Cisco live!

January 29 – February 2, 2018 · Barcelona

LTRDCN-2223

Implementing VXLAN In a Data Center

Rahul Parameswaran, Technical Marketing Engineer

Shyla Karimanye, Consulting Engineer, Advanced Services

Kevin Cook, Solutions Architect, Advanced Services

Cisco live!

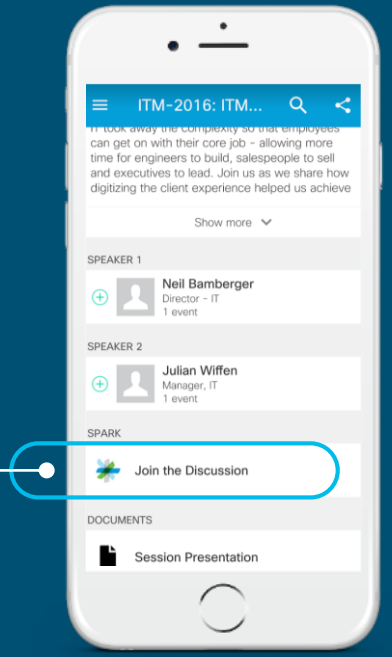
Cisco Spark

Questions?

Use Cisco Spark to communicate with the speaker after the session

How

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



cs.co/ciscolivebot#LTRDCN-2223

Agenda

- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

Prerequisites

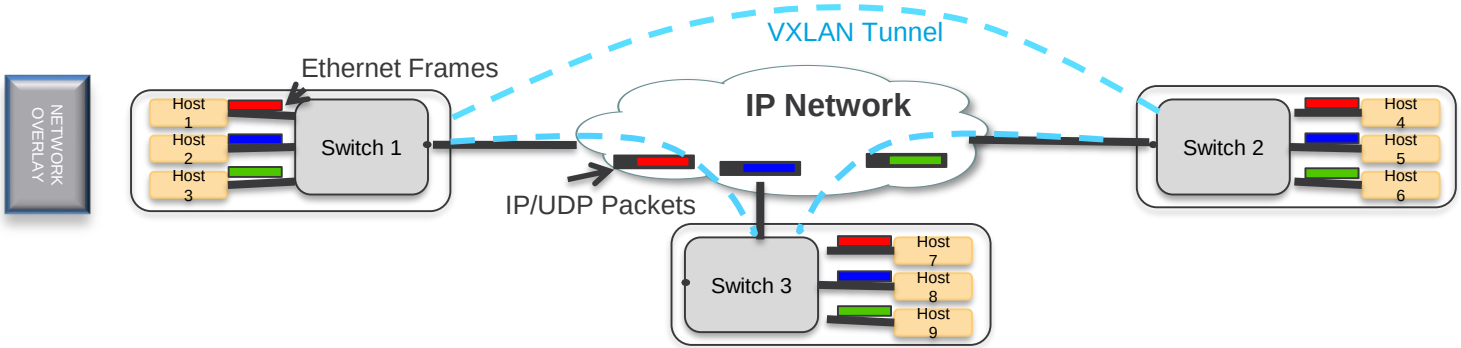
- Routing and Switching
- PIM
- MP-BGP
- NX-OS

Overlays

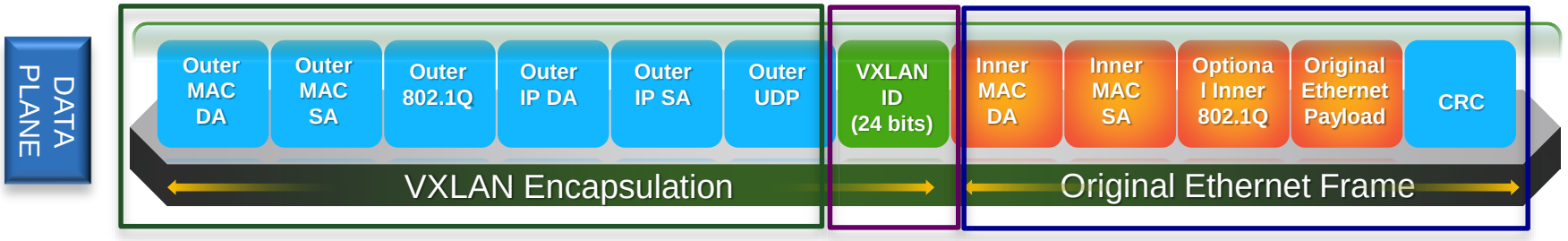
With growing adoption of virtualization in customer environments and a large number of workload mobility requirements in the data center, overlays are becoming a key technology. VXLAN is one of those overlay technologies.

Recap – What is VXLAN ?

- VXLAN is a 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)



Why VXLAN?

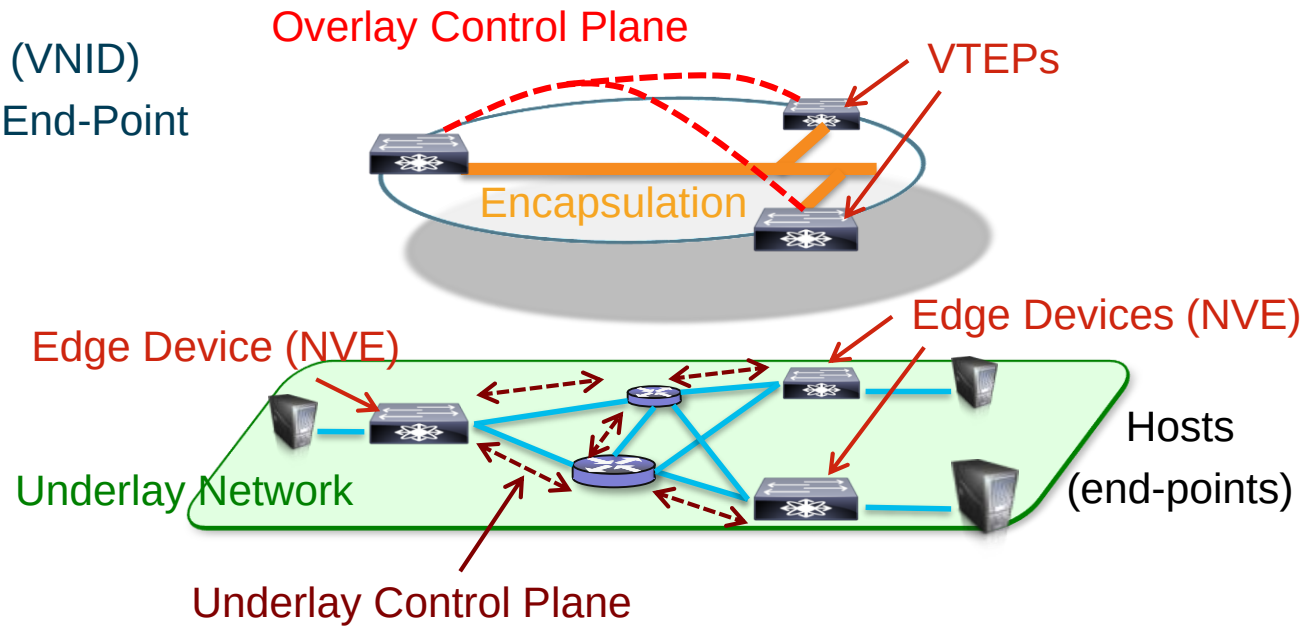
VXLAN provides a Network with Segmentation, IP Mobility, and Scale

- “Standards” based Overlay
- Leverages Layer-3 ECMP – all links forwarding
- Increased Name-Space to 16M identifier
- Segmentation and Multi-Tenancy
- Integration of Physical and Virtual

Overlay Taxonomy

Identifier = VN Identifier (VNID)

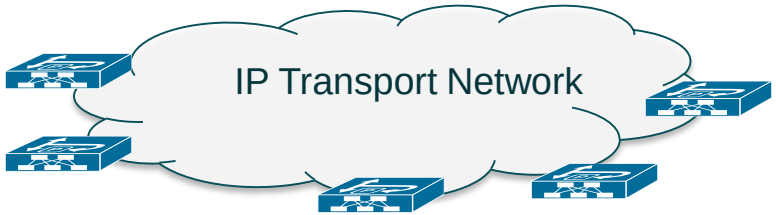
VTEP = VXLAN Tunnel End-Point



VXLAN Underlay Network – IP Routing

IP routed Network

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



Agenda

- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

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
- Security Risk

- Increased scale and stability
- Optimized workload mobility
- Increased Security

VXLAN BUM Traffic Handling

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

BU**M Traffic transport mechanisms**

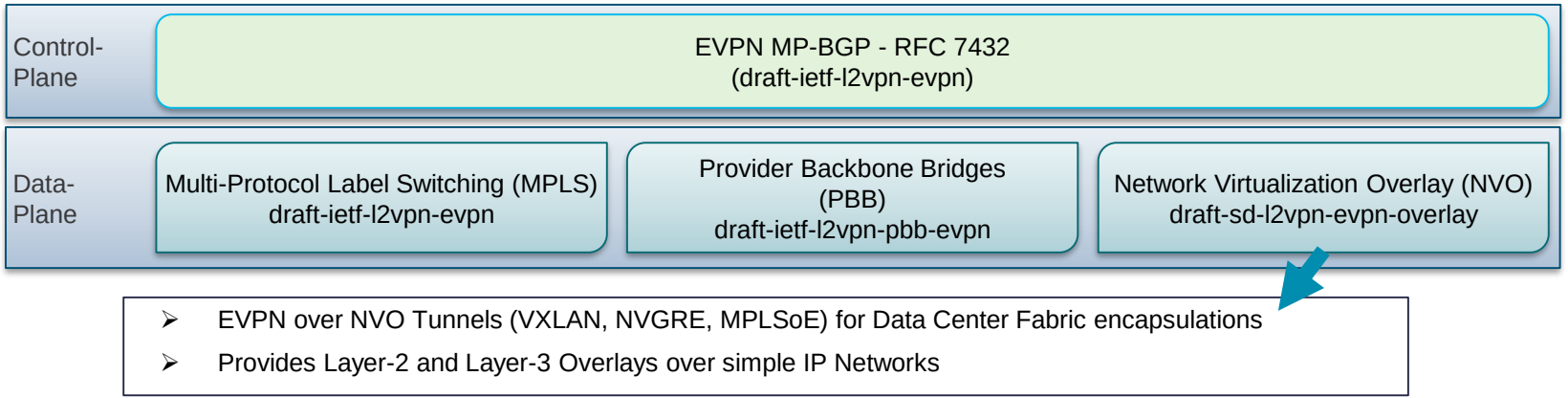
- Multicast replication
  - Requests the underlay network to run IP multicast
- Ingress unicast replication
  - One unicast replica per remote VTEP
  - Increase traffic load throughout the network

Agenda

- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

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



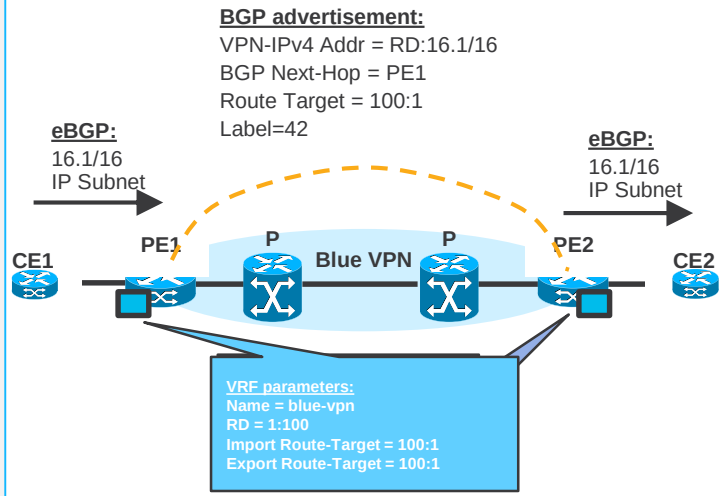
EVPN Primer --- MP-BGP Review

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

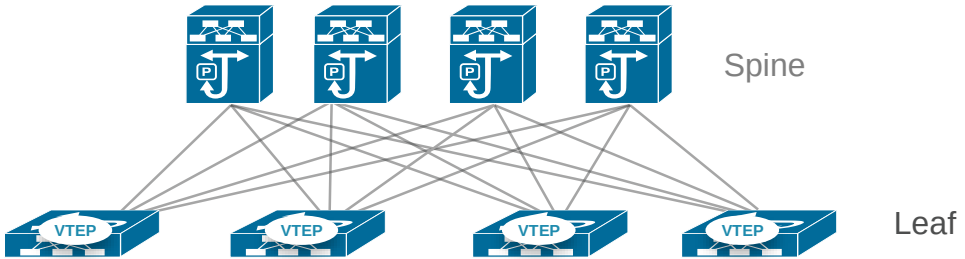
VPN Address-Family:
Distribute the MP-BGP VPN routes



EVPN Control Plane – Reachability Distribution

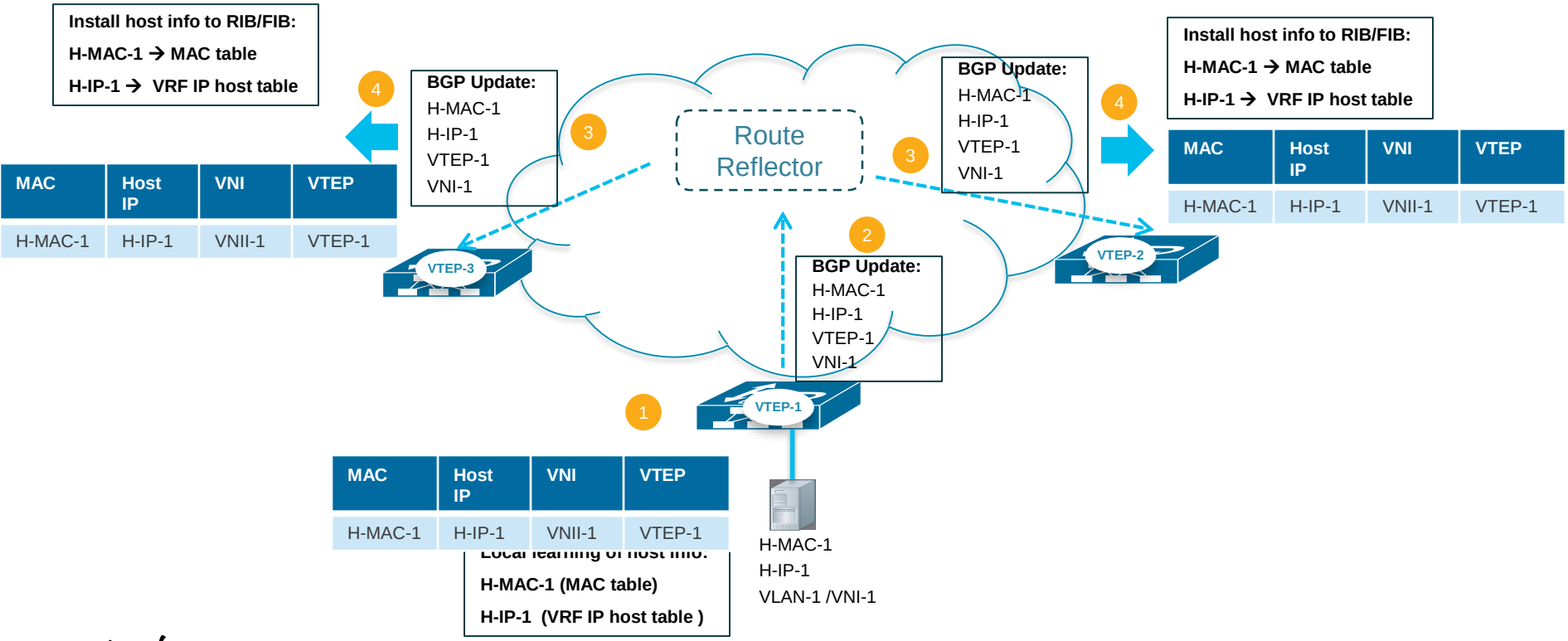
- EVPN Control Plane -- Host and Subnet Route Distribution

BGP Update
- Host-MAC
- Host-IP
- Internal IP Subnet
- External Prefixes



- 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

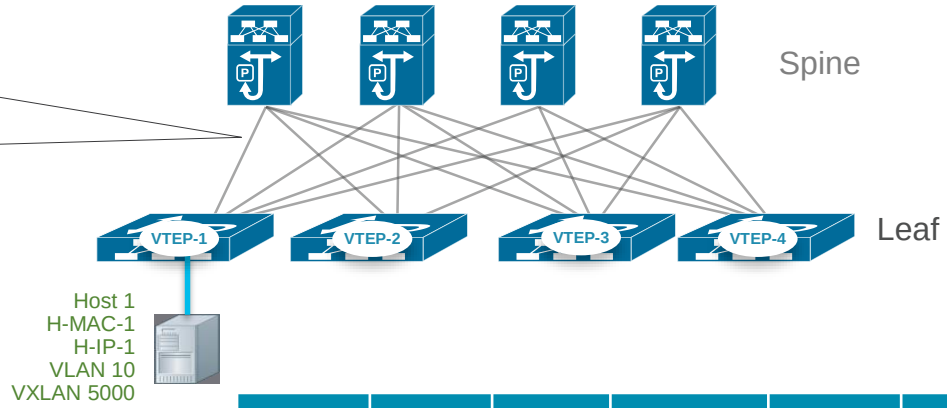
EVPN Control Plane -- Host Advertisement



EVPN Control Plane --- VM Mobility

NLRI:
- Host H-MAC-1, H-IP-1
- NVE VTEP-1
- VNI 5000

Ext. Community:
- Encapsulation: VXLAN
- Cost
- Sequence number :0



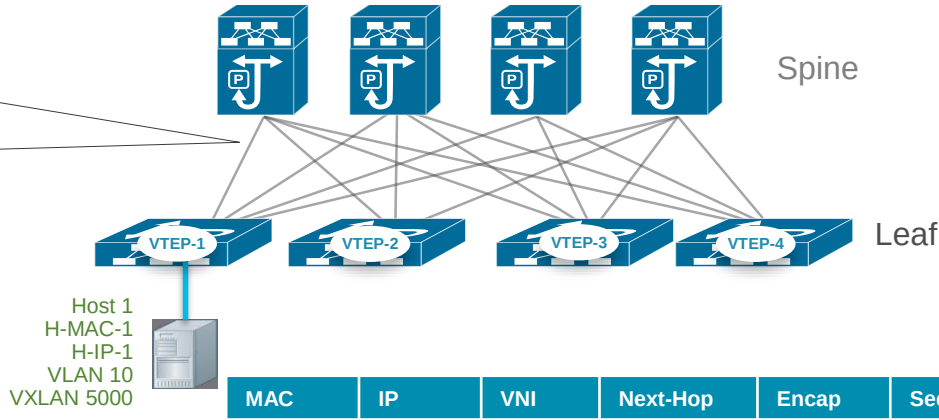
MAC	IP	VNI	Next-Hop	Encap	Seq#
H-MAC-1	H-IP-1	5000	VTEP-1	VXLAN	0

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

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



MAC	IP	VNI	Next-Hop	Encap	Seq#
H-MAC-1	H-IP-1	5000	VTEP-3	VXLAN	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

EVPN Control Plane --- ARP Suppression

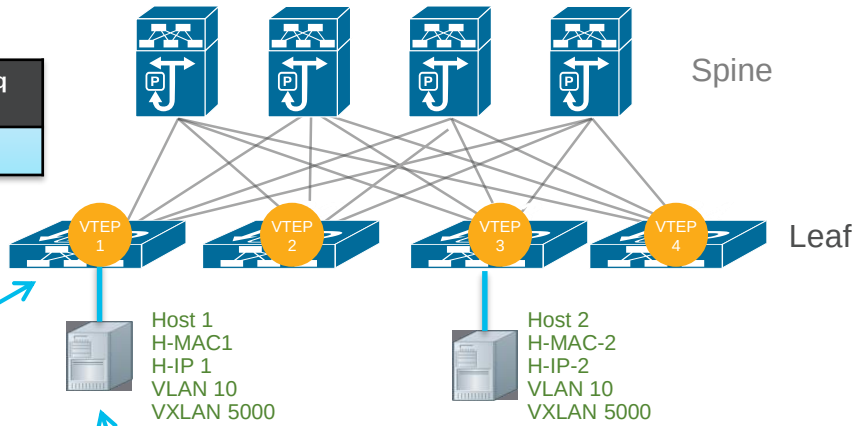
Minimize flood-&-learn behavior for host learning

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 behave 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

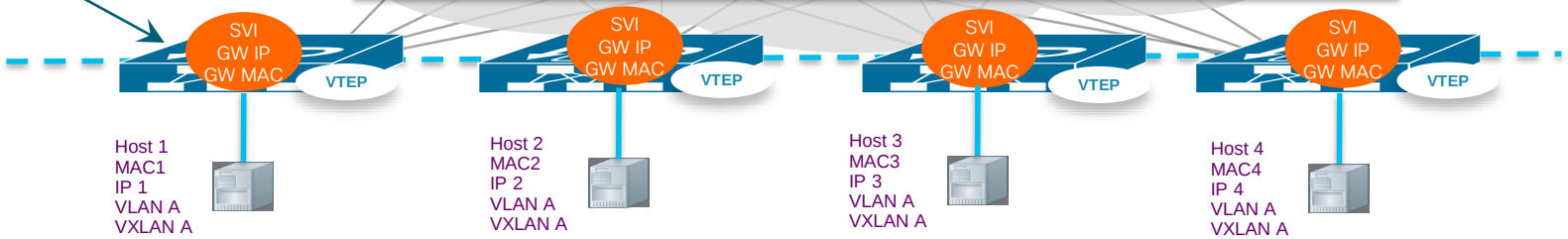


1 Host-1 sends ARP Request for H-IP-2

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.

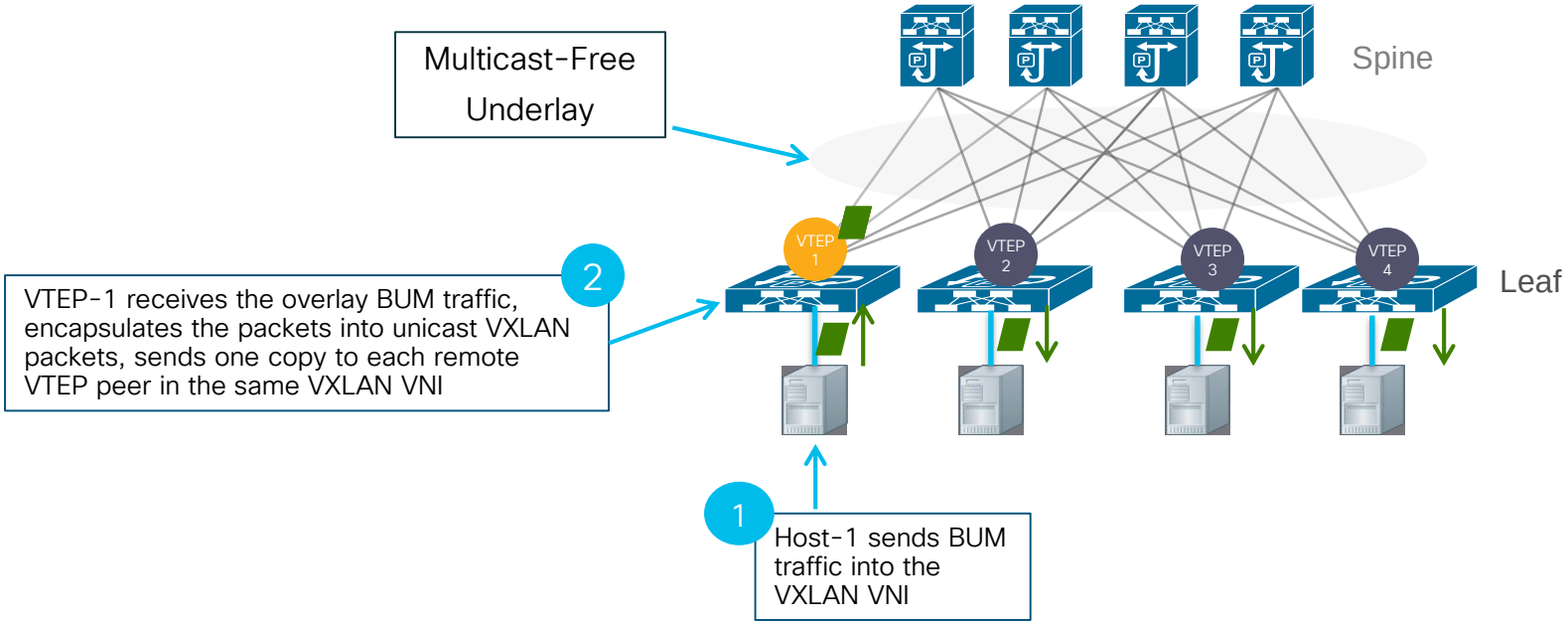
```
# VLAN to VNI mapping
vlan 200
 vn-segment 5200

# 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 Tenant-A
 ip address 20.0.0.1/24
 fabric forwarding mode anycast-gateway
```



EVPN Control Plane -- Head-end Replication

Head-end Replication (aka. Ingress replication):

Eliminate the need for underlay multicast to transport overlay BUM traffic



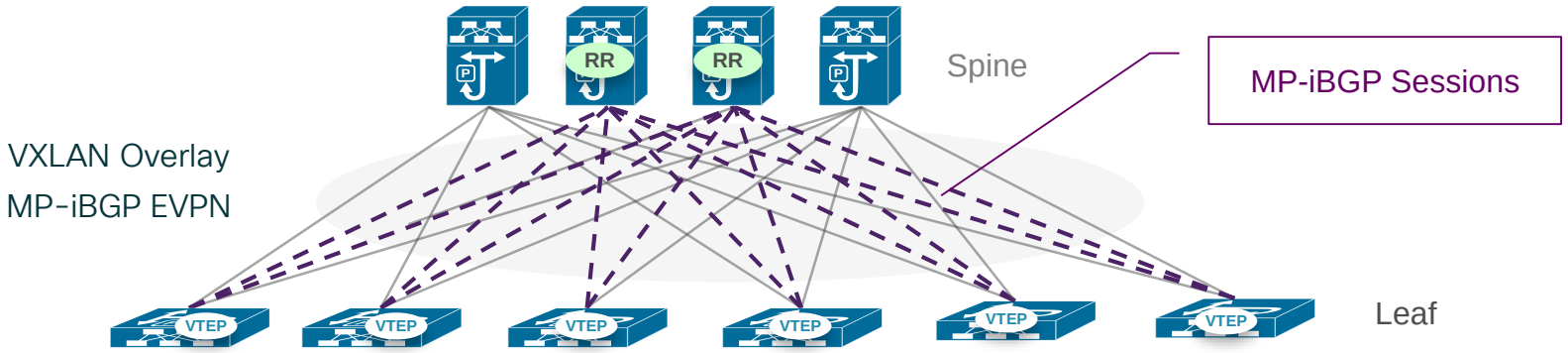
Functions of VXLAN/EVPN

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
Dynamic Ingress Replication	Unicast Alternative to Multicast underlay Dynamically discover remote peers for Ingress Replication

Agenda

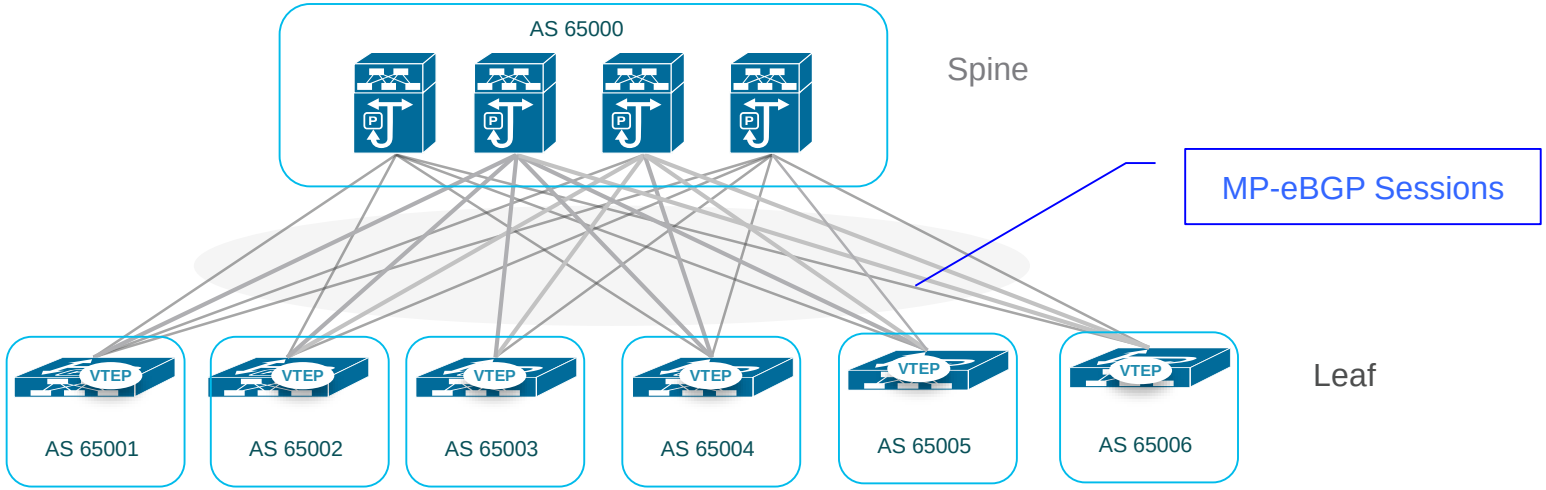
- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

VXLAN Fabric Design with MP-iBGP EVPN



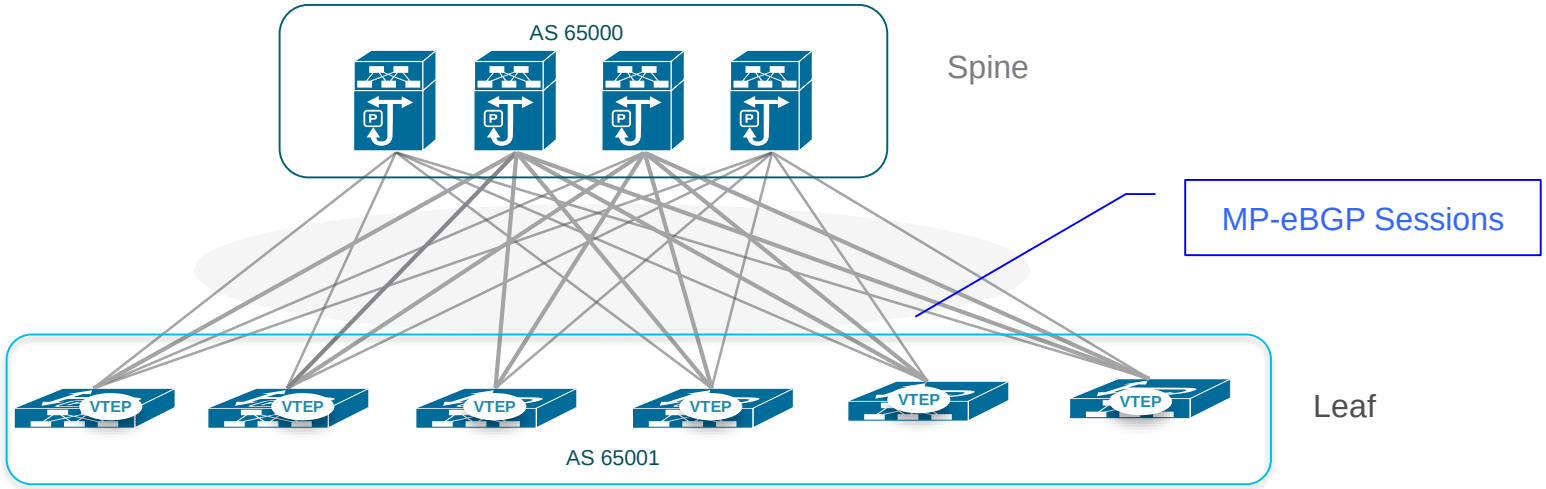
- VTEP Functions are on leaf layer
- Spine nodes are iBGP route reflector
- Spine nodes don’t need to be VTEP

VXLAN Fabric Design with MP-eBGP EVPN



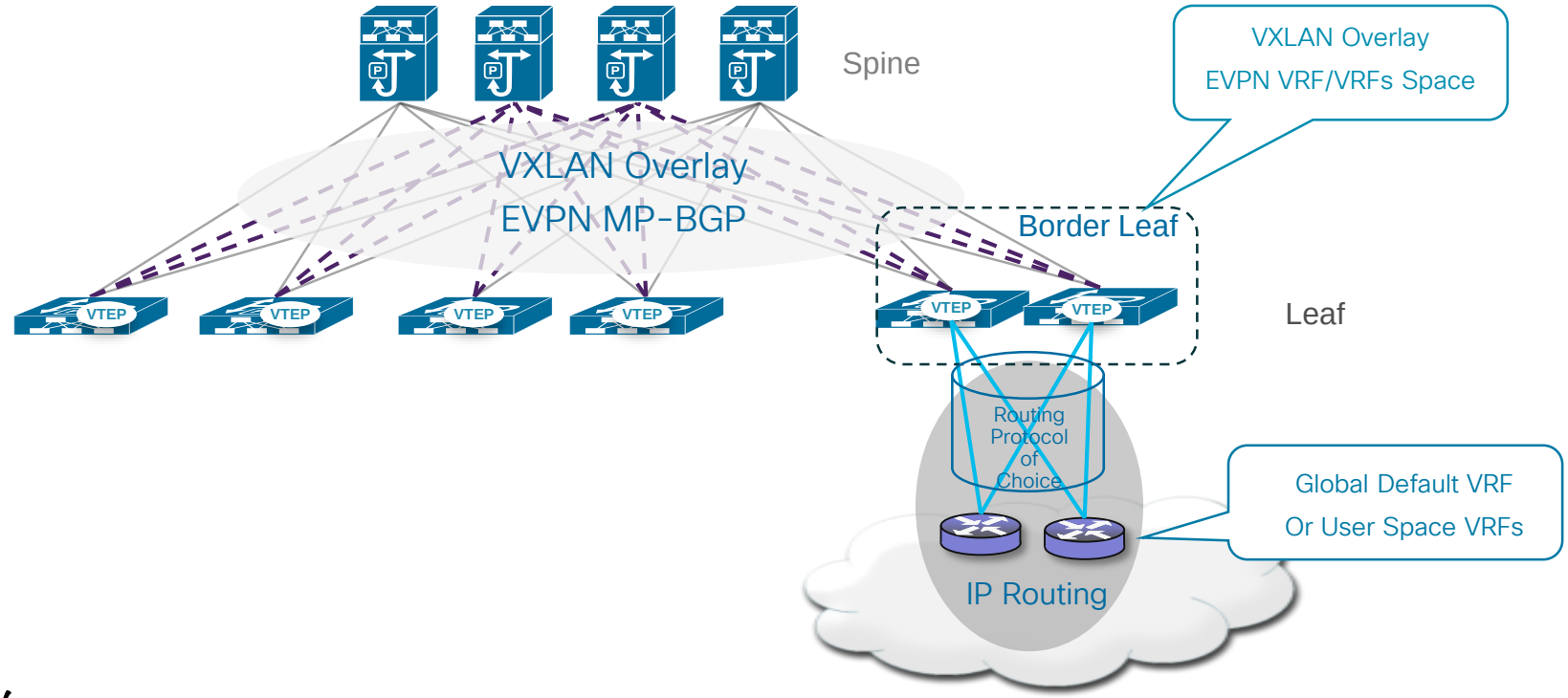
- VTEP Functions are on leaf layer
- Spine nodes are MP-eBGP Peers to VTEP leafs
- Spine nodes don’t need to be VTEP
- VTEP leafs can be in the same or different BGP AS’s

VXLAN Fabric Design with MP-eBGP EVPN



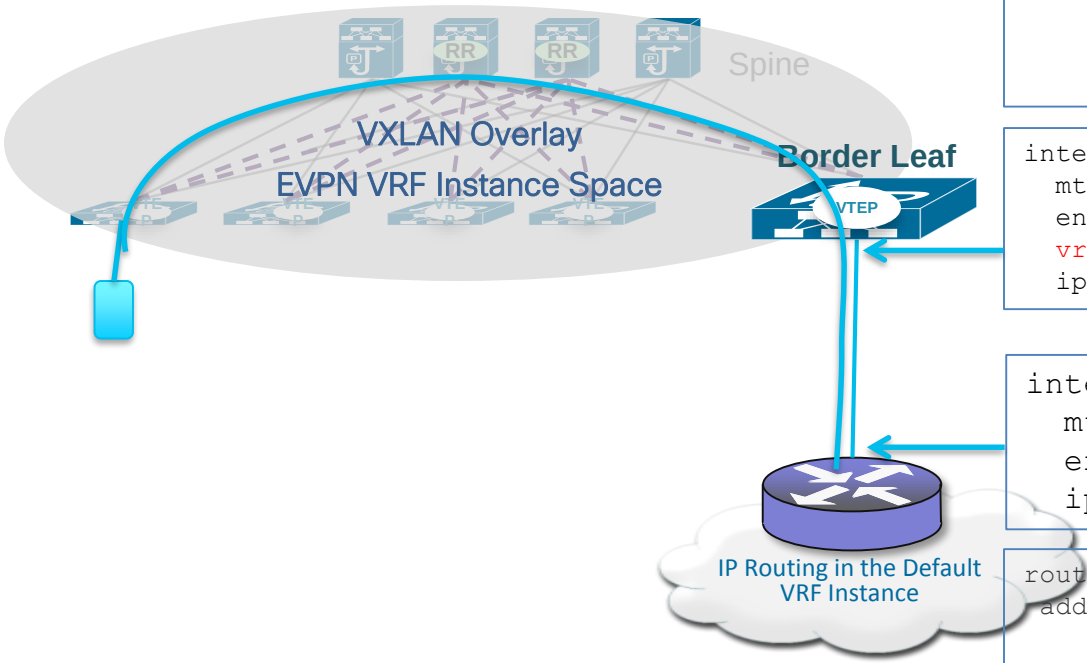
- VTEP Functions are on leaf layer
- Spine nodes are MP-eBGP Peers to VTEP leafs
- Spine nodes don’t need to be VTEP
- VTEP leafs can be in the same or different BGP AS’s

EVPN VXLAN Fabric External Routing



EVPN VXLAN External Routing with BGP Sample Configuration

FYI



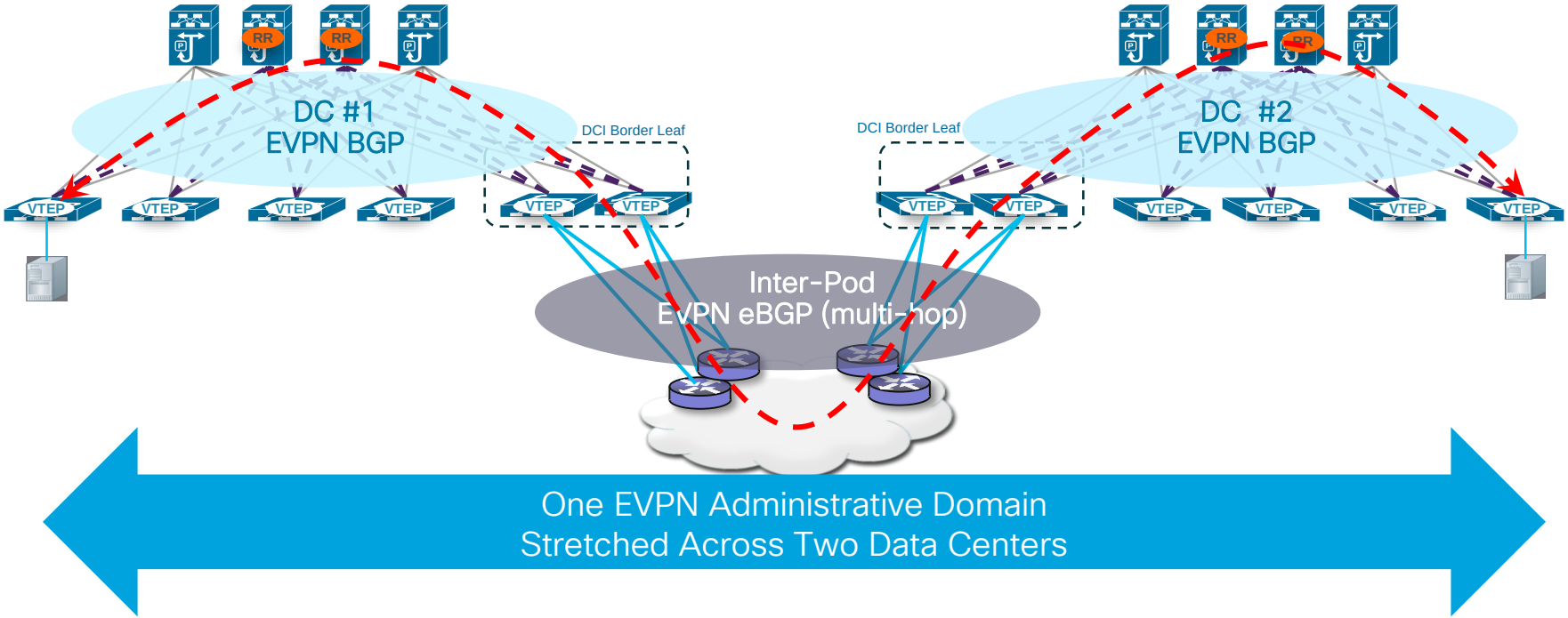

```
Router bgp 100
 vrf evpn-tenant-1
    address-family ipv4 unicast
      network 20.0.0.0/24
    neighbor 30.10.1.2 remote-as 200
      address-family ipv4 unicast
       prefix-list outbound-no-hosts out
```

```
interface Ethernet2/9.10
  mtu 9216
  encapsulation dot1q 10
  vrf member evpn-tenant-1
  ip address 30.10.1.1/30
```

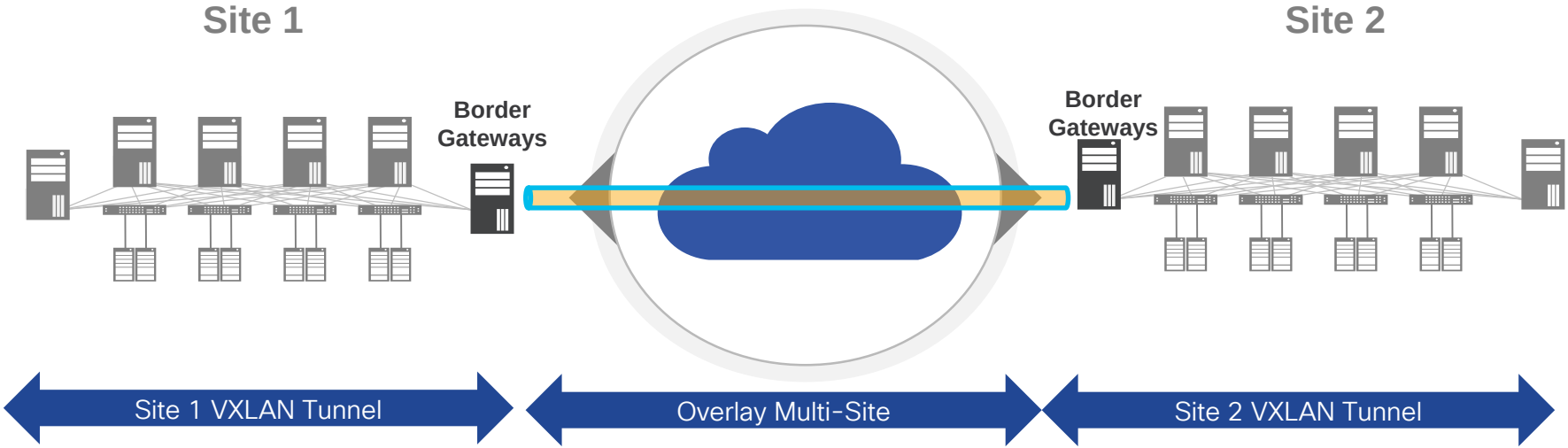
```
interface Ethernet1/50.10
  mtu 9216
  encapsulation dot1q 10
  ip address 30.10.1.2/30
```

```
router bgp 200
 address-family ipv4 unicast
    network 100.0.0.0/24
    network 100.0.1.0/24
  neighbor 30.10.1.1 remote-as 100
    address-family ipv4 unicast
```

EVPN Design for Multi-Pod



VXLAN EVPN Multi-Site



Scale through Hierarchical Forwarding

Fault Containment | Convergence independent of Network Size | Separate Admin Domains | Single Box

Agenda

- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

FYI

Building your VTEP (VXLAN Tunnel End-Point)

```
# Features & Globals
feature bgp
feature nv overlay
nv overlay evpn

# Spine (S1)

# Leaf (V1)
interface nve1
  source-interface loopback0
  host-reachability protocol bgp
```

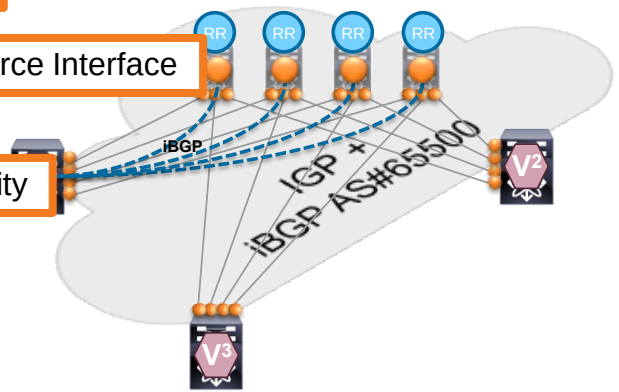
Enables VTEP (only required on Leaf or Border)

Enables EVPN Control-Plane in BGP

Configure the VTEP interface

Use a Loopback for Source Interface

Enable BGP for Host reachability



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

FYI

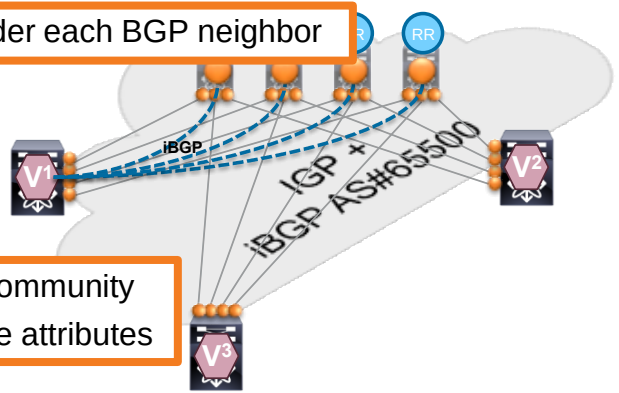
Building your EVPN MP-BGP Control-Plane

```
# Features & Globals
feature bgp
feature nv overlay
nv overlay evpn

# Spine (S1)
router bgp 65500
  router-id 10.10.10.1
  address-family ipv4 unicast
  address-family l2vpn evpn
  neighbor 10.10.10.10 remote-as 65500
    update-source loopback0
    address-family l2vpn evpn
      send-community both
      route-reflector-client

# 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
*
```

Enables EVPN Control-Plane in BGP

Activate L2VPN EVPN under each BGP neighbor

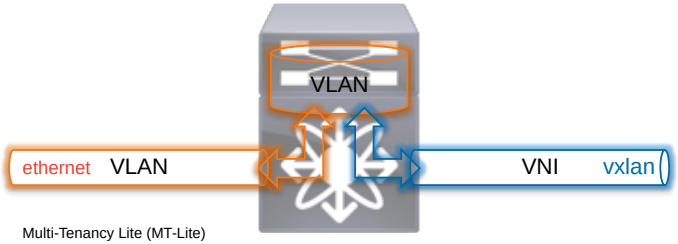
Send Extended BGP Community to distribute EVPN route attributes



FYI

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.



Multi-Tenancy Lite (MT-Lite)

```
# Features
feature vn-segment-vlan-based

# VLAN to VNI mapping (MT-Lite)
Vlan 10
 vn-segment 5010

# Activate Layer-2 VNI for EVPN
evpn
  vni 5010 l2
    rd auto
    route-target import auto
    route-target export auto

# 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
```

VLAN to Layer-2 VNI mapping

Enables EVPN Control-Plane for Layer-2 Services

Alternative is to use “ingress-replication protocol bgp”

Enables Layer-2 VNI on VTEP and suppress ARP

FYI

Distributed Anycast Gateway for Extended VLANs

- 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 with an identical anycast gateway virtual IP address

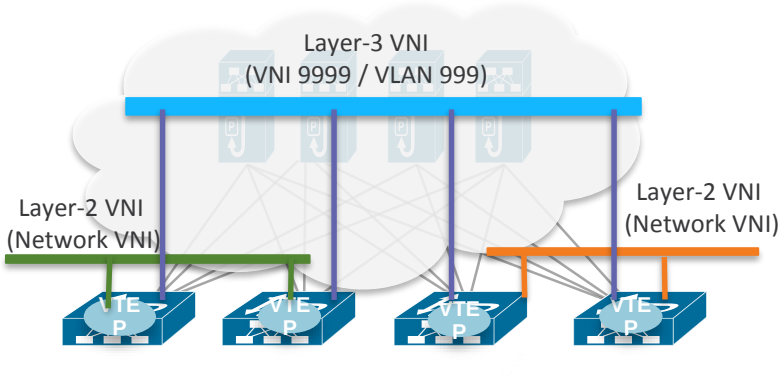
One gateway virtual MAC per VTEP

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 Tenant-A
 ip address 20.0.0.1/24
 fabric forwarding mode anycast-gateway
```

FYI

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

# Define SVI for VRF routing instance
interface Vlan999
  no shutdown
  mtu 9216
  vrf member VRF-A
  ip forward

# 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
```

VLAN to Layer-3 VNI mapping

VLAN to Layer-3 VNI mapping

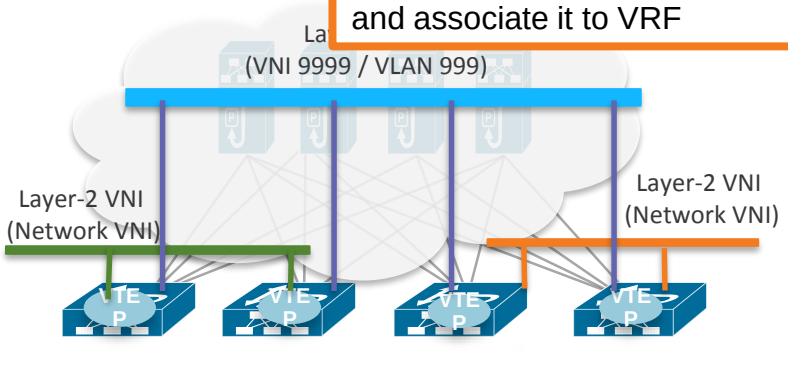
- ip forward required for prefix-based routing

VRF context definition

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

FYI

Routing in VXLAN – Configure the Routing



Enables Layer-3 VNI on VTEP and associate it to VRF

1:1 mapping between L3 VNI and tenant VRF

VRF/Tenant definition within Overlay Control-Plane

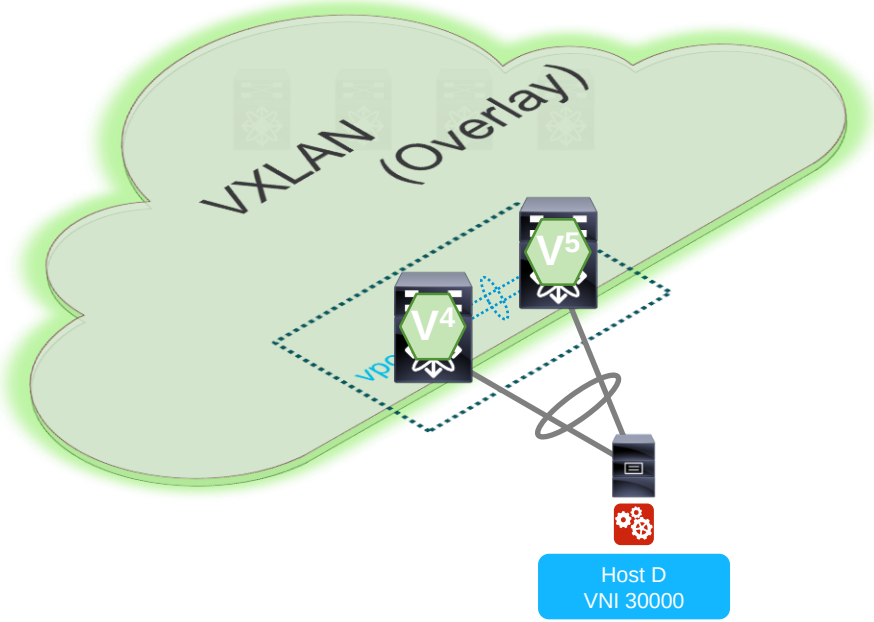
Configuration Example for VRF-A

```
# Activate Layer-3 VNI on VTEP
interface nve1
  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
```

FYI

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



FYI

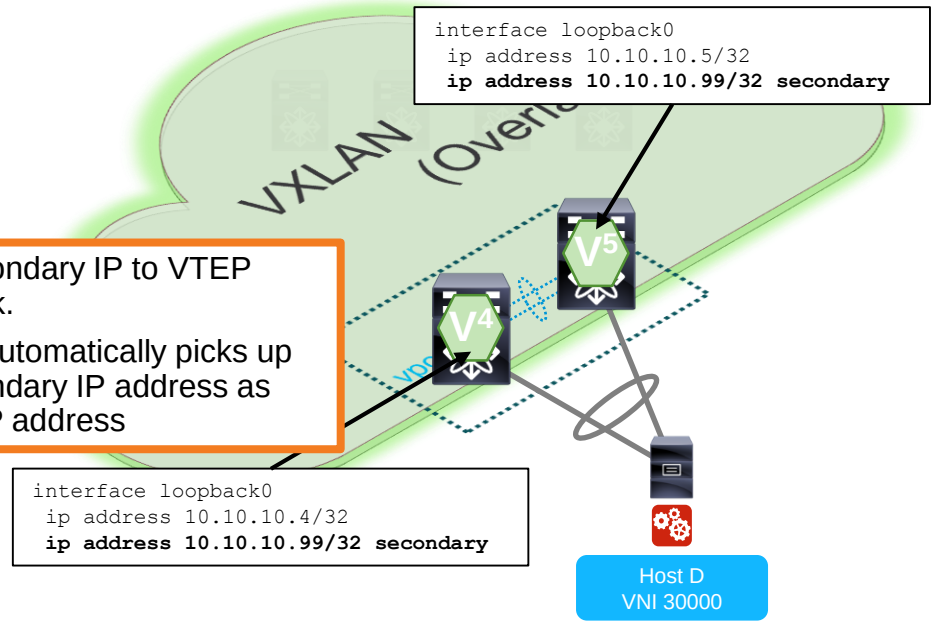
VXLAN Hardware Gateway Redundancy (vPC)

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



FYI

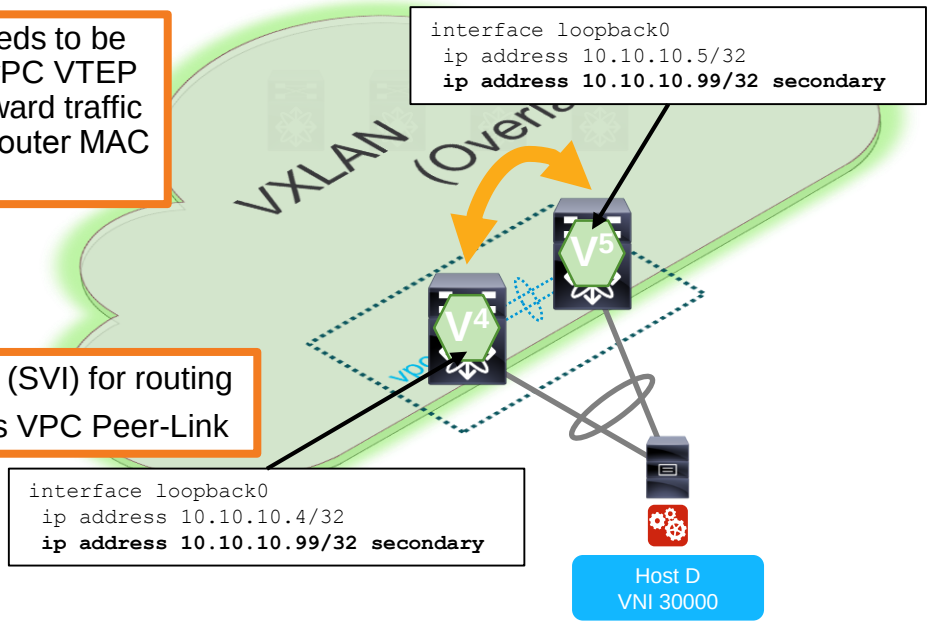
VXLAN Hardware Gateway Redundancy (vPC)

vPC VTEP Configuration Example

```
# VPC Domain Configuration
vpc domain 99
  peer-switch
  peer-keepalive destination V4-mgmt sou
  peer-gateway
  ip arp synchronize

# 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
```

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

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



FYI

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 to by default. Need to change the policy to next-hop unchanged

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
```

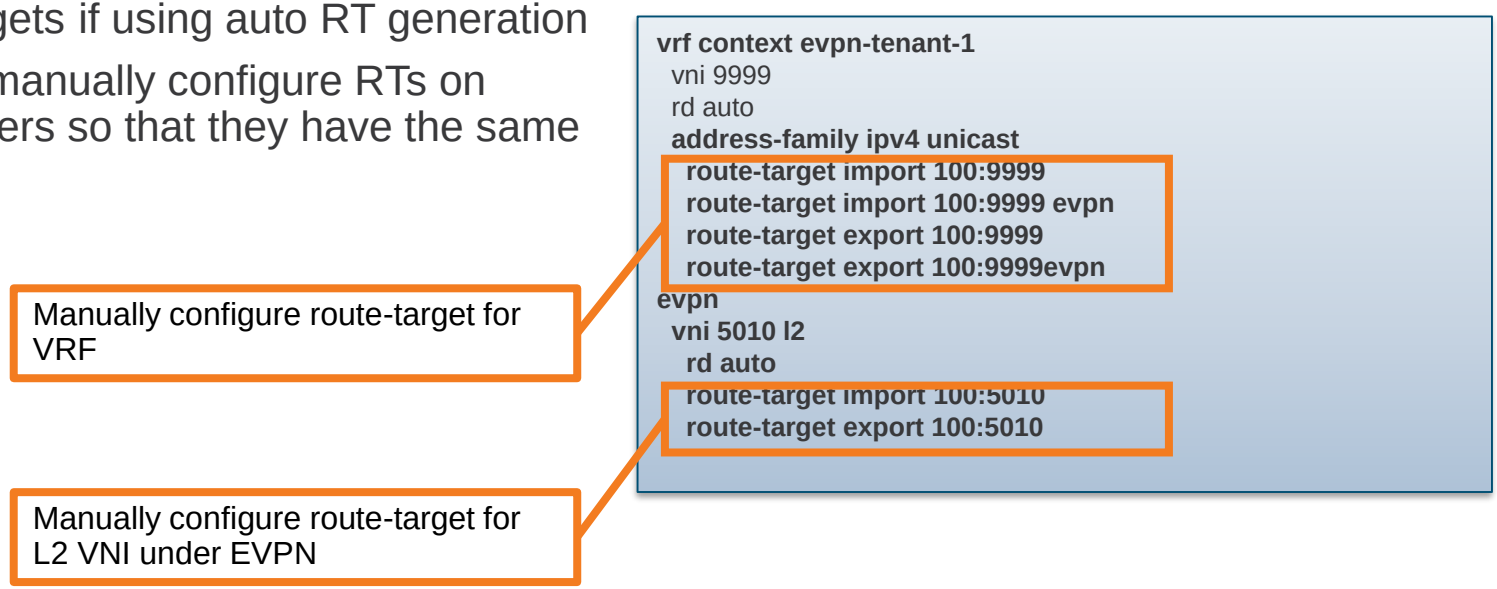
Set next-hop policy not to change the next-hop attribute

FYI

eBGP EVPN Configuration(2)

Manually configure import/export route-target

- With eBPG, 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



More Learning on VXLAN EVPN….

BRKDCT-3378: Building DataCenter networks with VXLAN BGP-*EVPN (Lukas Krattiger)*

BRKDCN-2304: L4-L7 Service Integration in Multi-Tenant VXLAN *EVPN* Data Center Fabrics (Matthias Wessendorf)

BRKSPG-2030: VXLAN troubleshooting and debugging in a mixed Service Provider - Data Center network (Nikolay Karpyshev, Ivan Shirshin)

BRKDCN-3505: Extending VXLAN Fabrics across geographically dispersed sites (Yves Louis)

TECDCT-2181: Deployment Considerations for Interconnecting Distributed Virtual Data Centers (Yves Louis, Victor Moreno)

LTRDCT-2224: Enhancing VXLAN/*EVPN* Fabrics with LISP (Lukas Krattiger, satish kondalam)

Agenda

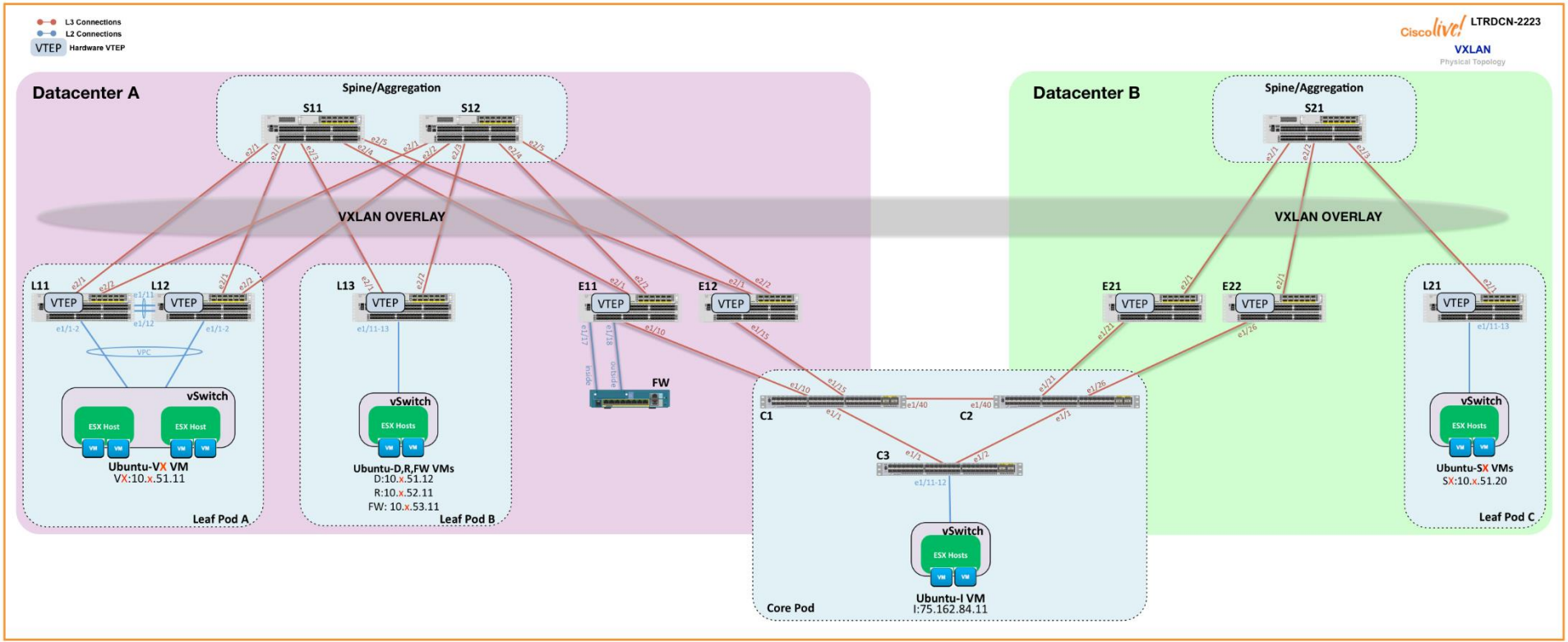
- VxLAN Overview
- Flood-&-Learn VXLAN
- VXLAN with MP-BGP EVPN Control Plane
- VXLAN Design Options
- MP-BGP EVPN VXLAN Configuration
- Lab Introduction

Lab Overview

Module Details

- **Module 1 – Network Based Overlay [DC1]**: In this module, students will configure a network based overlay with Nexus 9000 switches and use them as VTEPs. Students will also learn how to extend the anycast gateway using BGP EVPN.
- **Module 2 – FW-Security Zone [DC1]**: In this module, students will create a secure zone by placing a transparent mode FW between the VXLAN fabric and core pod.
- **Module 3 – MultiPOD [DC1 and DC2]:** In this module, students will stretch the VXLAN fabric from DC 1 to DC2 and extend VLANs from one DC to the other.

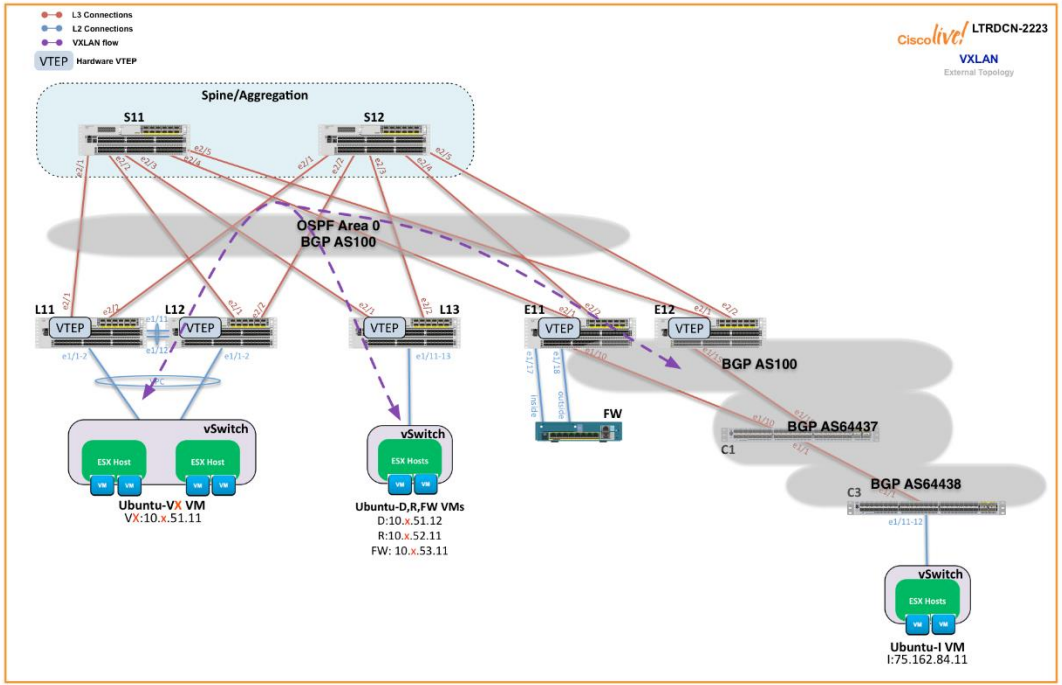
Lab Topology



Network Overview - DC1

Using BGP EVPN for Control Plane

- Using OSPF for underlay
- Using IBGP for overlay - Route Reflectors are on Spines S11 and S12
- Using VXLAN EVPN for control plane
- VXLAN Ingress Replication for the BUM Traffic
- Each student has their own VRF that will be representative of multi-tenancy.

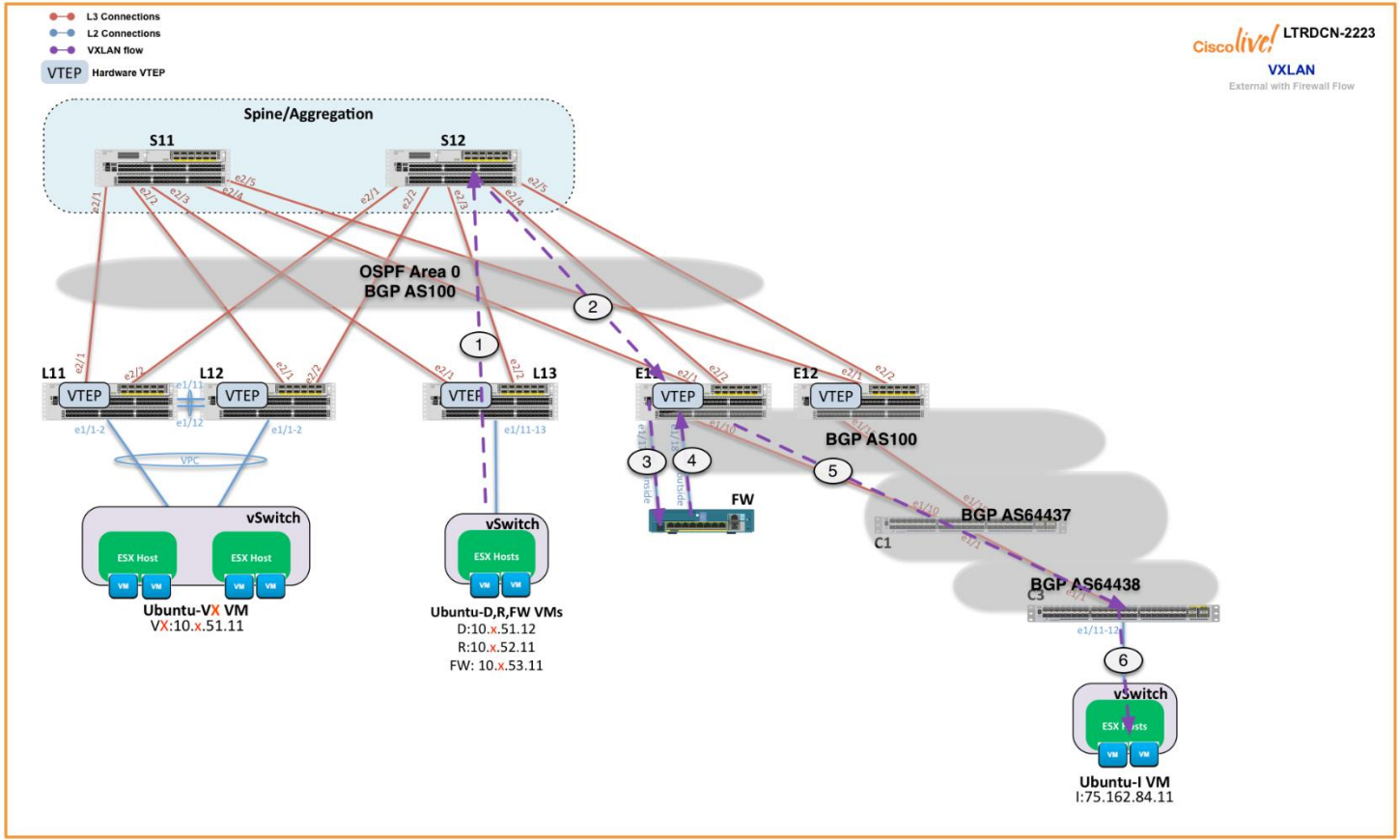


Security Module

Creating a Security Zone via VXLAN

- The Hosts in the Secured Zone are in VXLAN X53.
- Transparent FW is attached to Edge E11.
- No redundancy in this portion of the lab for simplicity.

Firewall Usecase Flow



Network Overview – DC2

Using EBGP EVPN for Underlay/Overlay

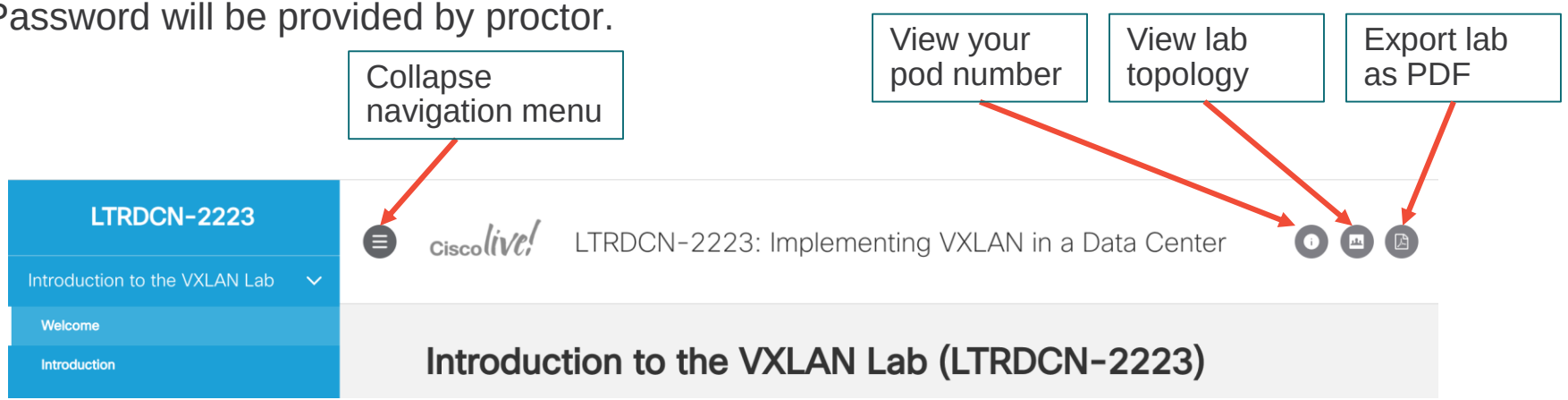
- Using EBGP for Underlay/Overlay
- Ingress Replication for the BUM Traffic.
- VLAN X51 will be stretched from DC1 to DC2.
- Each student has their own VRF that will be representative of multi-tenancy.

Manual Overview (1)

- Manual available at

  http://cs.co/ltrdcn-2223

- Password will be provided by proctor.



- Navigate either using the navigation menu on the left or the back/forward buttons at the bottom of the page.
- Only type commands that are shown in a box. Commands shown under “Configuration Sample” are not meant to be typed into the devices.

Manual Overview (2)

- Manual available at

  http://cs.co/ltrdcn-2223

- RDP Server: vxlanlab.ciscolive.com:3390
- Username: vxlan\PODxuser
- Password:

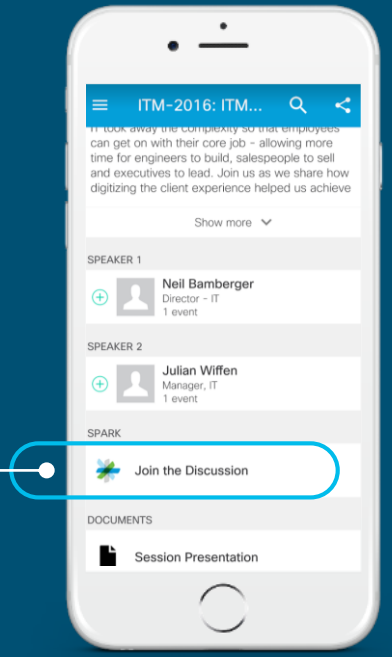
Cisco Spark

Questions?

Use Cisco Spark to communicate with the speaker after the session

How

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

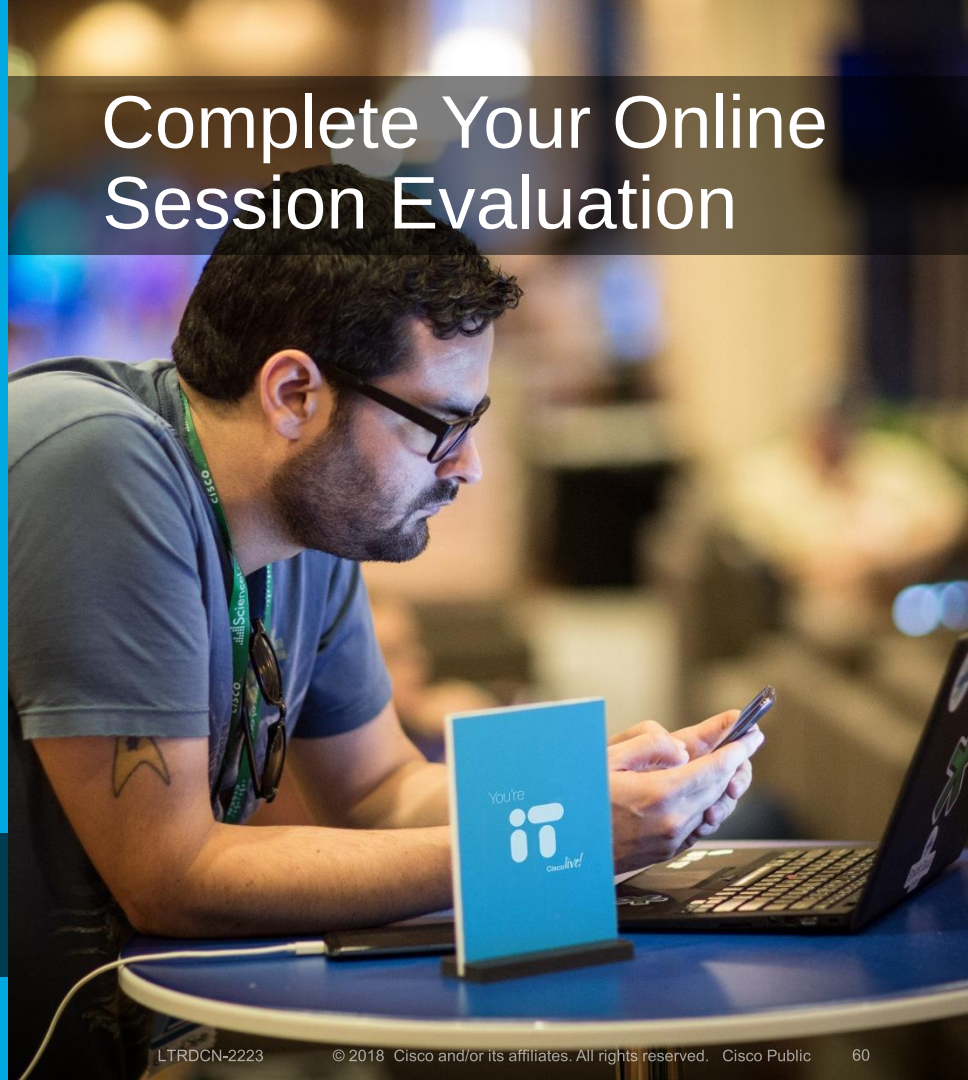


cs.co/ciscolivebot#LTRDCN-2223

Complete Your Online Session Evaluation

- Please complete your Online Session Evaluations after each session
- Complete 4 Session Evaluations & the Overall Conference Evaluation (available from Thursday) to receive your Cisco Live T-shirt
- All surveys can be completed via the Cisco Live Mobile App or the Communication Stations

Don’t forget: Cisco Live sessions will be available for viewing on-demand after the event at www.ciscolive.com/global/on-demand-library/.



Continue Your Education

- Demos in the Cisco campus
- Walk-in Self-Paced Labs
- Tech Circle
- Meet the Engineer 1:1 meetings
- Related sessions

Thank you


You're
Cisco live!