNI to NVE

Associate L3VNI to NVE

```
vrf context VRF-RED
  vni 9999
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
  evpn
    vni 5010 12
    rd auto
    route-target both auto
```

```
vrf context VRF-BLUE
  vni 9998
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
  evpn
    vni 5020 12
    rd auto
    route-target both auto
```



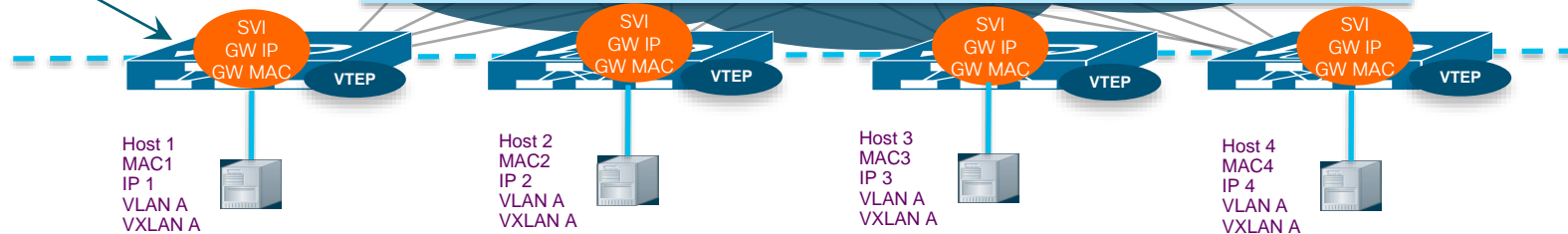
Distributed Anycast Gateway in MP-BGP EVPN

The same anycast gateway virtual IP address and MAC address are configured on all VTEPs in the VNI.

```
# VLAN to VNI mapping
vlan 20
  vn-segment 5020

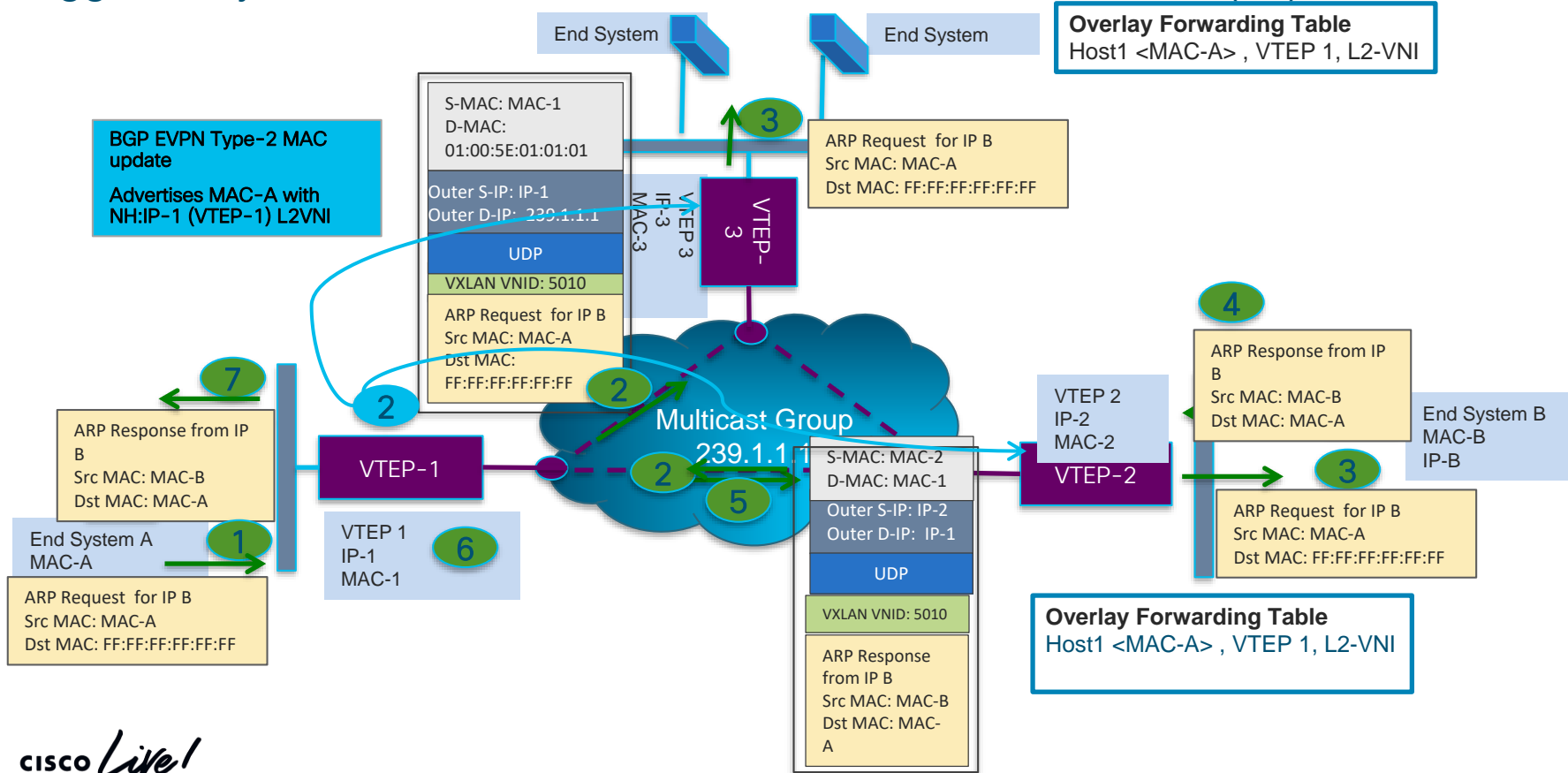
# Anycast Gateway MAC, identically configured on all VTEPs
fabric forwarding anycast-gateway-mac 0002.0002.0002

# Distributed IP Anycast Gateway (SVI)
# Gateway IP address needs to be identically configured on all VTEPs
interface vlan 20
  no shutdown
  vrf member VRF-BLUE
  ip address 192.168.20.254/24
  ipv6 address 2002::1/64
  fabric forwarding mode anycast-gateway
```



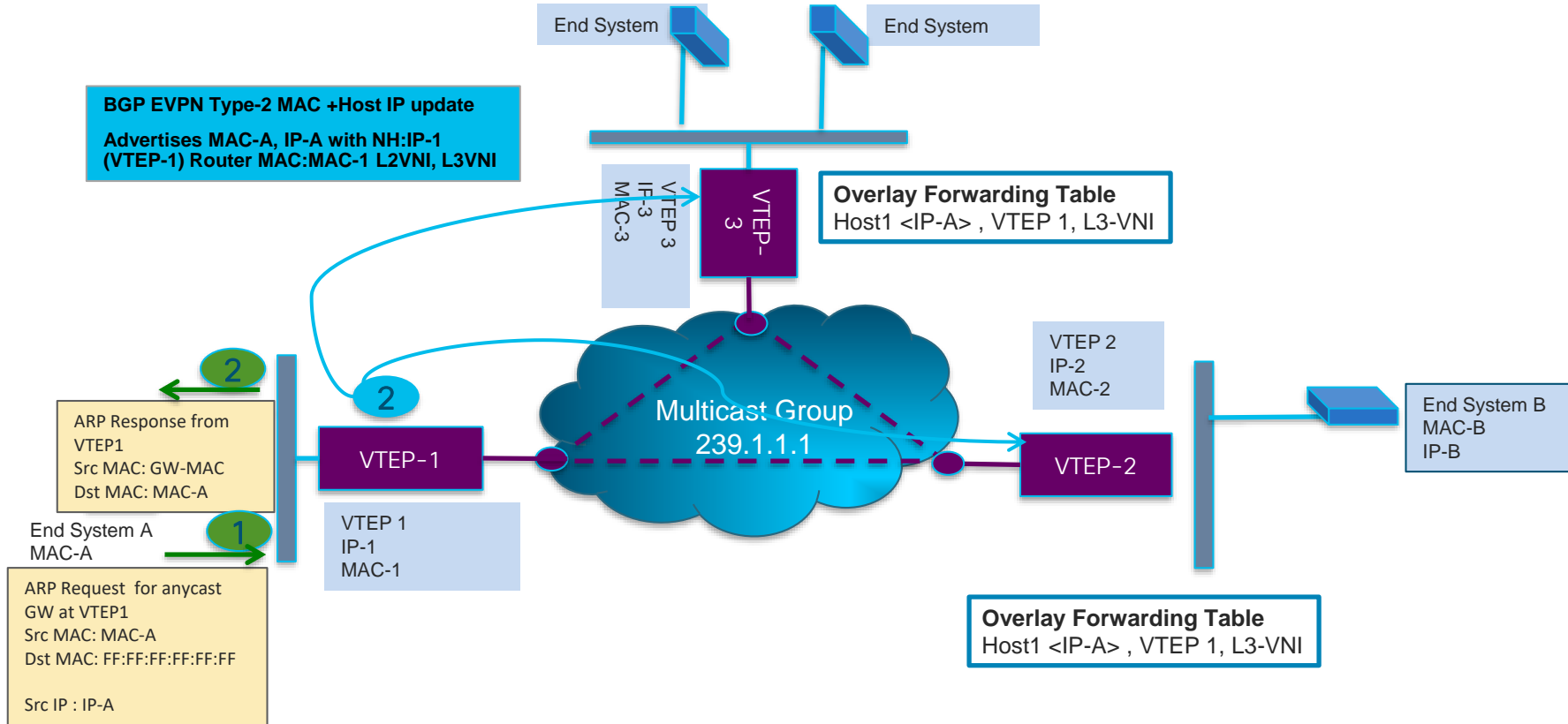
EVPN Peer and Endpoint(Host) Discovery

Triggered by Host Communication across the same VLAN/VNI (L2)



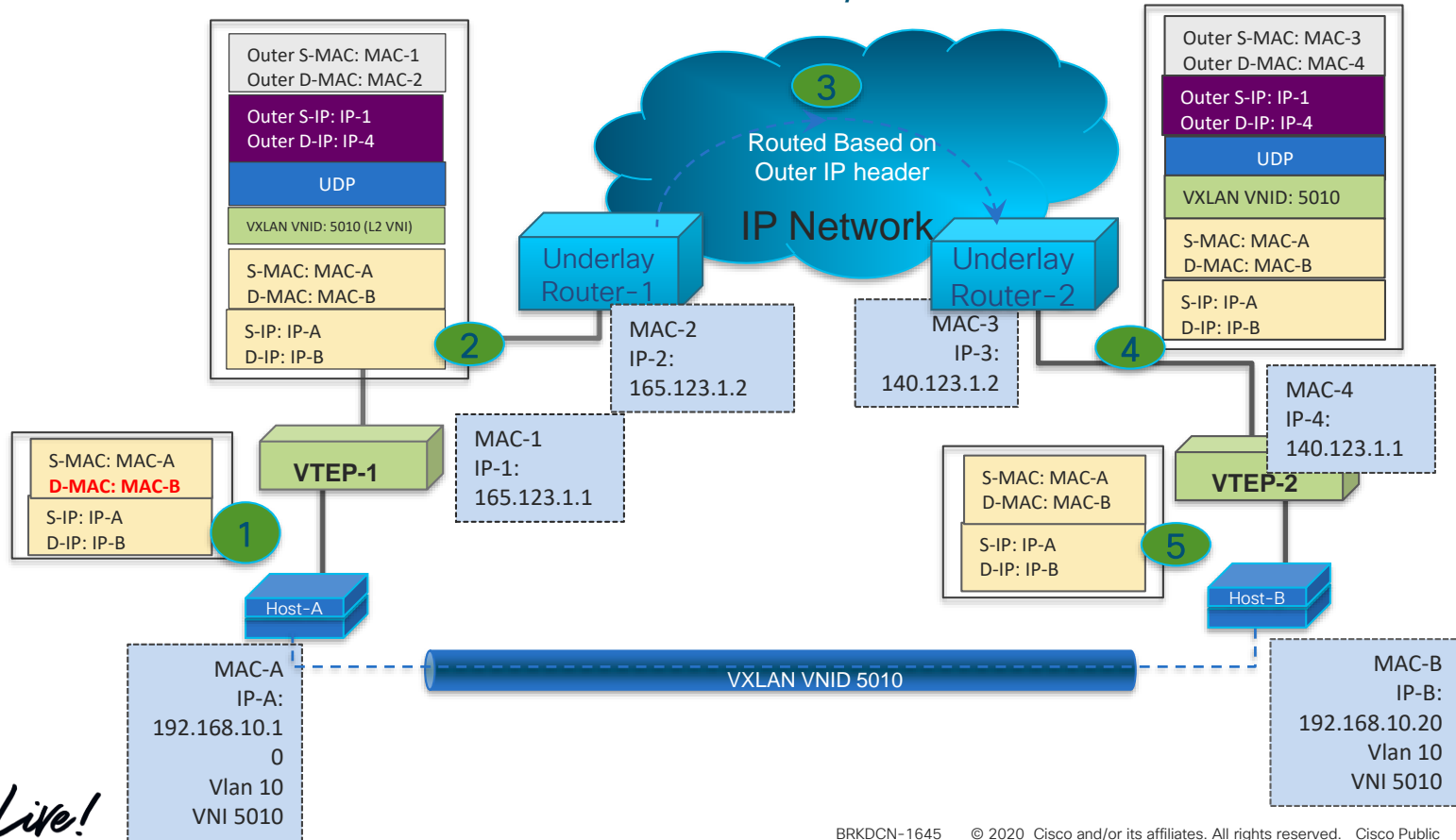
EVPN Peer and Endpoint(Host) Discovery

Triggered by Host Communication between VLAN/VNI (L3)



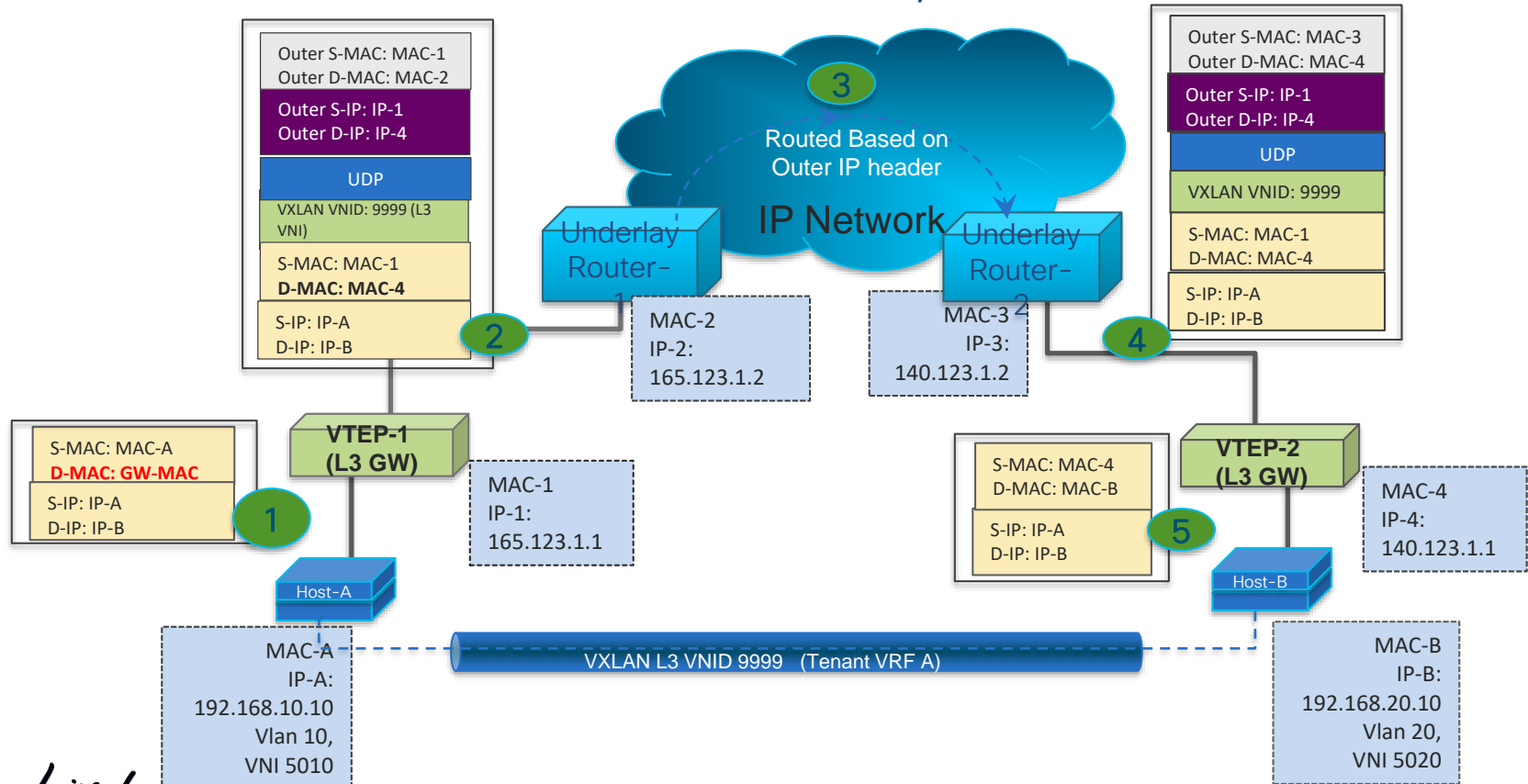
Packet Walk

Communication between hosts in same VLAN/VNI

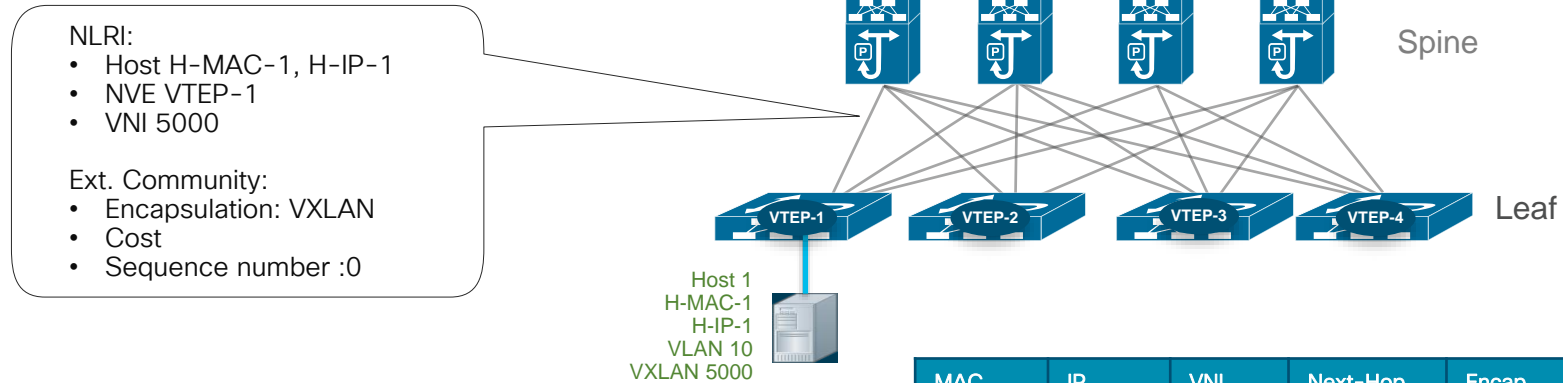


Packet Walk

Communication between hosts in different VLAN/VNI



EVPN Control Plane --- VM Mobility



1. Host 1 attaches to VTEP-1
2. VTEP-1 detects Host1 and advertises H1 with seq #0
3. Other VTEPs learn about the host route of Host 1

MAC	IP	VNI	Next-Hop	Encap	Seq#
H-MAC-1	H-IP-1	5000	VTEP-1	VXLAN	0

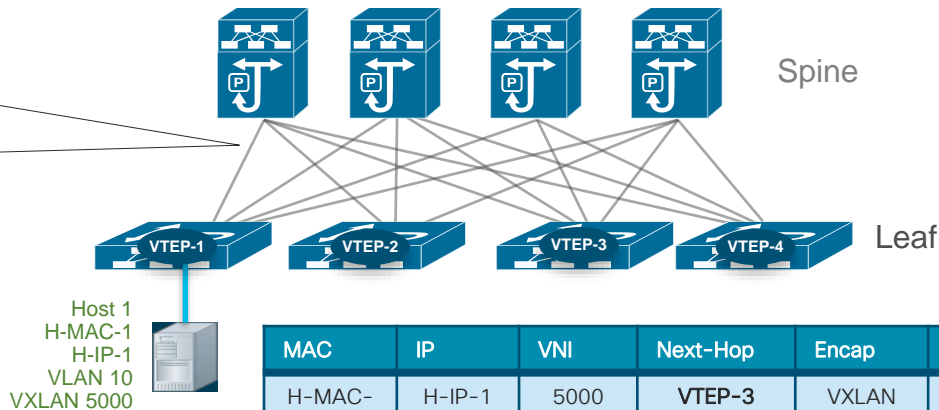
EVPN Control Plane --- VM Mobility

NLRI:

- Host H-MAC-1, H-IP-1
- NVE VTEP-3
- VNI 5000

Ext. Community:

- Encapsulation: VXLAN
- Cost
- Sequence number: 1



1. Host 1 moves to VTEP-3 from VTEP-1
2. VTEP-3 detects Host 1, sends MP-BGP update for Host 1 with its own VTEP address and a new seq #1
3. Other VTEPs learn about the new route of Host 1 from VTEP 3 with a higher sequence number and prefer that update

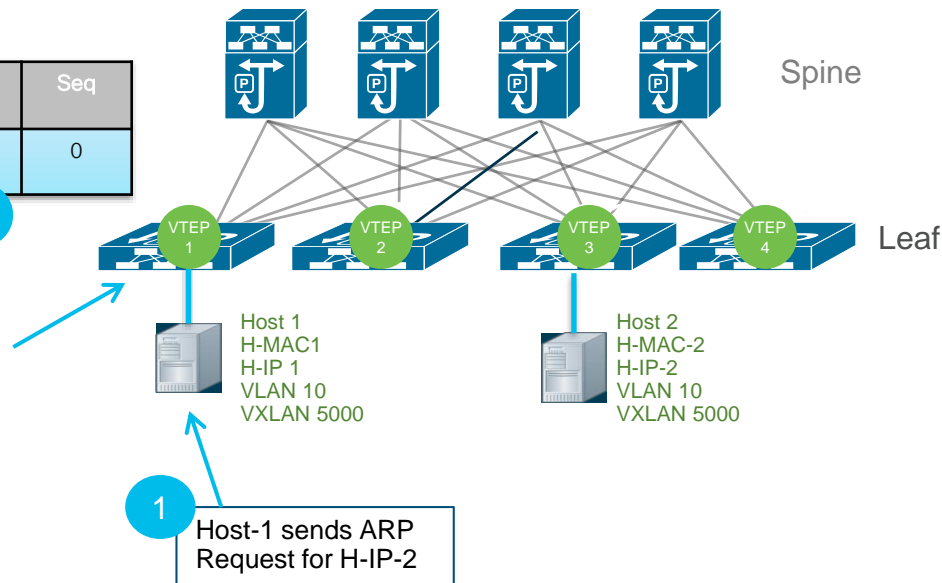
EVPN Control Plane --- ARP Suppression

Minimize flood-&-learn behavior for host learning

MAC	IP	VNI	Next-Hop	Encap	Seq
H-MAC-2	H-IP-2	5000	VTEP-3	VXLAN	0

2
VTEP-1 receives and intercepts the ARP Request. Checks in its own host table.

- If it has a match for H-IP-2, it'll send ARP response on behalf of Host-2
- If it doesn't have a match for H-IP-2, it'll forward the ARP request to remote VTEPs via multicast encap or head-end replication



Functions of VXLAN/EVPN

Host/Network
Reachability
Advertisement

Advertise host/network reachability information through control protocol (MP-BGP)

VTEP Security &
Authentication

Authenticate VTEPs through BGP peer authentication

Distributed
Anycast Gateway

Seamless and Optimal vm-mobility

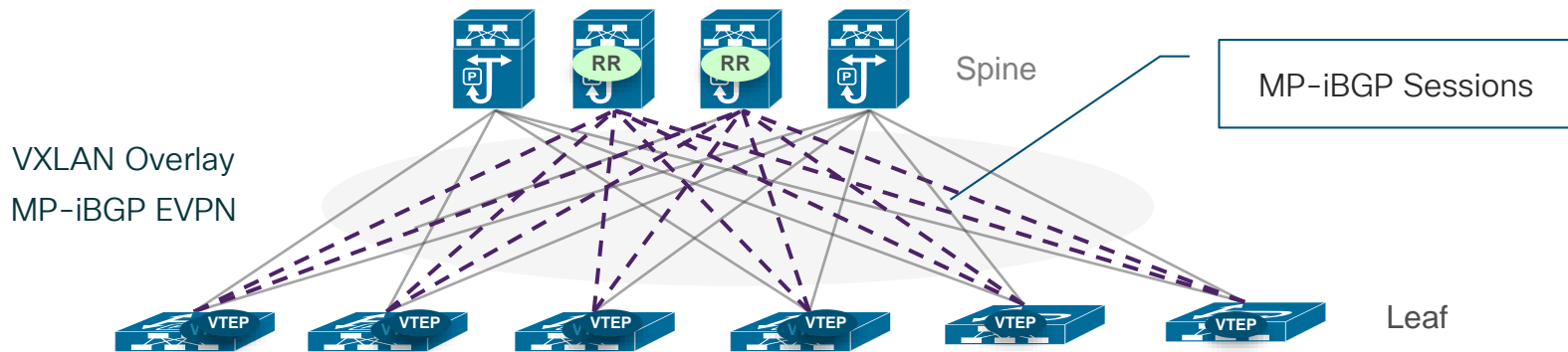
ARP Suppression

Early ARP termination
Localize ARP learning process
Minimize network flooding



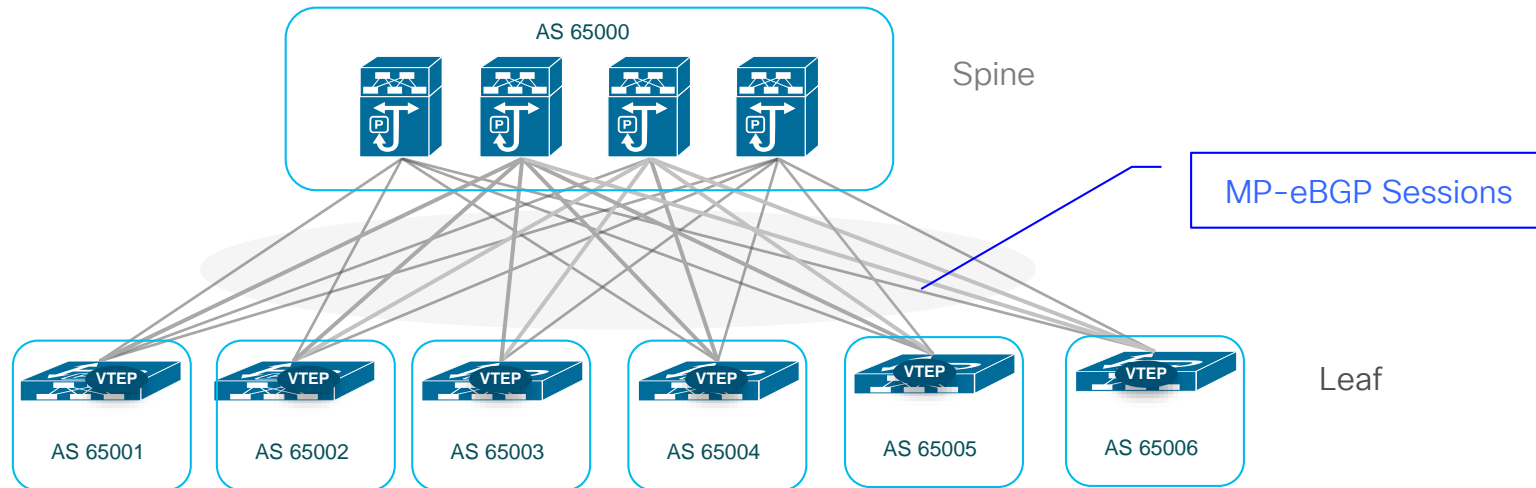
Design Options and Use case

VXLAN Fabric Design with MP-iBGP EVPN



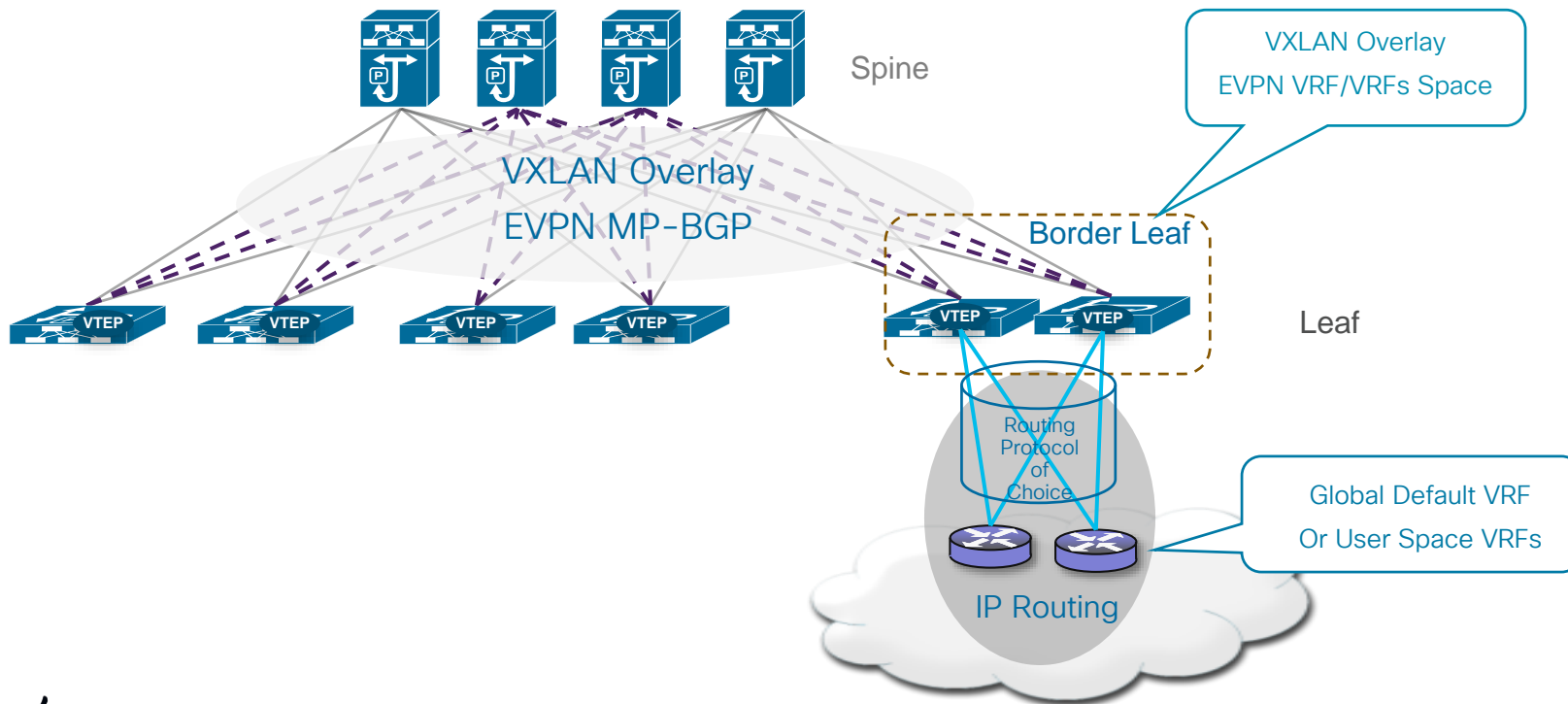
- VTEP Functions are on leaf layer
- Spine nodes are iBGP route reflector
- Spine nodes don't need to be VTEP

VXLAN Fabric Design with MP-eBGP EVPN

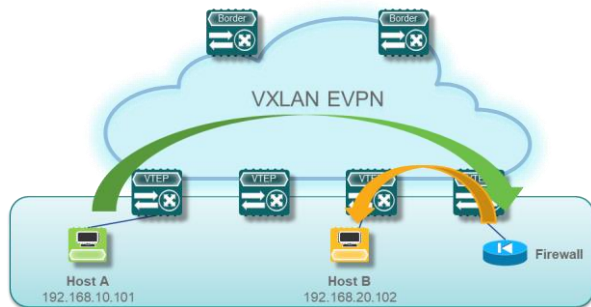


- VTEP Functions are on leaf layer
- Spine nodes are MP-eBGP Peers to VTEP leafs
- Spine nodes don't need to be VTEP
- VTEP leafs can be in the same or different BGP AS's

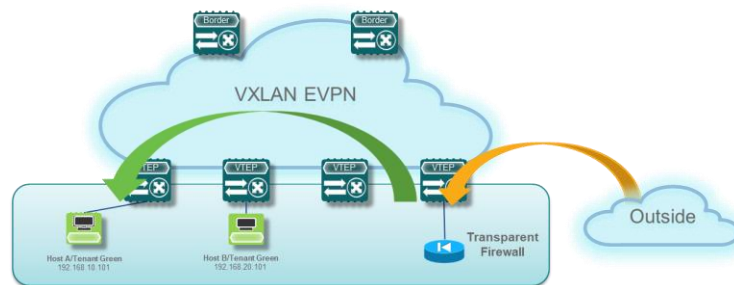
VXLAN Fabric - External Routing



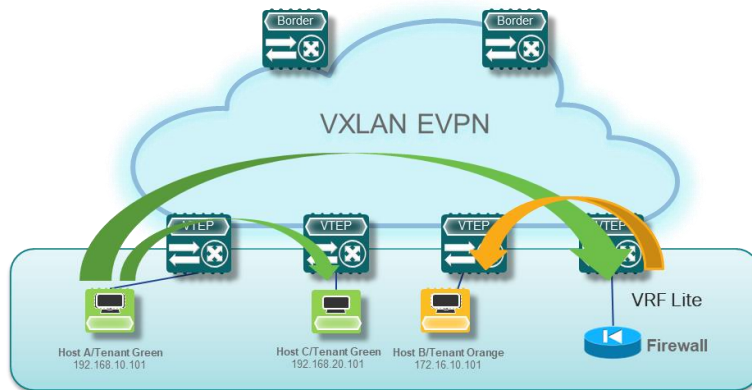
VXLAN Fabric – Service Insertion



Firewall as a default gateway : Centralized Gateway - Firewall bottleneck

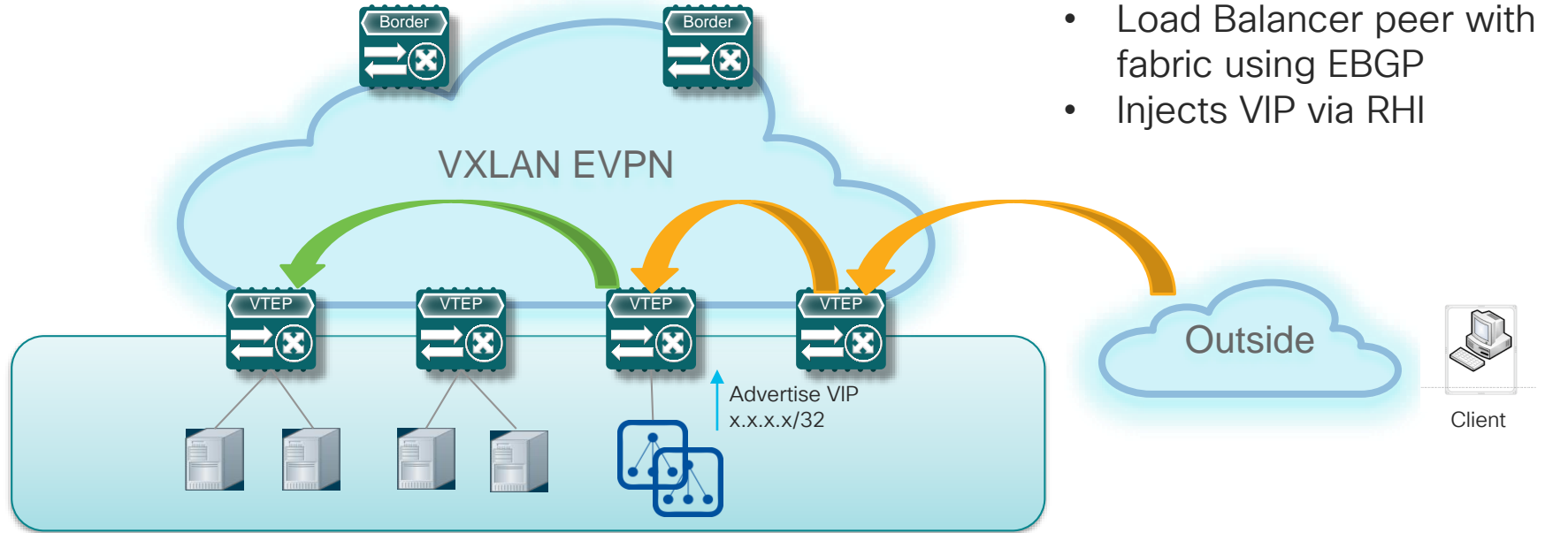


Transparent Firewall : Inspect and then bridge Traffic between “dirty” VLAN and “clean” VLAN

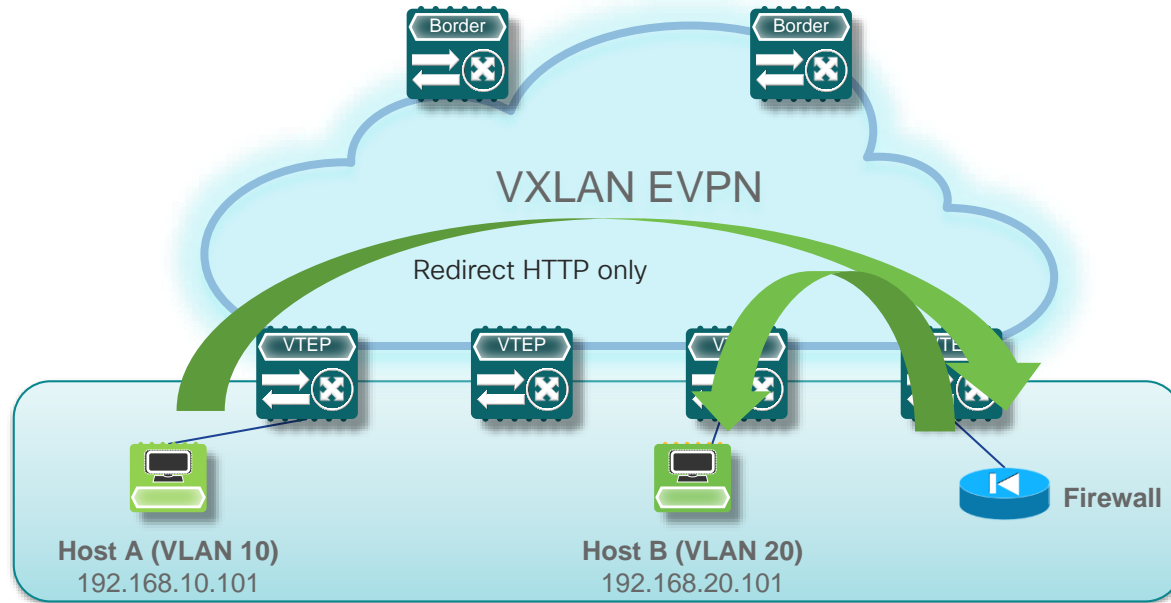


Tenant Edge Firewall: Traffic between Tenants/VRFs routed via the firewall

VXLAN Fabric – Service Insertion



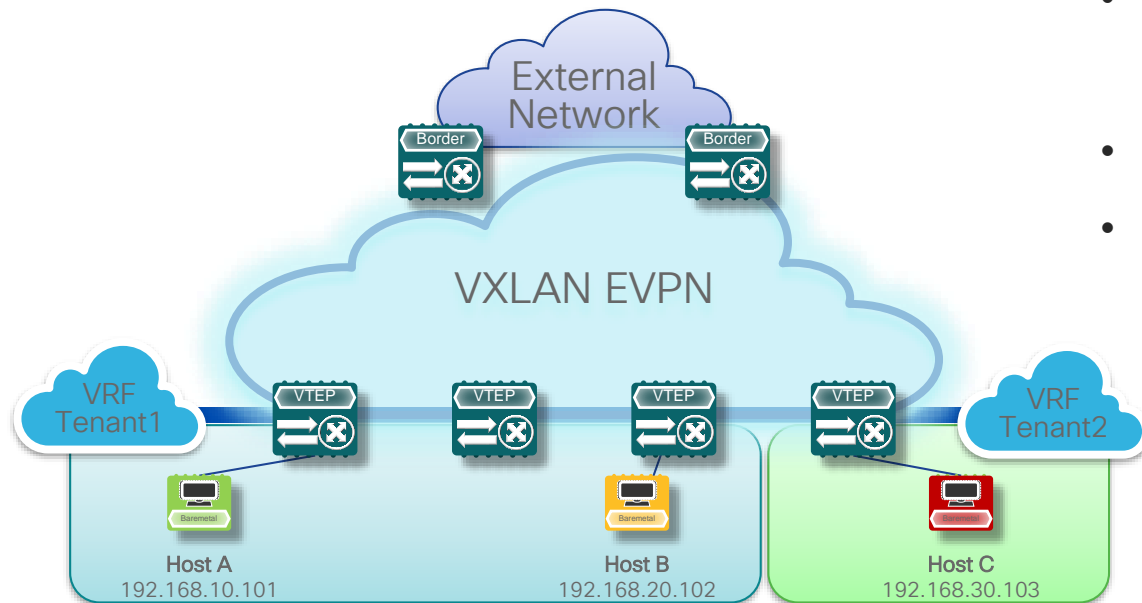
VXLAN Fabric – Selective Traffic Redirection



- Leverages Policy Based Redirect
- Inter VLAN traffic bypass default routing lookup and redirected
- Service Redirection to Load Balancers, Firewalls etc.

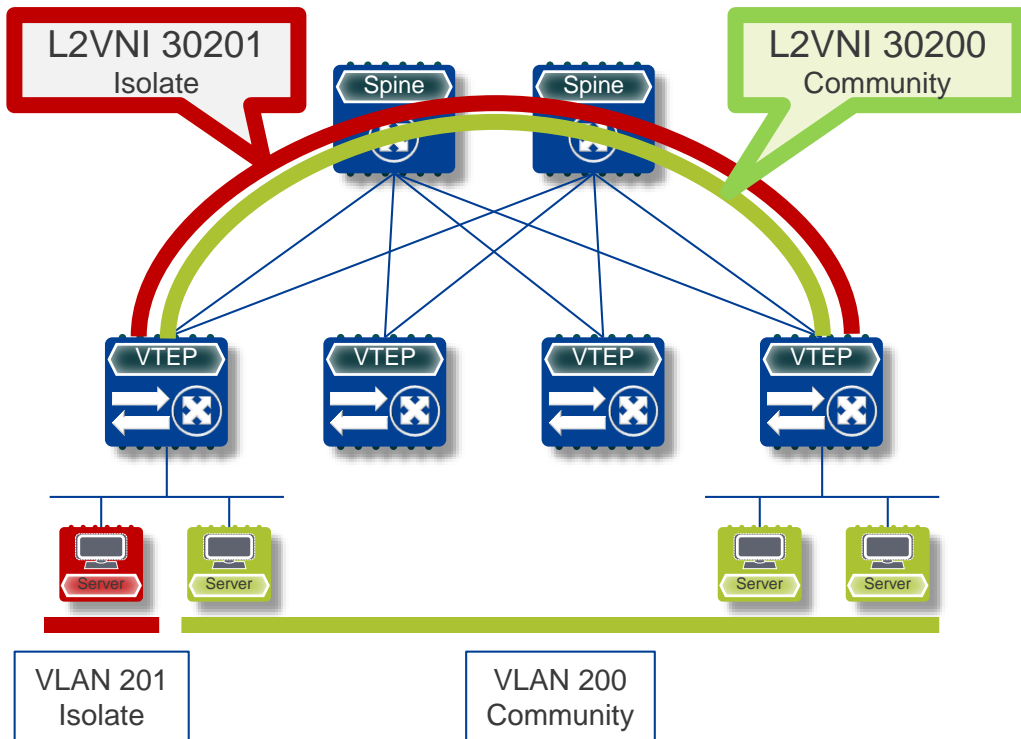
VXLAN Fabric – Centralized Route Leaking

Extranet Support



- Use Cases – Shared Services, External Connectivity
- VRF to VRF or VRF to Default
- Centralize Location for leaking routes

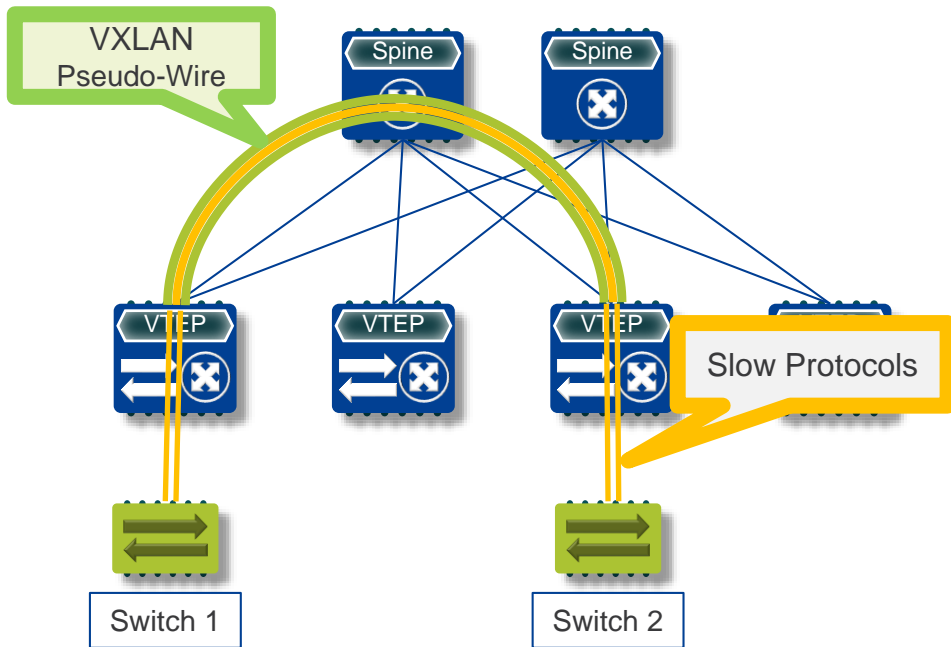
VXLAN Fabric – Private VLAN over VXLAN



Private VLAN with VXLAN

- Extending Private VLAN over VXLAN
- Sub-VLAN Segmentation
- Availability of 2nd VLAN Modes
 - Community VLAN across VXLAN
 - Promiscuous VLAN across VXLAN
 - Isolate VLAN localized but extended across VXLAN

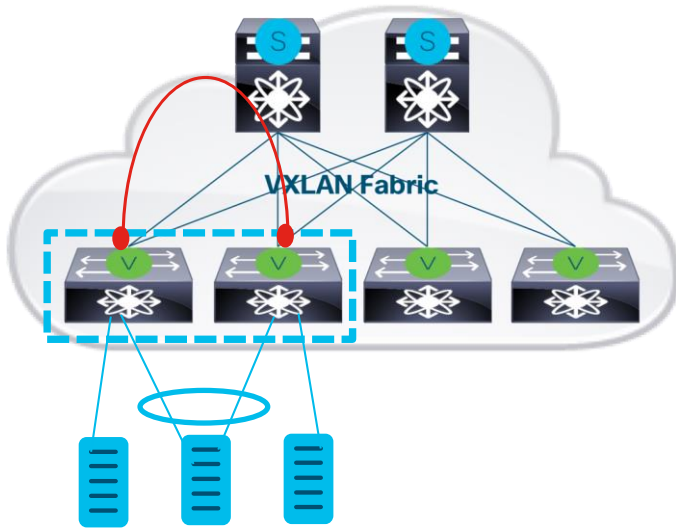
VXLAN Fabric – VXLAN Pseudo wire(Xconnect)



VXLAN Pseudo-Wire

- Cross-Connect (X-Connect) concept
 - Point-2-Point
- Enables Protocol Tunneling for
 - STP, CDP, LLDP, PAGP, LACP, BFD

Peerlink-Less VPC

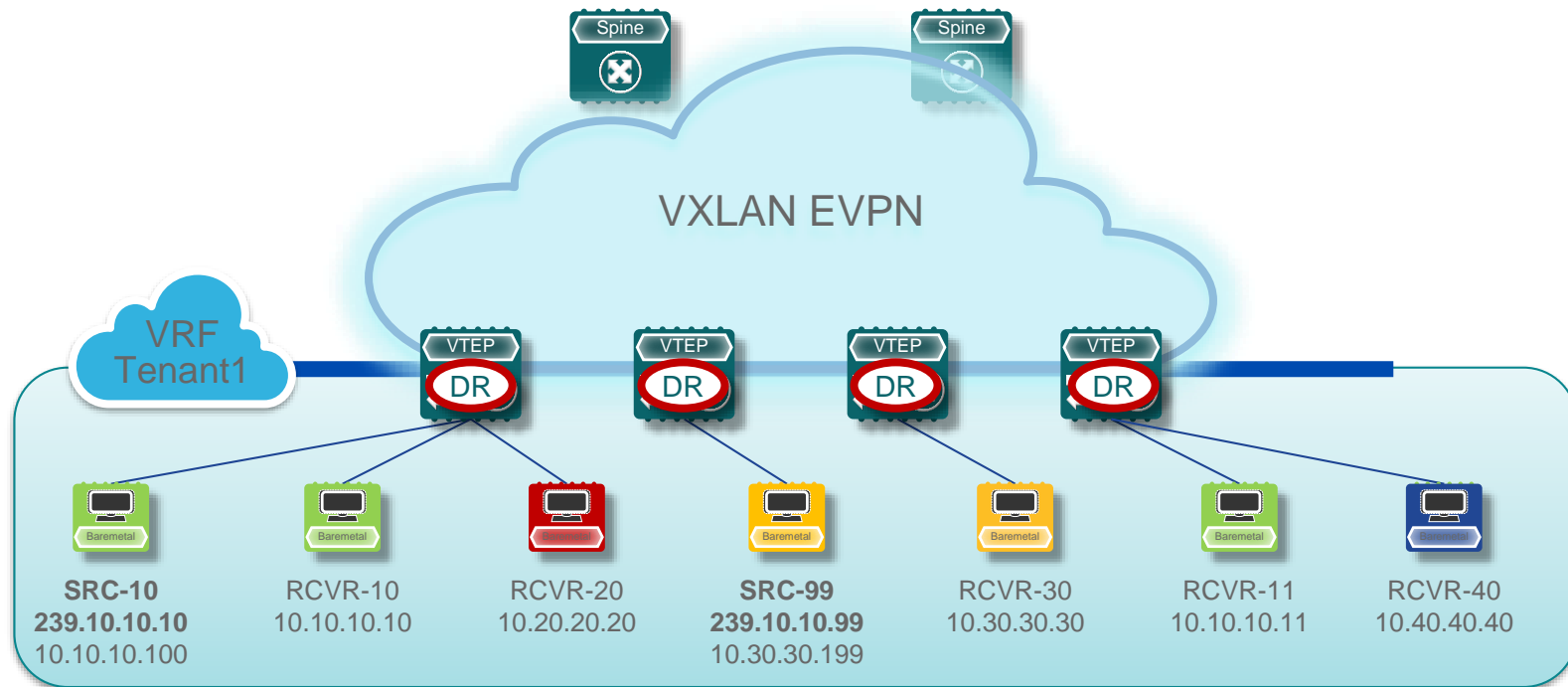


Enhanced dual-homing solution without wasting physical ports

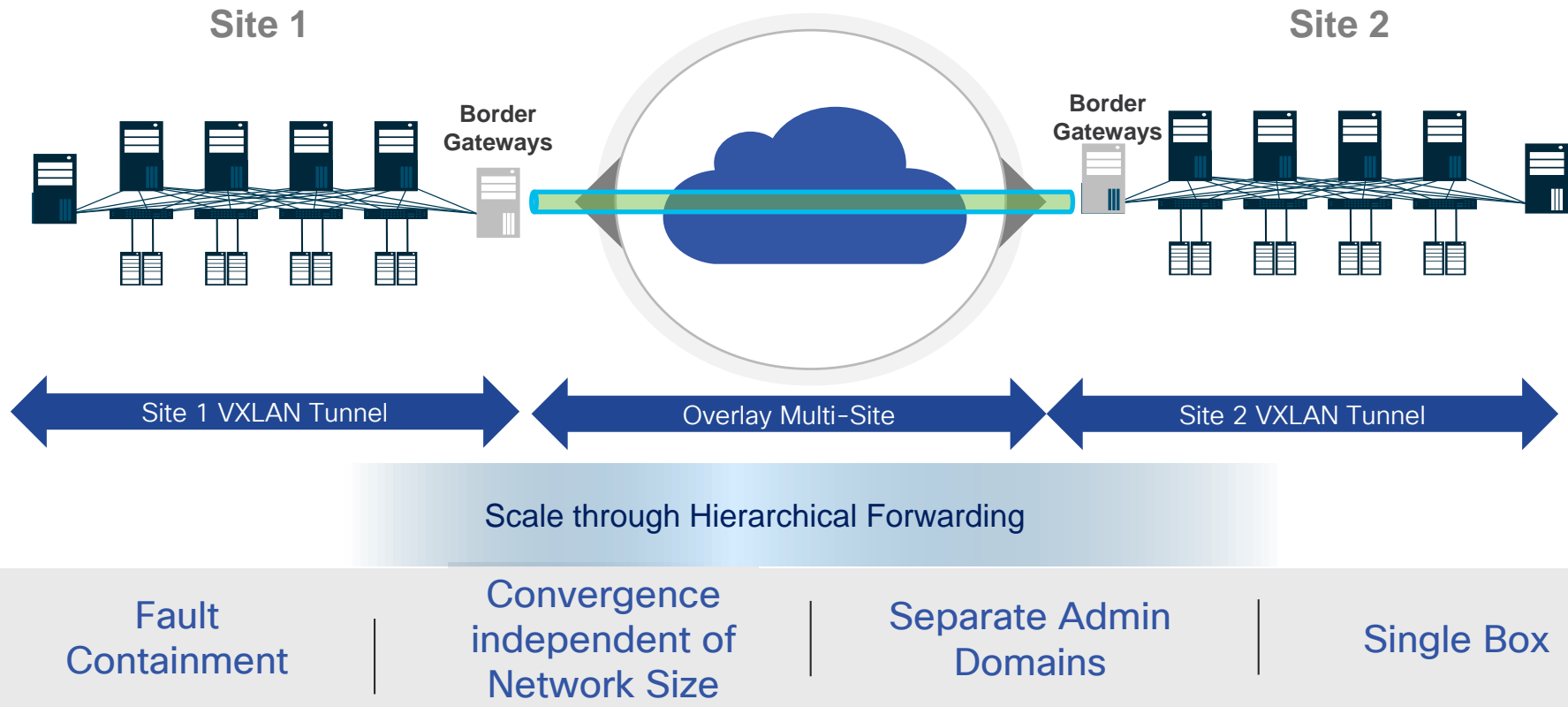


Preserve traditional vPC characteristics

VXLAN Fabric – Tenant Routed Multicast



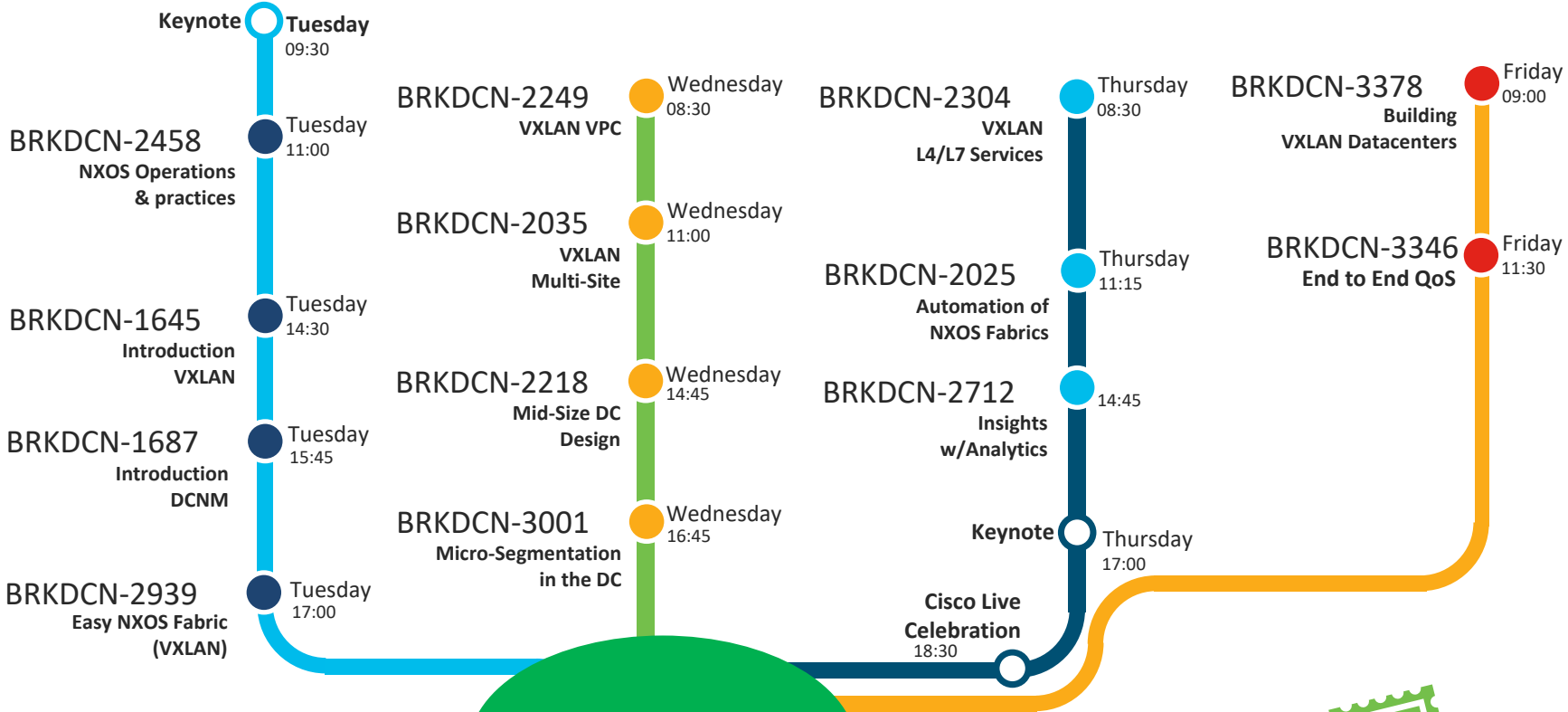
VXLAN EVPN Multi-Site



Summary

Summary

- VXLAN enables scalable Data Center fabrics
- BGP EVPN with VXLAN provides a robust control plane enabling multi-tenancy, VM mobility , optimizing traffic forwarding
- Seamless integration with service nodes such as Firewalls and Load balancers and ability to provide shared services
- Fabric can cater to multicast traffic in the overlay
- VXLAN as a DCI with Multi-Site



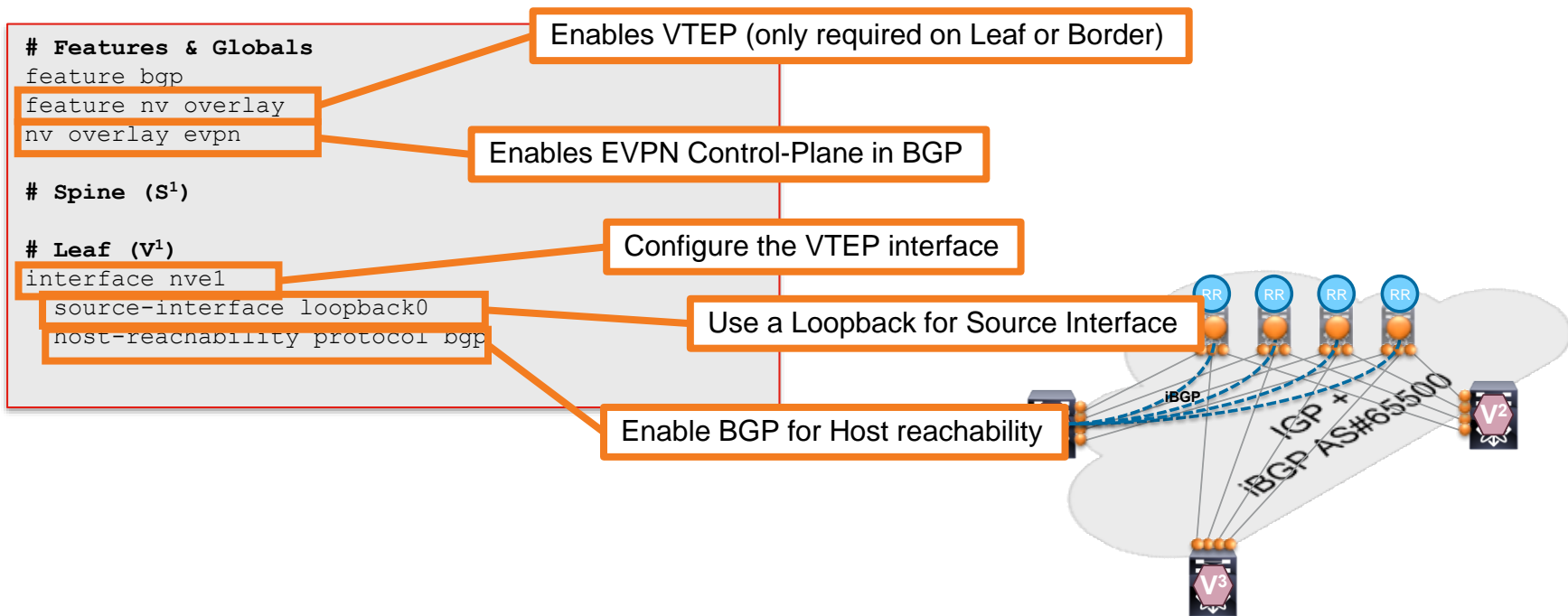
CISCO *Live!*

NXOS

Fabric Technology



Building your VTEP (VXLAN Tunnel End-Point)



*Simplified BGP configuration; would have 4 BGP peers (RR)
IGP not shown

Building your EVPN MP-BGP Control-Plane

Features & Globals

```
feature bgp
feature nv overlay
nv overlay evpn
```

Enables EVPN Control-Plane in BGP

Spine (S¹)

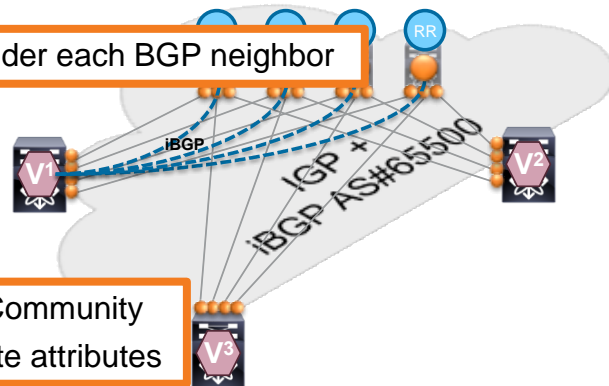
```
router bgp 65500
router-id 10.10.10.1
address-family ipv4 unicast
address-family l2vpn evpn
neighbor 10.10.10.0/24 remote-as 65500
update-source loopback0
address-family l2vpn evpn
send-community both
route-reflector-client
```

Activate L2VPN EVPN under each BGP neighbor

Leaf (V¹)

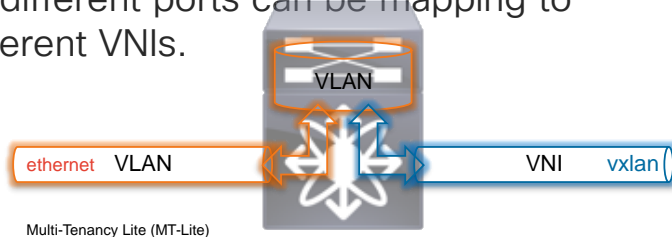
```
router bgp 65500
router-id 10.10.10.10
address-family ipv4 unicast
neighbor 10.10.10.1 remote-as 65500
update-source loopback0
address-family l2vpn evpn
send-community both
```

Send Extended BGP Community to distribute EVPN route attributes



Extend your VLAN to VXLAN

- VLAN to VNI configuration on a per-Switch based
- VLAN becomes “Switch Local Identifier”
- VNI becomes “Network Global Identifier”
- 4k VLAN limitation per-Switch does still apply
- 4k Network limitation has been removed
- VLAN can be port-significant. The same vlan on different ports can be mapping to different VNIs.



Features

```
feature vn-segment-vlan-based
```

VLAN to VNI mapping (MT-Lite)

```
Vlan 10
vn-segment 5010
```

VLAN to Layer-2 VNI mapping

Activate Layer-2 VNI for EVPN

```
evpn
vni 5010 12
rd auto
route-target import auto
route-target export auto
```

Enables EVPN Control-Plane for Layer-2 Services

Activate Layer-2 VNI on VTEP

```
interface nve1
source-interface loopback0
host-reachability protocol bgp
member vni 5010
mcast-group 239.239.239.100
suppress-arp
```

Alternative is to use “ingress-replication protocol bgp”

Enables Layer-2 VNI on VTEP and suppress ARP

Distributed Anycast Gateway for Extended V

FYI

- All VTEPs in a VXLAN are the distributed anycast gateway for its IP subnet.
- All VTEPs in a VXLAN need to be configured with an identical anycast gateway virtual MAC address
- All VTEPs in a VXLAN need to be configured

One gateway virtual MAC per VTEP

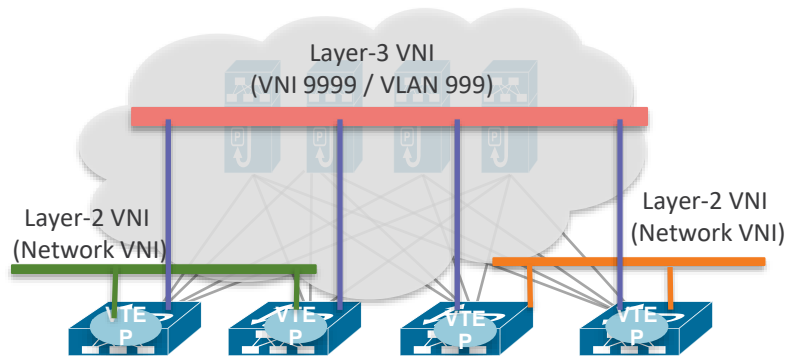
One gateway virtual IP per VLAN/VXLAN

```
# VLAN to VNI mapping
vlan 200
  vn-segment 5200

# Anycast Gateway MAC, identically configured on all VTEPs
fabric forwarding anycast-gateway-mac 0002.0002.0002

# Distributed IP Anycast Gateway (SVI)
# Gateway IP address needs to be identically configured on all VTEPs
interface vlan 200
  no shutdown
  vrf member VRF-A
  ip address 20.0.0.1/24
  fabric forwarding mode anycast-gateway
```

Routing in VXLAN – Define the Resources



1:1 mapping between L3 VNI
and tenant VRF

Configuration Example for VRF-A

Define VLAN for VRF routing instance

```
Vlan 999
  vn-segment 9999
```

VLAN to Layer-3 VNI mapping

Define SVI for VRF routing instance

```
interface Vlan999
  no shutdown
  mtu 9216
  vrf member VRF-A
  ip forward
```

VLAN to Layer-3 VNI mapping
- ip forward required for prefix-based routing

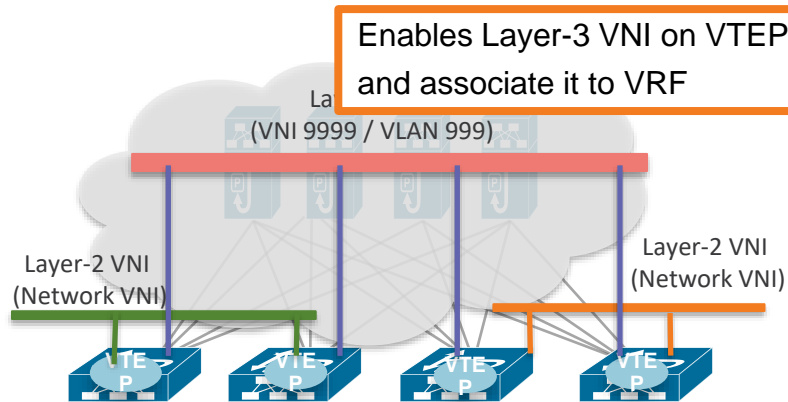
VRF configuration for "customer" VRF

```
vrf context VRF-A
  vni 9999
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
```

VRF context definition

- VNI
- Route-Distinguisher
- Route-Targets
- IPv4 and/or IPv6

Routing in VXLAN – Configure the Routing



1:1 mapping between L3 VNI and tenant VRF

VRF/Tenant definition within Overlay Control-Plane

Configuration Example for VRF-A

Activate Layer-3 VNI on VTEP

```
interface nvel
  source-interface loopback0
  host-reachability protocol bgp
  member vni 5010
  mcast-group 239.239.239.100
  suppress-arp
  member vni 9999 associate-vrf
```

Route-Map for Redistribute Subnet

```
route-map REDIST-SUBNET permit 10
  match tag 12345
```

Control-Plane configuration for VRF (Tenant)

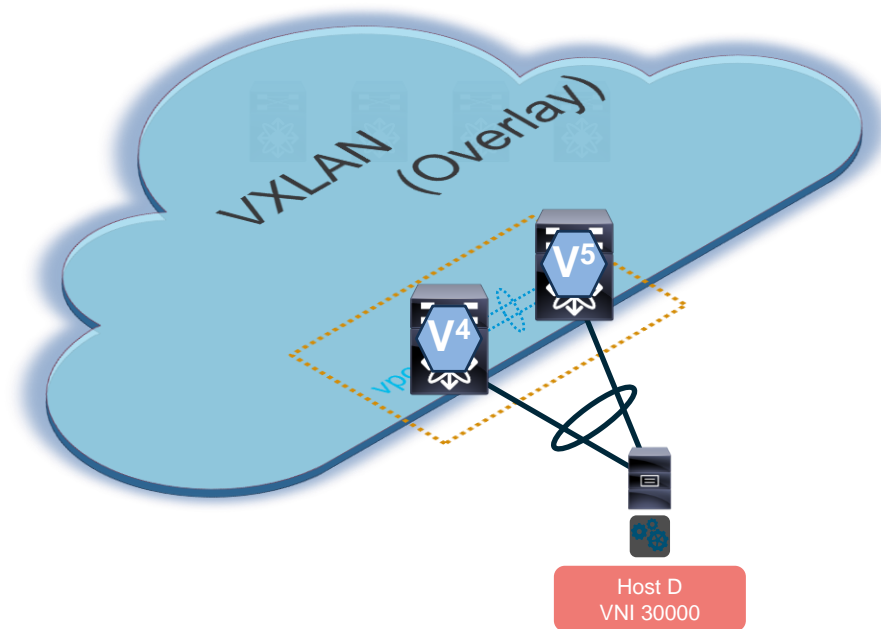
```
router bgp 65500
```

vrf VRF-A

```
address-family ipv4 unicast
  advertise l2vpn evpn
  redistribute direct route-map REDIST-SUBNET
  maximum-paths ibgp 2
```

VXLAN Hardware Gateway Redundancy (vPC)

- Redundant connectivity for classic Ethernet hosts
- Extend the IP Interface (Loopback) configuration for the vPC VTEP
 - Secondary IP address (anycast) is used as the anycast VTEP address
 - Both vPC VTEP switches need to have the identical secondary IP address configured under the loopback interface



VXLAN Hardware Gateway Redundancy (vPC)

vPC VTEP Configuration Example

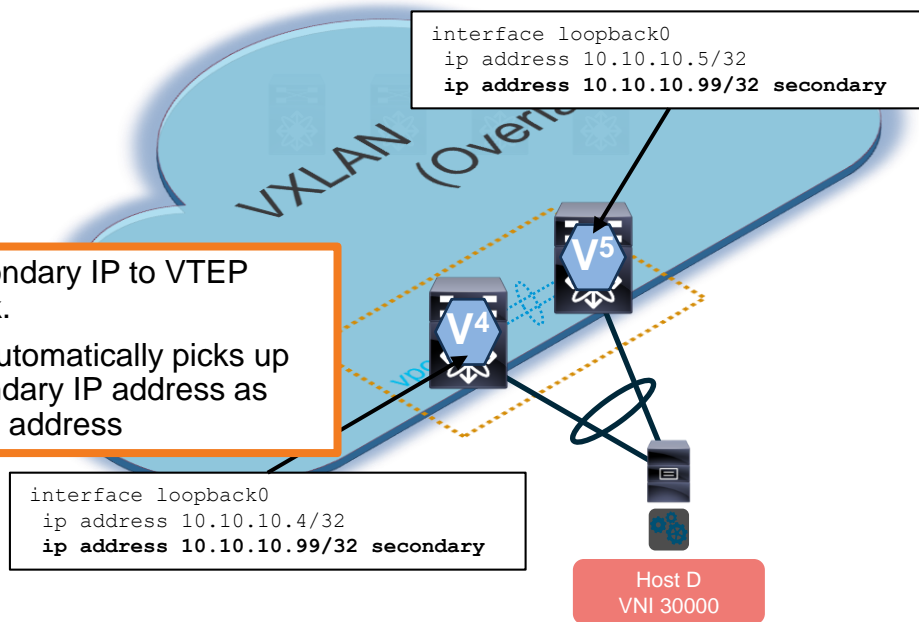
```
# VLAN to VNI mapping (MT-Lite)
vlan 55
  vn-segment 30000
# VTEP IP Interface; Source/Destination for all
# VXLAN Encapsulated Traffic.
  Primary IP address is used for Orphan Hosts
  Secondary IP is for vPC Hosts (same IP on both
  vPC Peers)
interface loopback0
  ip address 10.10.10.5/32
  ip address 10.10.10.99/32 secondary
# VTEP configuration using Loopback as source.
interface nve1
  source-interface loopback0
  host-reachability protocol bgp
  member vni 5010
  mcast-group 239.239.239.100
  suppress-arp
  member vni 9999 associate-vrf
```

Add Secondary IP to VTEP Loopback.

VXLAN automatically picks up the secondary IP address as the VTEP address

```
interface loopback0
  ip address 10.10.10.4/32
  ip address 10.10.10.99/32 secondary
```

```
interface loopback0
  ip address 10.10.10.5/32
  ip address 10.10.10.99/32 secondary
```



VXLAN Hardware Gateway Redundancy (vPC)

vPC VTEP Configuration Example

VPC Domain Configuration

```
vpc domain 99
peer-switch
peer-keepalive destination V4-mgmt source V4-mgmt
peer-gateway
ip arp synchronize
```

peer-gateway needs to be enabled so that vPC VTEP switches can forward traffic for each other's router MAC address

VPC Peer-Link

```
interface port-channelXX
switchport mode trunk
vpc peer-link
```

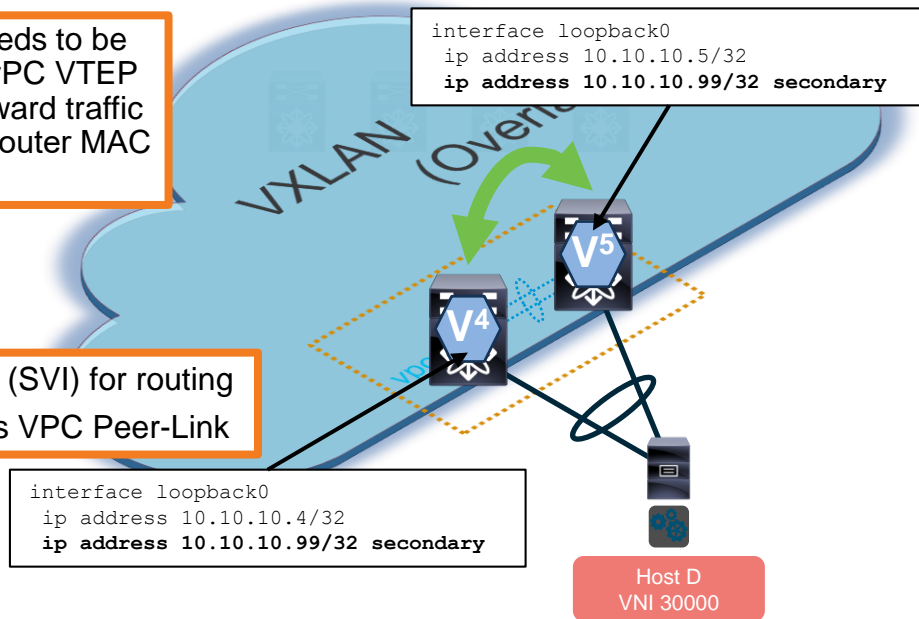
VPC Domain Routing Adjacency

```
interface Vlan3999
no shutdown
ip address 10.254.254.1/30
ip router ospf 1 area 0.0.0.0
ip ospf network point-to-point
ip pim sparse-mode
```

Routed Interface (SVI) for routing adjacency across VPC Peer-Link

```
interface loopback0
ip address 10.10.10.4/32
ip address 10.10.10.99/32 secondary
```

```
interface loopback0
ip address 10.10.10.5/32
ip address 10.10.10.99/32 secondary
```



eBGP EVPN Configuration (1)

Next-hop Unchange

- BGP next-hop is used as the tunnel tail end address. It shall be the advertising VTEP's address.
- Ensure the next-hop in the BGP route isn't changed during the route distribution
- eBGP changes next-hop by default. Need to change the policy to next-hop unchanged

Set next-hop policy not to change the next-hop attribute

eBGP configuration on a spine switch

```
route-map permit-all permit 10
route-map nh-unchange permit 10
  set ip next-hop unchanged
router bgp 65000
  router-id 10.1.1.1
  address-family ipv4 unicast
  address-family l2vpn evpn
  nexthop route-map nh-unchange
  retain route-target all
  neighbor 192.167.11.2 remote-as 65001
  address-family ipv4 unicast
  address-family l2vpn evpn
    send-community extended
  route-map permit-all out
```

eBGP EVPN Configuration(2)

Manually configure import/export route-target

- With eBGP, VTEPs will have different route-targets if using auto RT generation
- Need to manually configure RTs on eBGP peers so that they have the same RTs

Manually configure route-target for VRF

Manually configure route-target for L2 VNI under EVPN

```
vrf context evpn-tenant-1
vni 9999
rd auto
address-family ipv4 unicast
route-target import 100:9999
route-target import 100:9999 evpn
route-target export 100:9999
route-target export 100:9999 evpn
evpn
vni 5010 l2
rd auto
route-target import 100:5010
route-target export 100:5010
```

Complete your online session survey



- Please complete your session survey after each session. Your feedback is very important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (starting on Thursday) to receive your Cisco Live t-shirt.
- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Content Catalog on ciscolive.com/emea.

Cisco Live sessions will be available for viewing on demand after the event at ciscolive.com.

Continue your education



Demos in the
Cisco campus



Walk-in labs



Meet the engineer
1:1 meetings



Related sessions



Thank you





You make **possible**