

CISCO *Live!*

GO BEYOND

#CiscoLiveAPJC



Introduction to VXLAN:

The Future Path of Your Data Center

Richard Licon, Principal Technical Marketing Engineer
BRKDCN-1621

Cisco Webex App

Questions?

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

How

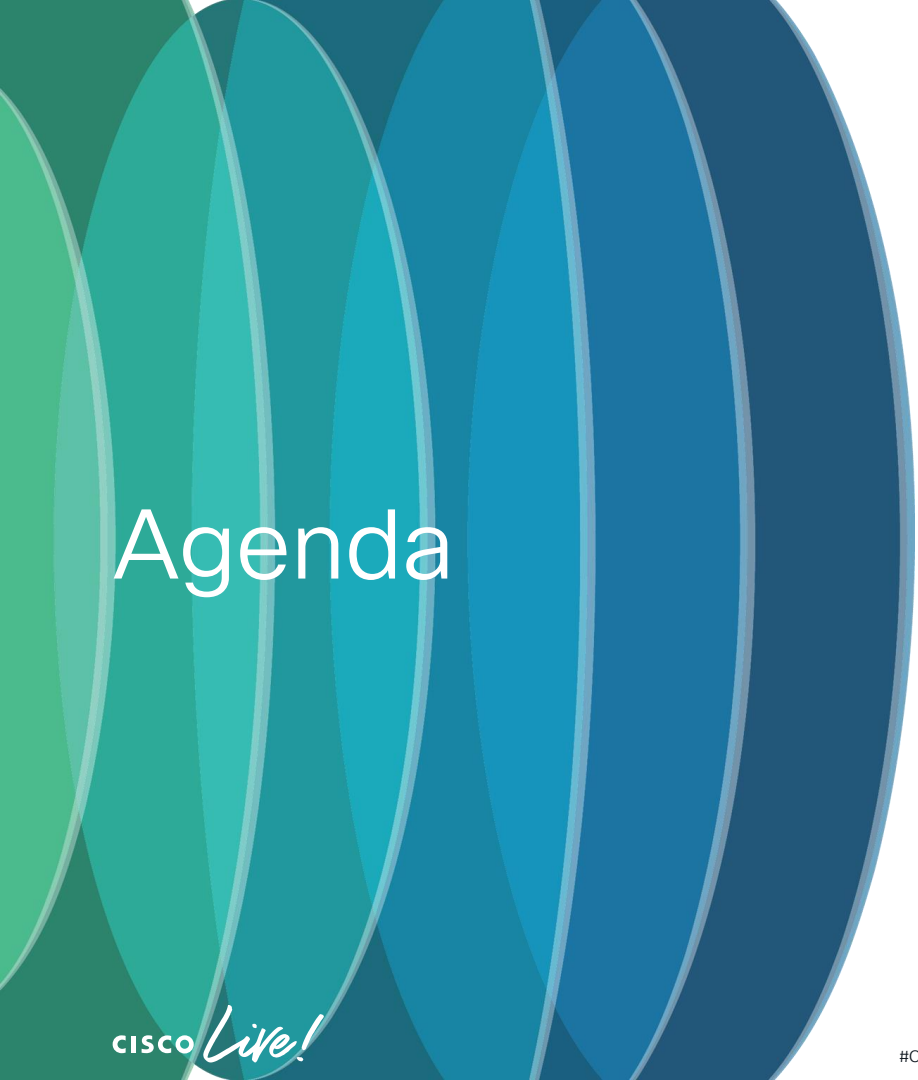
- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until November 15, 2024.



<https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-1621>





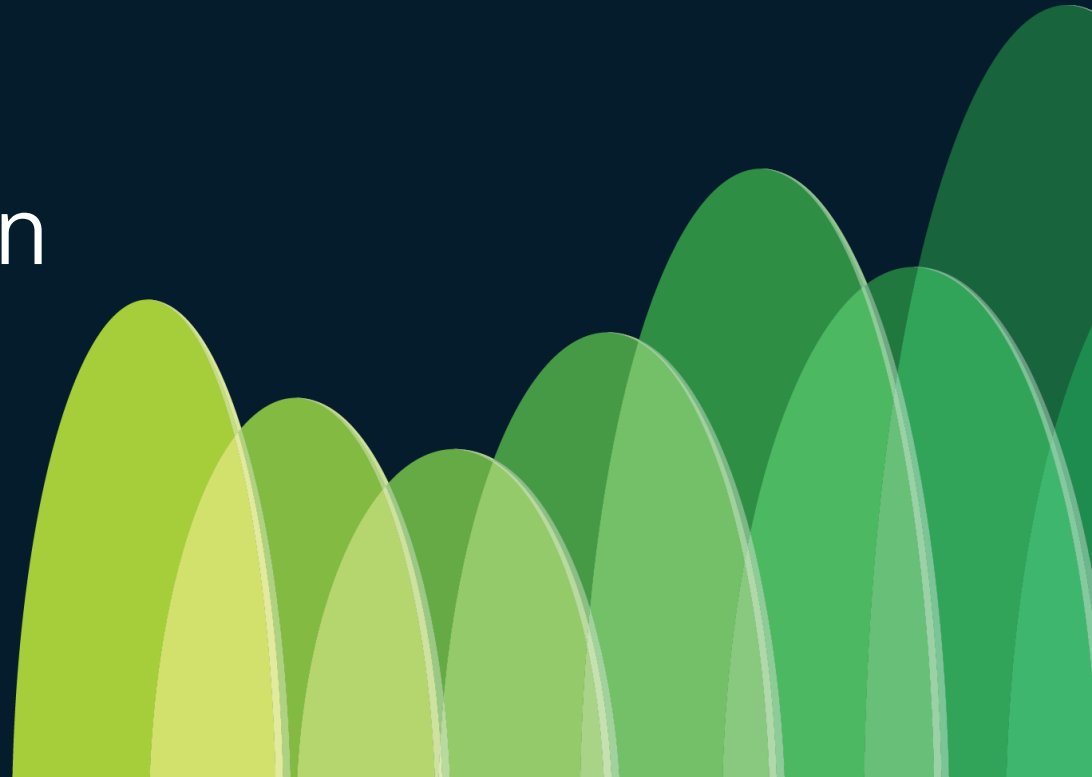
Agenda

- Evolution of the Data Center
- The Need for network overlays
- What is VXLAN?
- The Need for a Control Plane
- Packet Walks
- Conclusion

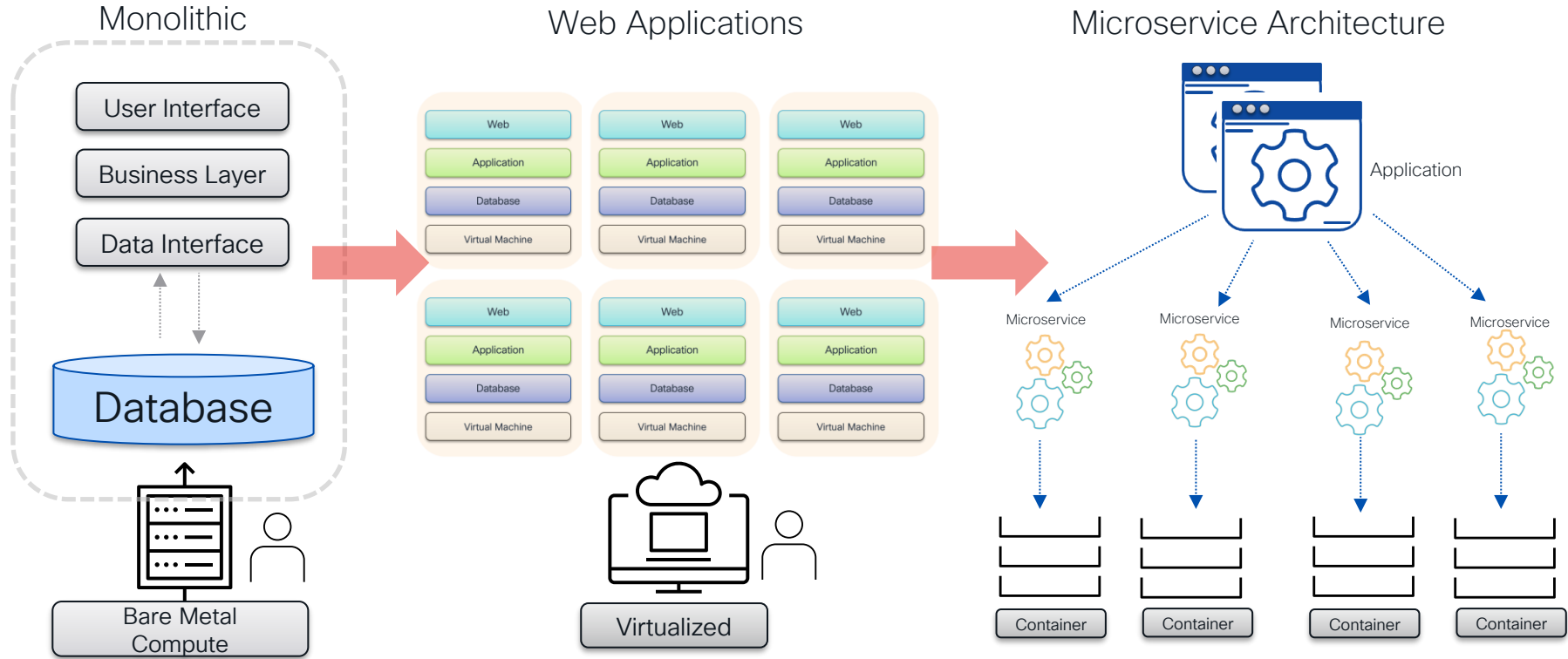
Session Abstract

- The need for flexibility in the data center opens the door to network overlays. Several standards have been proposed and implemented to enable virtualized networks to be layered over a physical network. The data center network needs to be flexible enough to support workloads that can move from one host to another, and for new services to be deployed rapidly. VXLAN is one such technology that provides massively scalable virtual network overlays on top of existing IP infrastructures. Let's learn more about VXLAN and delve deeper into it.
- This session aims at covering what is VXLAN, how it really works, and what problems does it solve. This session will have configuration examples of how we enable a VXLAN Fabric. No prior knowledge of VXLAN is required. A basic understanding of Unicast and Multicast routing protocol is needed along with an understanding of MP BGP

Evolution of the Application Architecture



Evolution of the Application Architectures

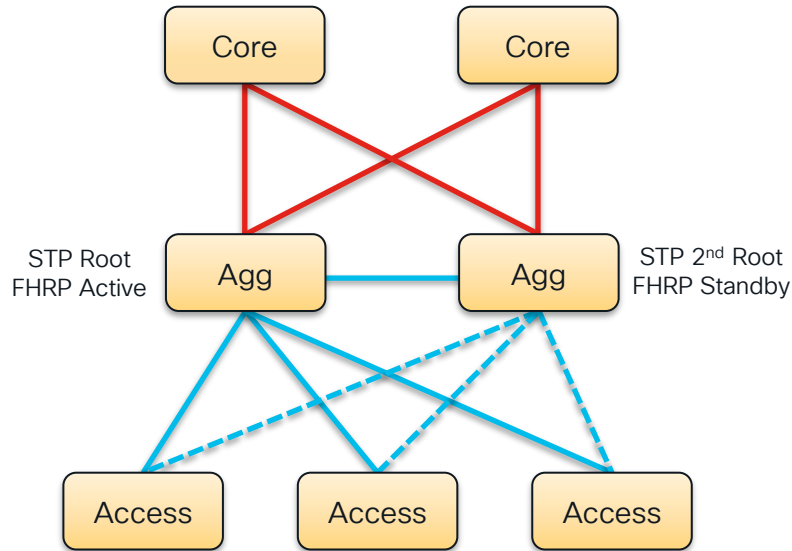


Evolution of the Data Center Network

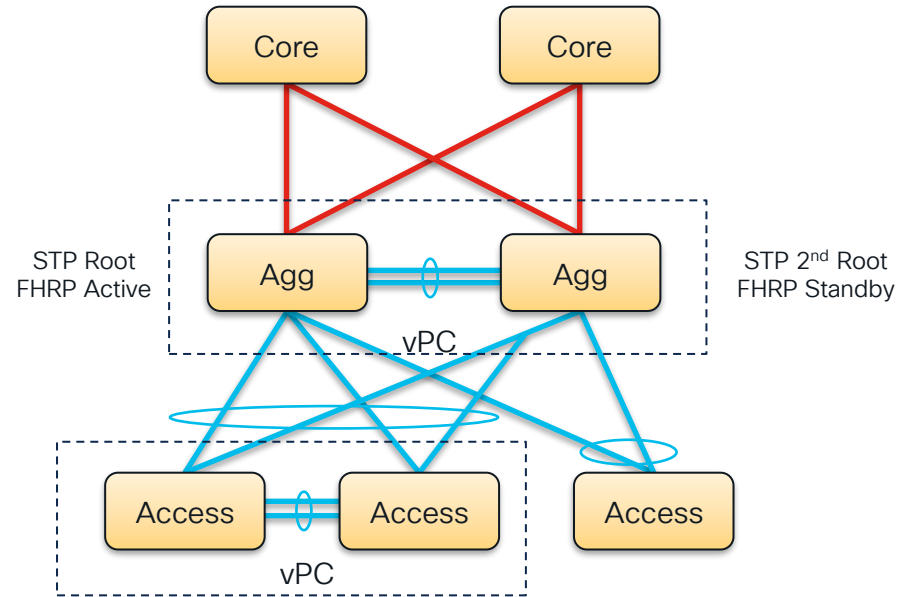


Evolution of the Data Center Network

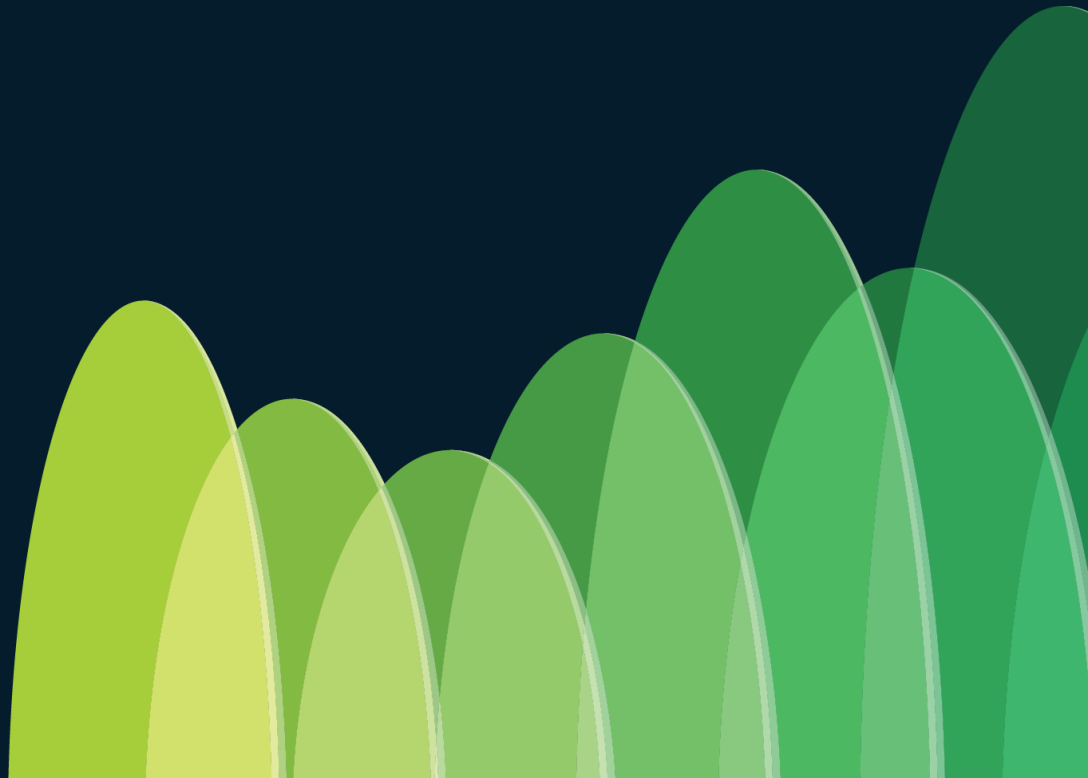
Classic Spanning-Tree



vPC and Spanning-Tree

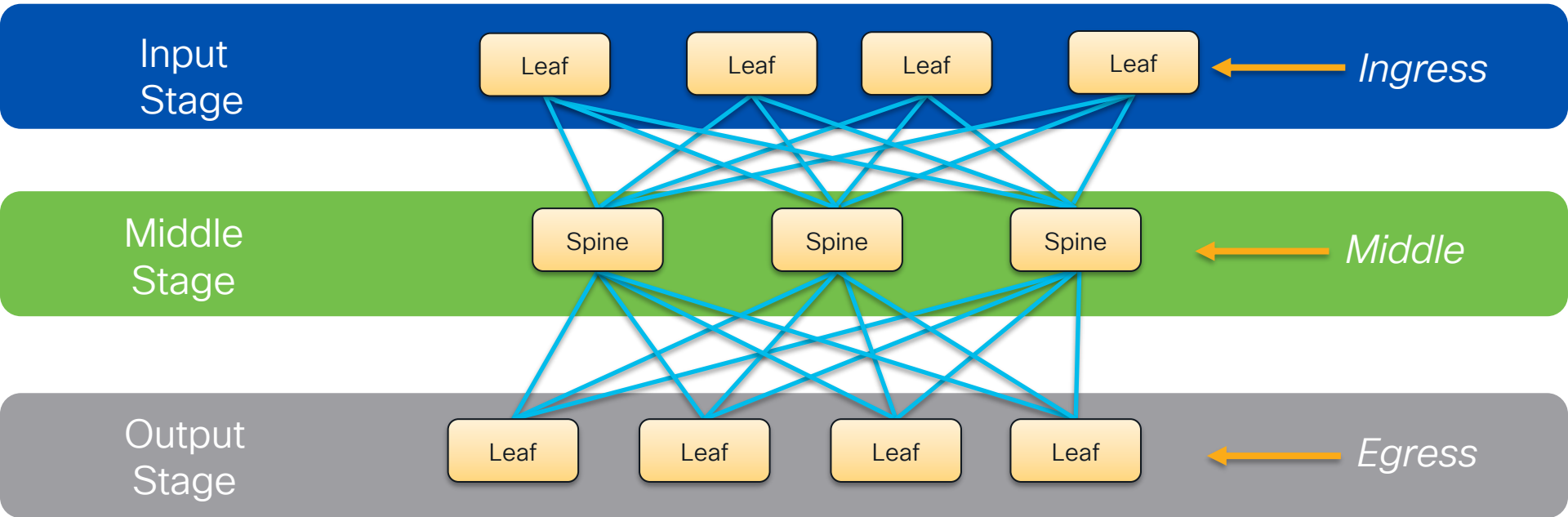


Evolution of the Data Center Network Architecture: Clos Networks



Evolution of the Data Center Network

3-Stage Clos Fabric

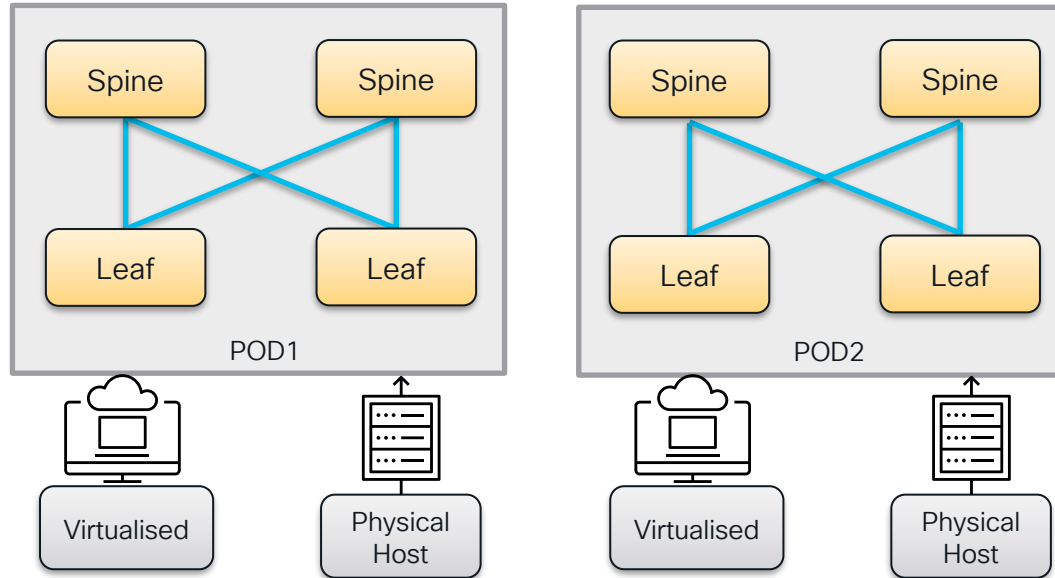


Invented by Edson Erwin in 1938

March 1953: A Study of Non-Blocking Switching Networks. (Clos, Charles)

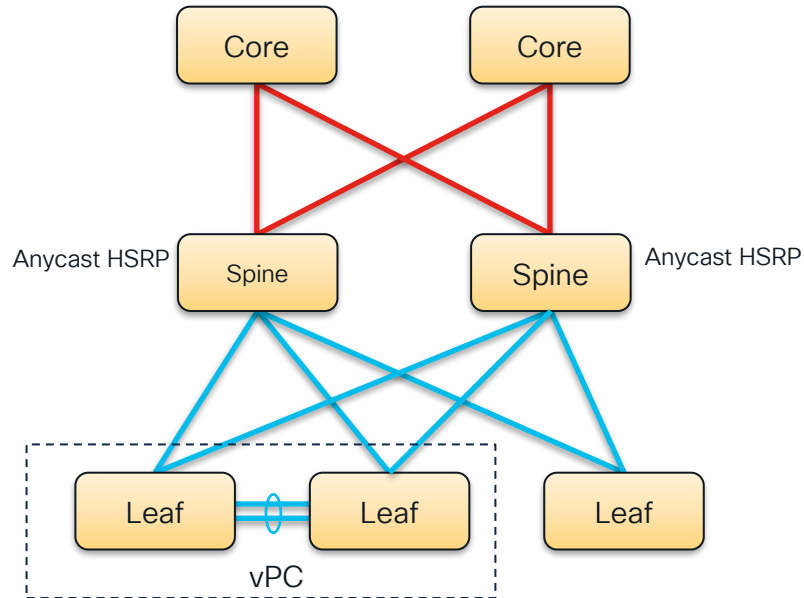
Evolution of the Data Center Network

3-Stage Fabric

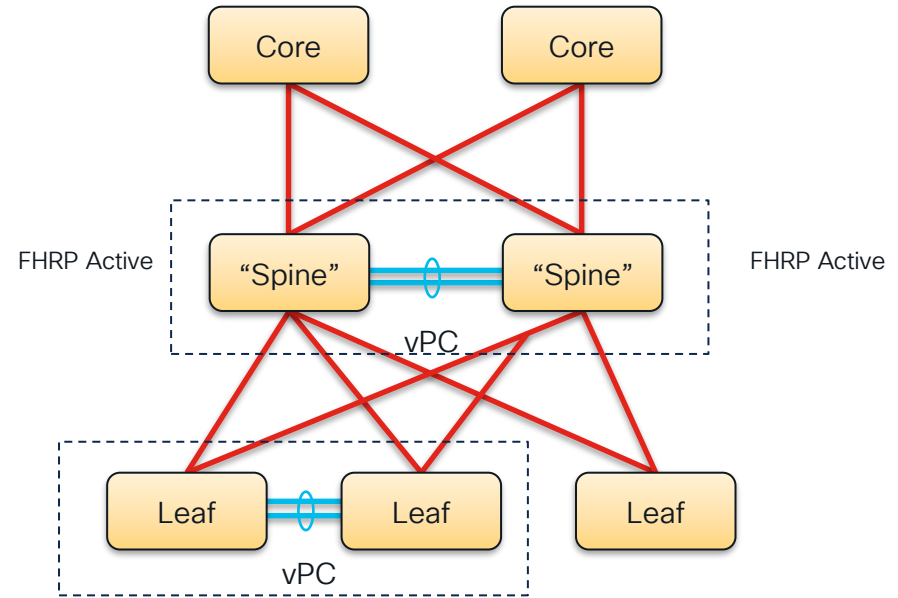


Evolution of the Data Center Network

FabricPath (MAC-in-MAC)



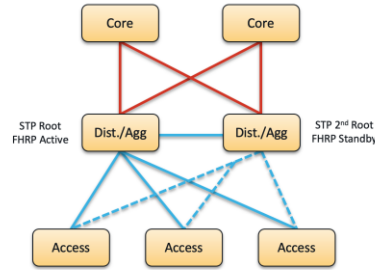
VXLAN Flood & Learn (MAC-in-IP)



Data Center Network Challenges

Legacy Architectures

Classic Spanning-Tree

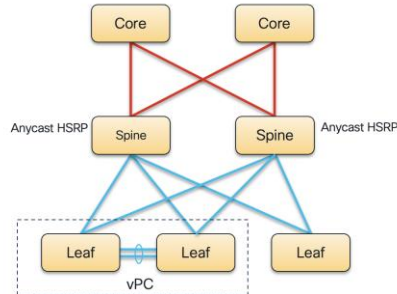


Convergence Time:
STP high convergence time
TCN MAC Flush

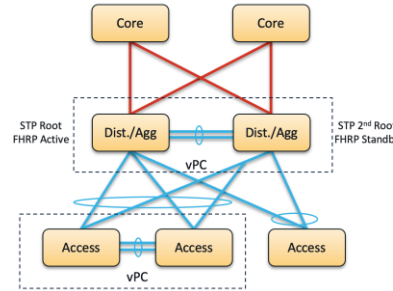
Unused Links:
STP Redundant links in
blocked state

Rigid Network Services:
L4-L7 Services placed at
Distribution Layer

FabricPath (MAC-in-MAC)



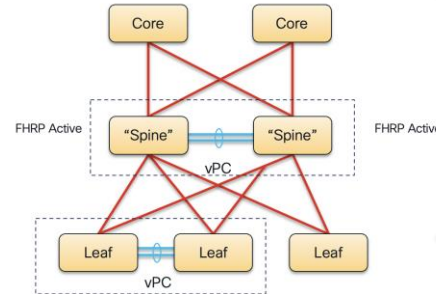
vPC and Spanning-Tree



Suboptimal Forwarding:
Path defined by
Root Switch or FHRP Active

OpEx/CapEx:
Expensive to upgrade
Scale up with large Chassis

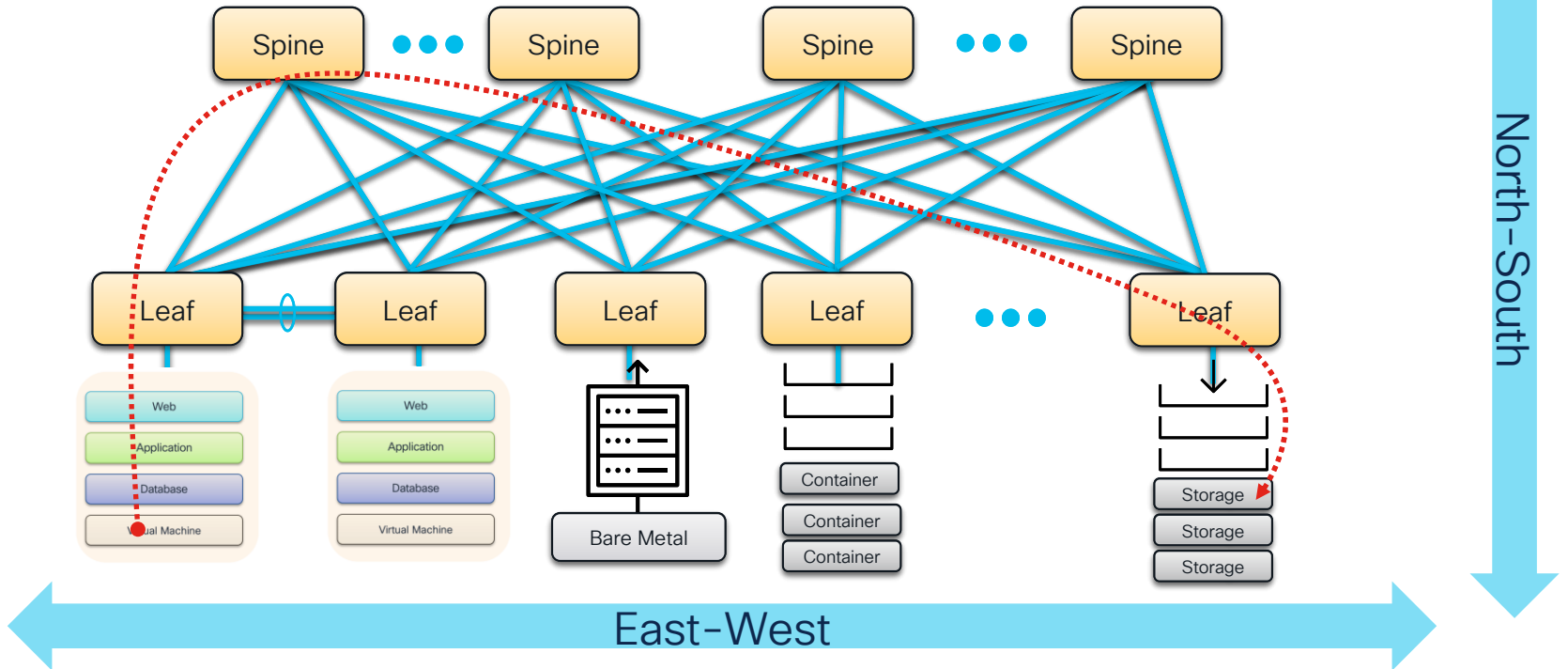
VXLAN Flood & Learn (MAC-in-IP)



Limited Scale:
No Control Plane
Limited Workload Mobility

Evolution of the Data Center Network

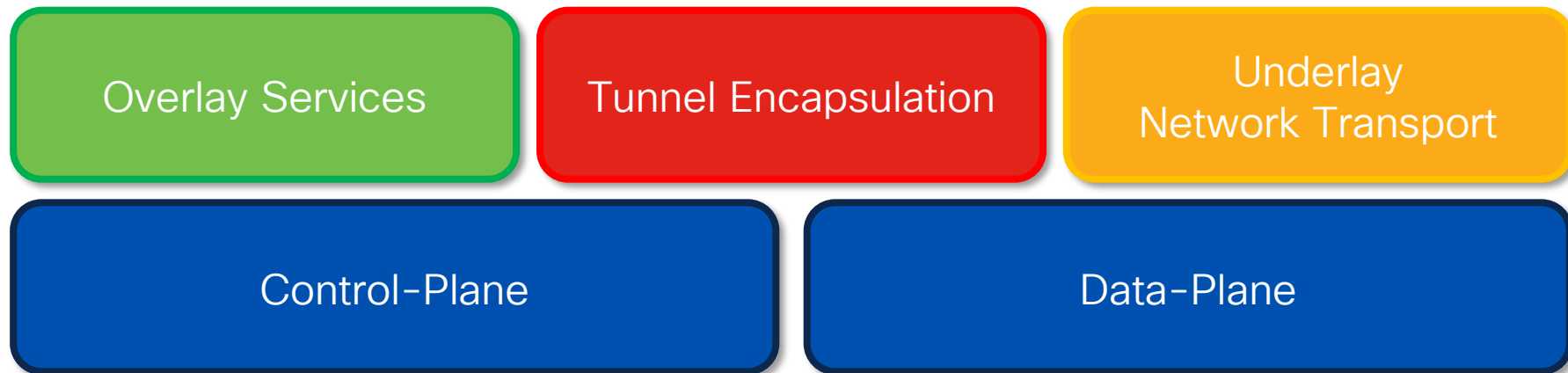
The need for network overlays



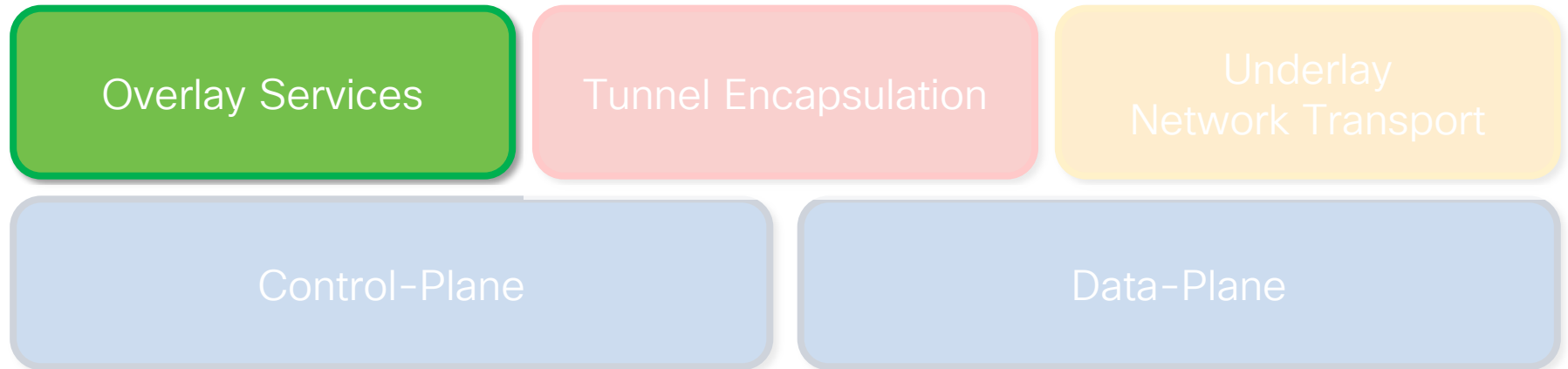
The Need for Network Overlays



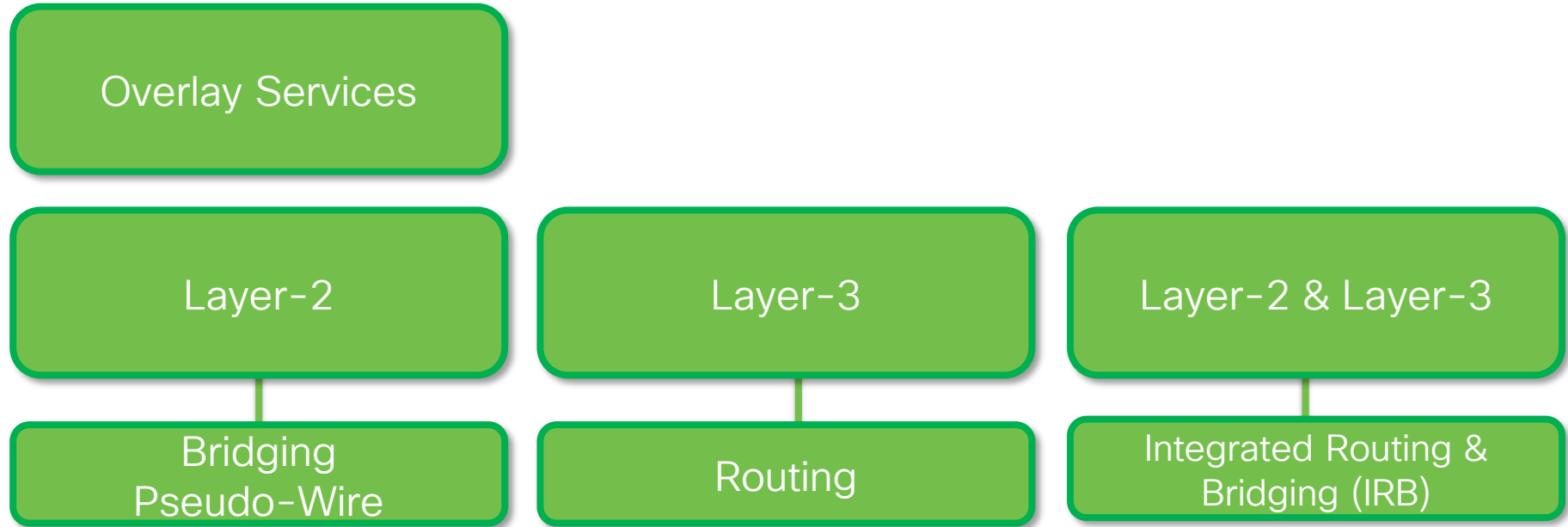
Overlay Taxonomy



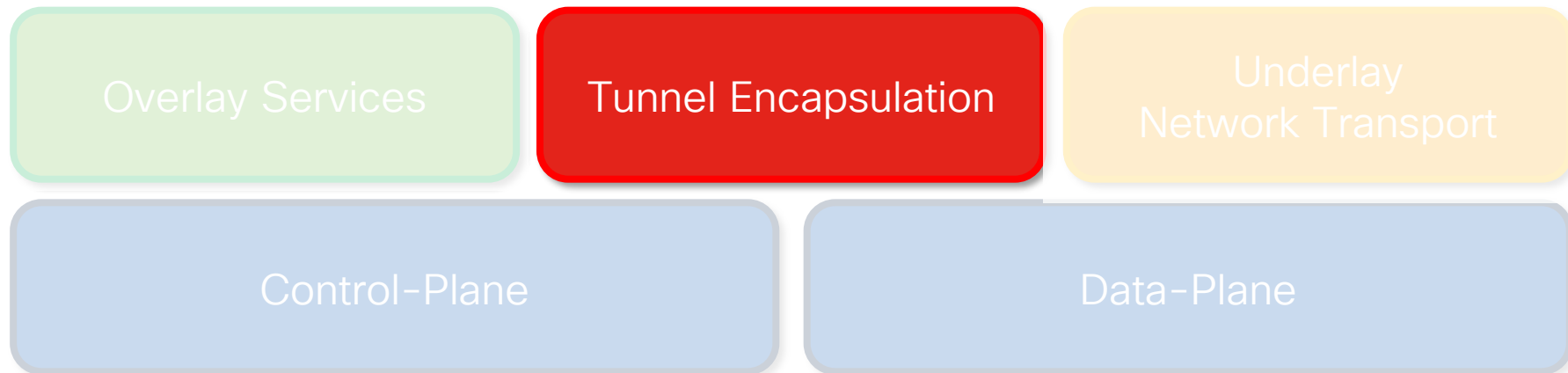
Overlay Services



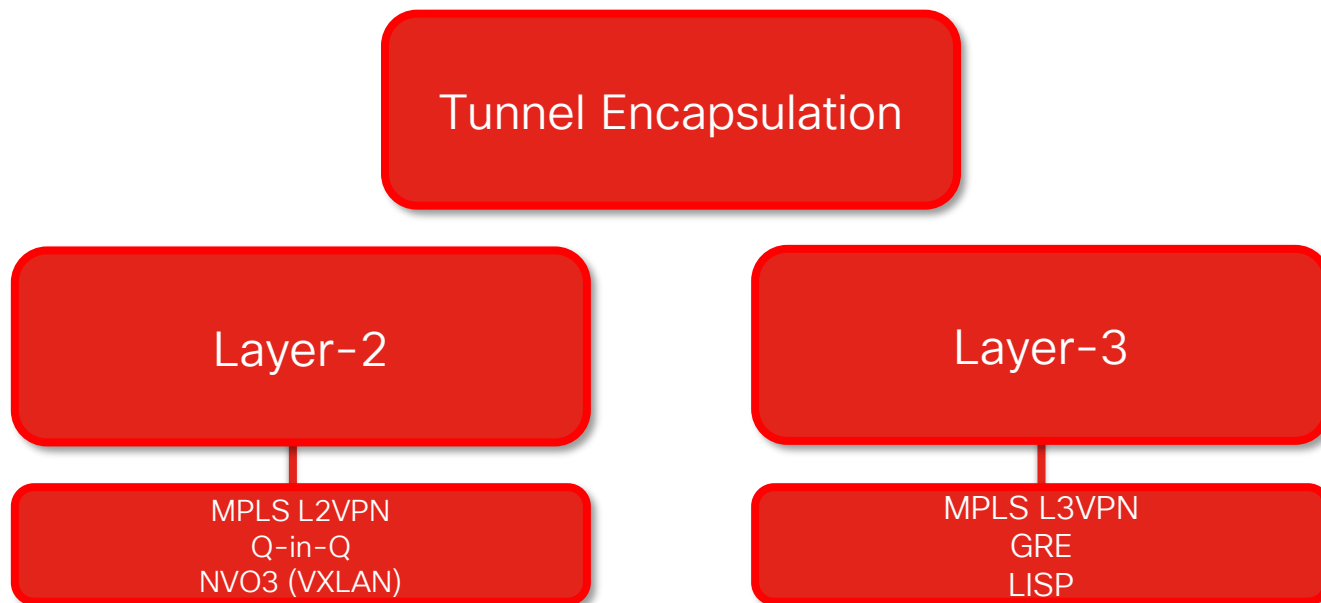
Overlay Services



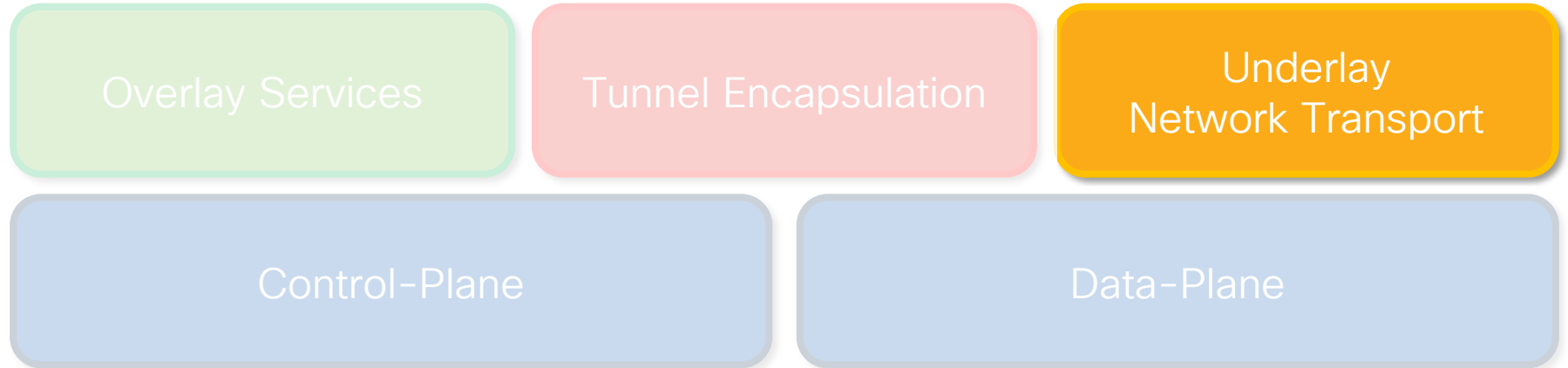
Tunnel Encapsulation



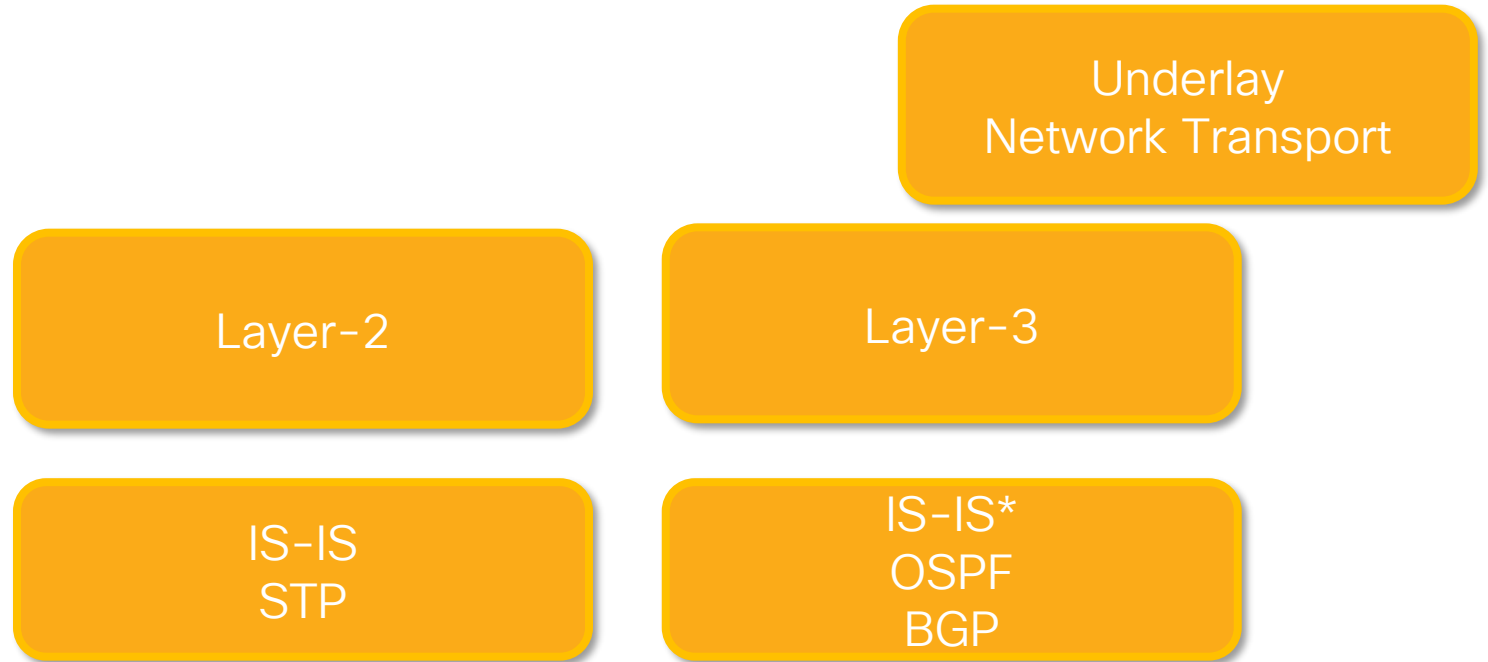
Tunnel Encapsulation



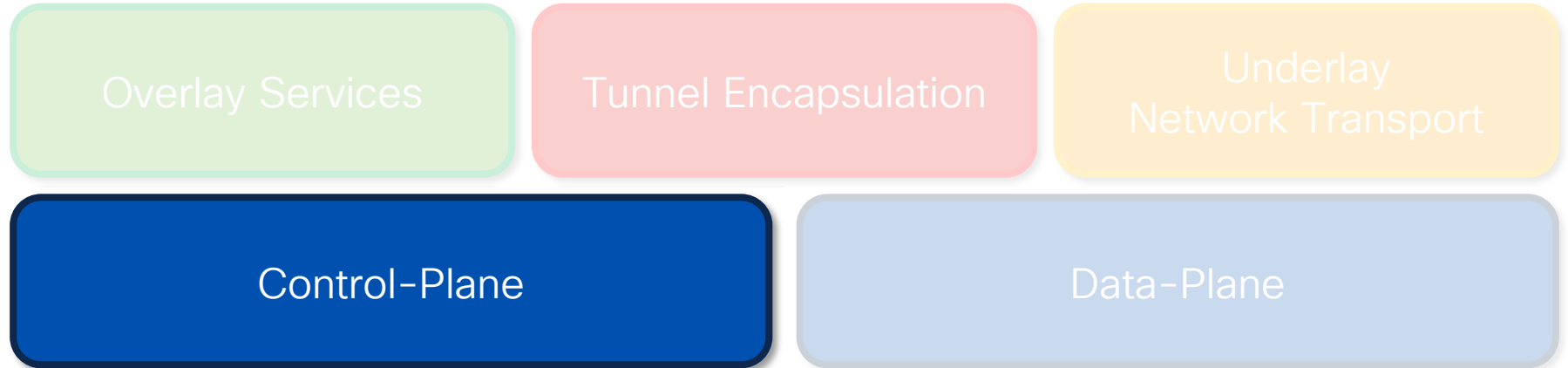
Underlay Network Transport



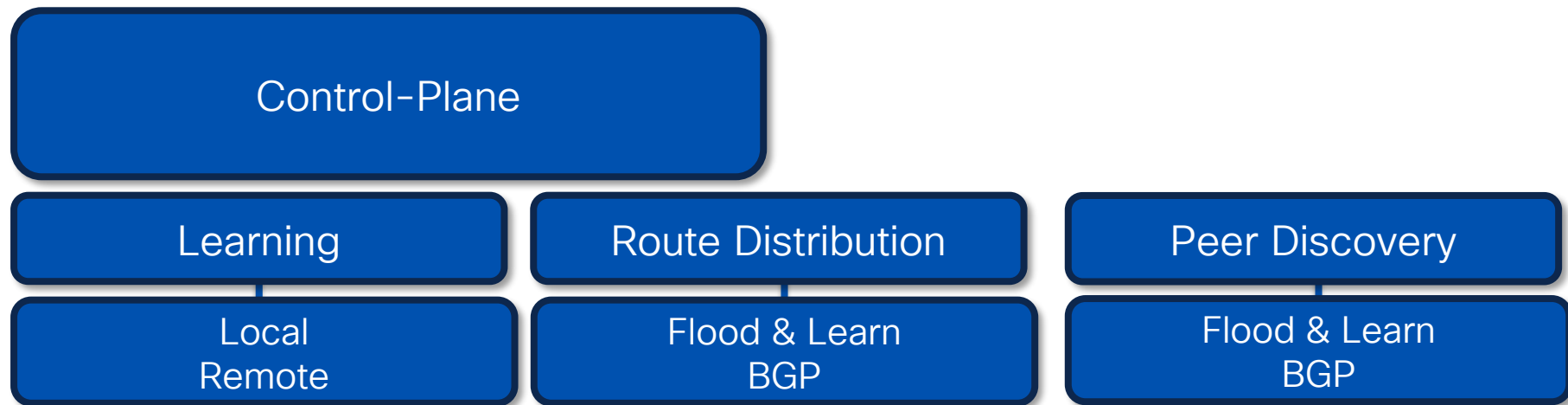
Underlay Network Transport



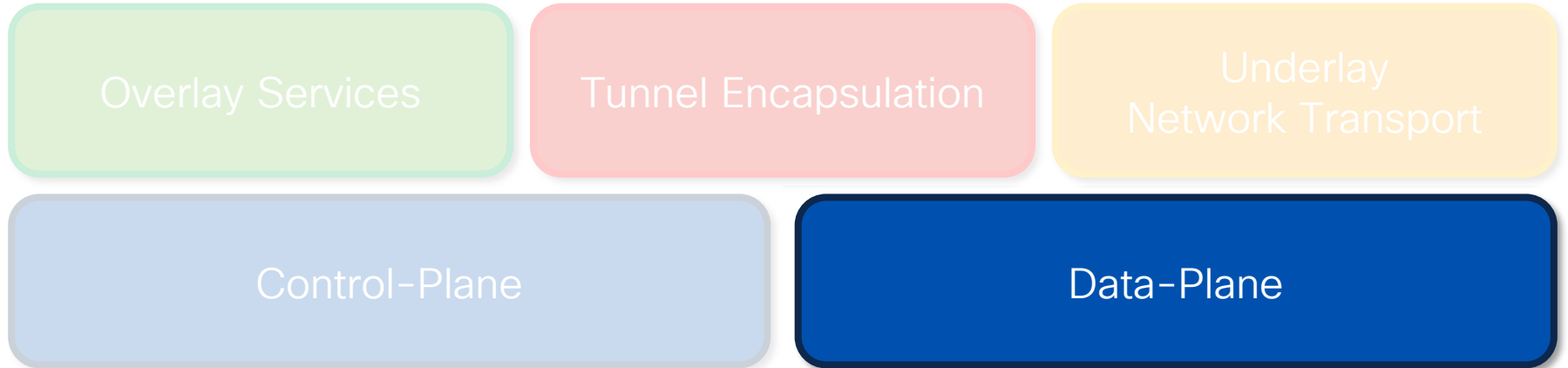
Control-Plane



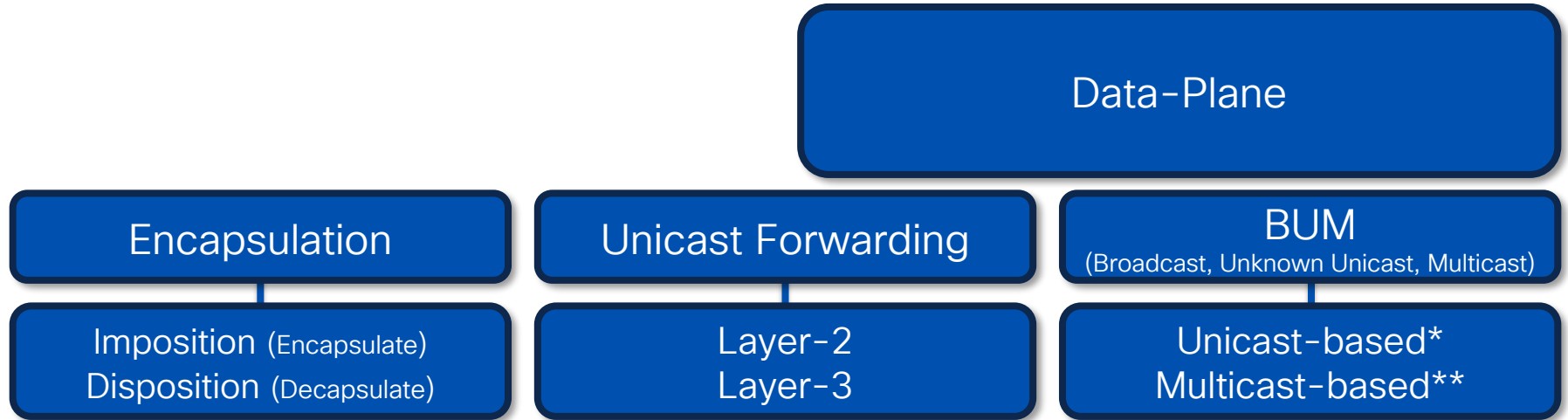
Control-Plane



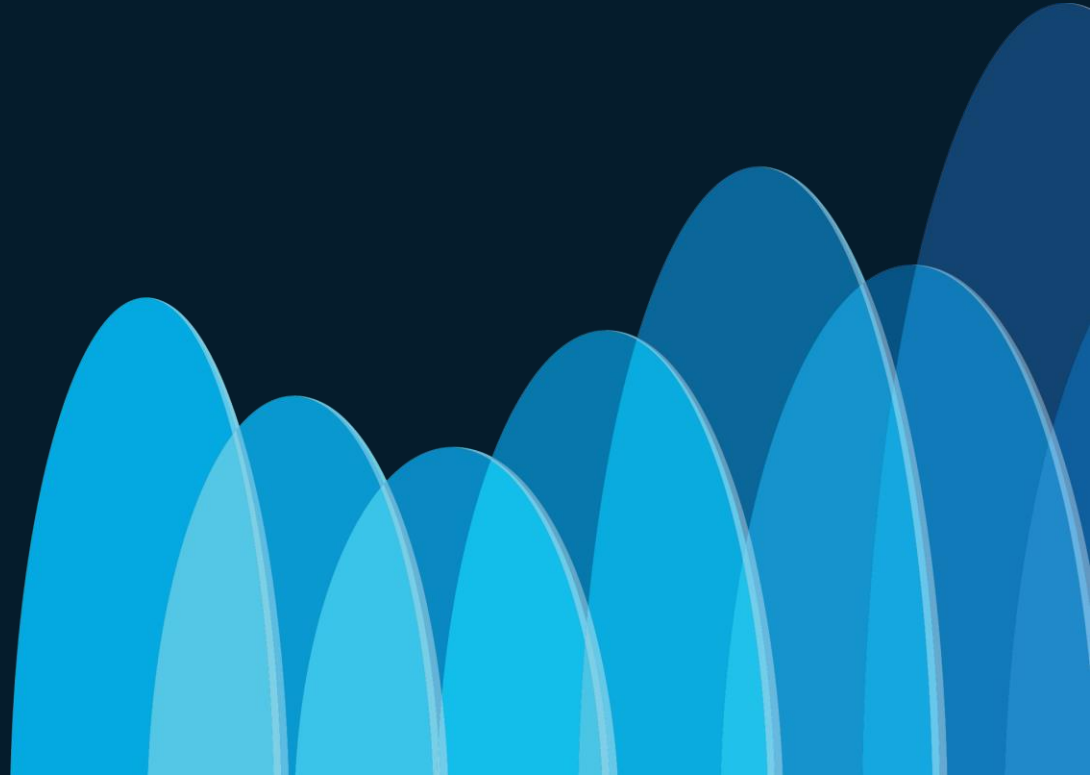
Data-Plane



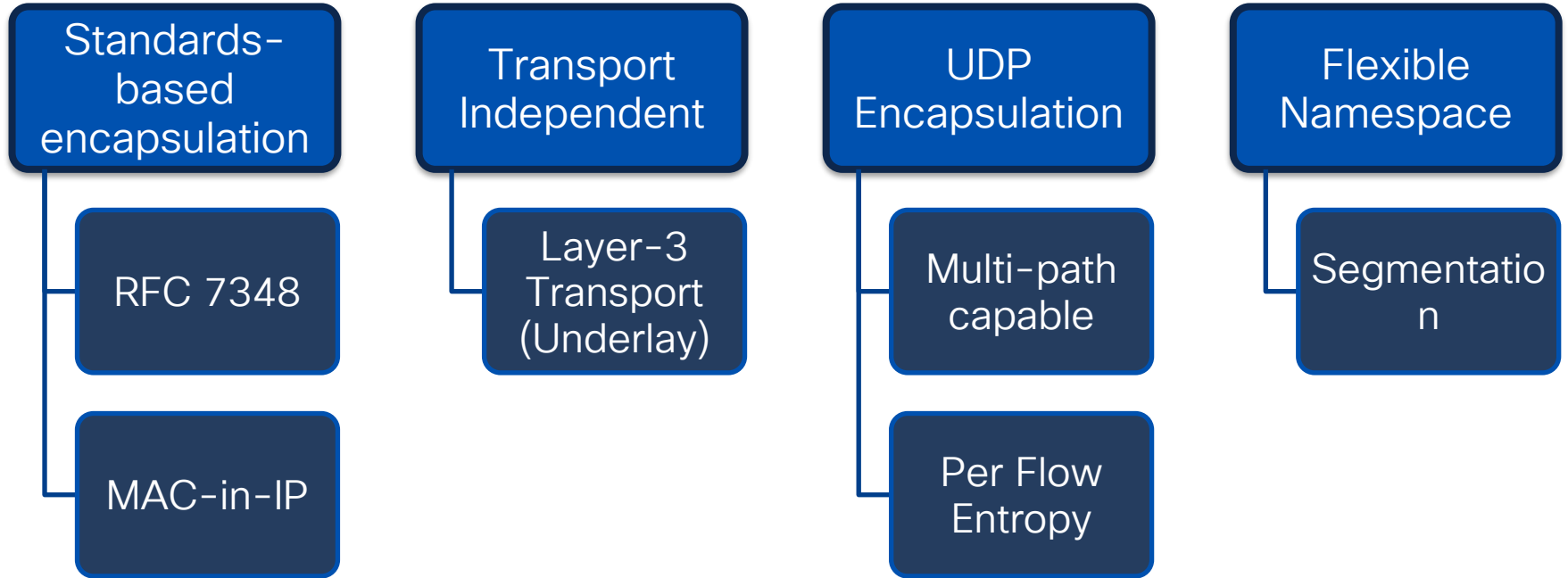
Data-Plane



VXLAN: Virtual Extensible LAN



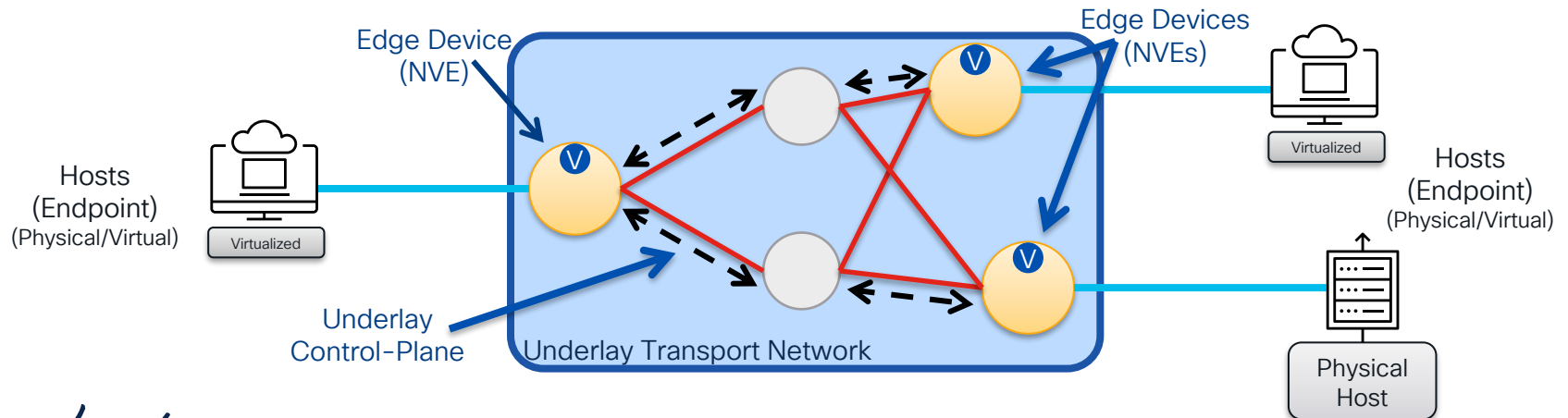
What is VXLAN?



Underlay Taxonomy

- Edge Devices host the VTEP
- Responsible for the encapsulation and decapsulation of the VXLAN Header

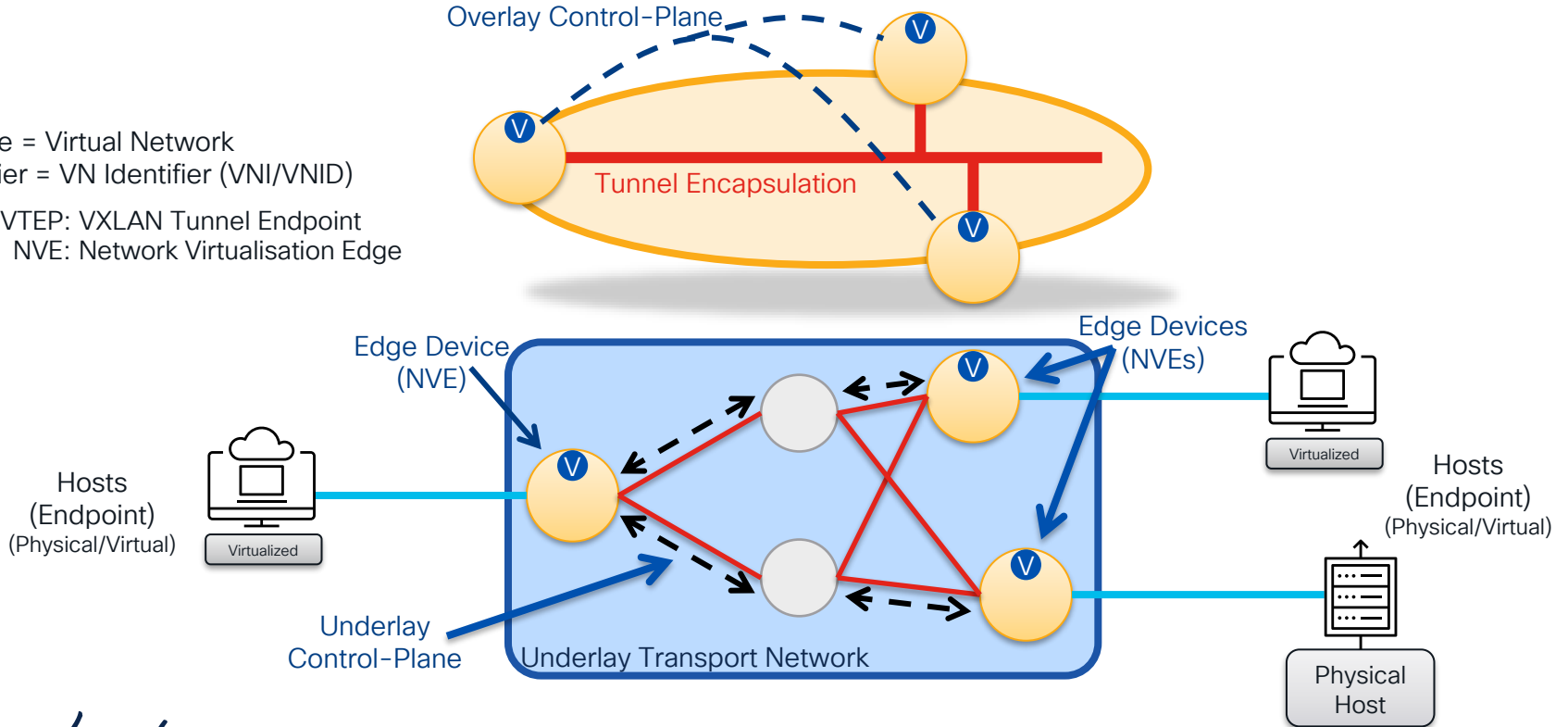
V VTEP: VXLAN Tunnel Endpoint
NVE: Network Virtualisation Edge



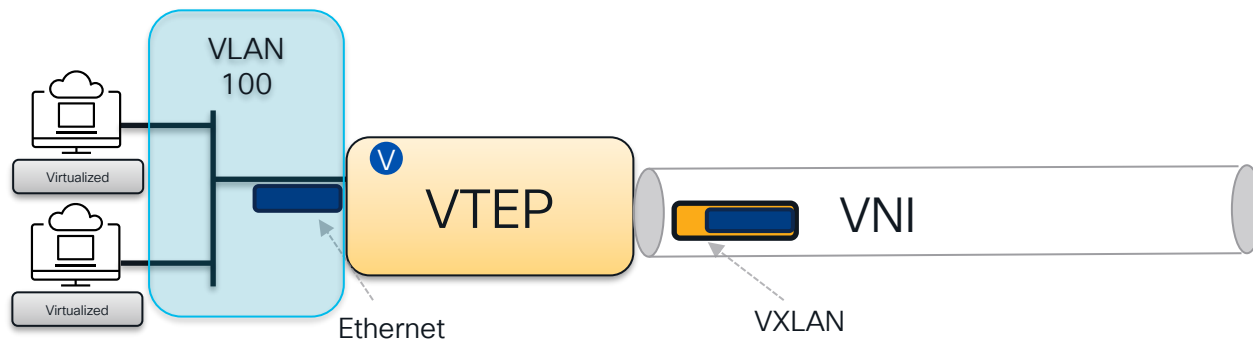
Overlay Taxonomy

Service = Virtual Network
Identifier = VN Identifier (VNI/VNID)

● VTEP: VXLAN Tunnel Endpoint
NVE: Network Virtualisation Edge

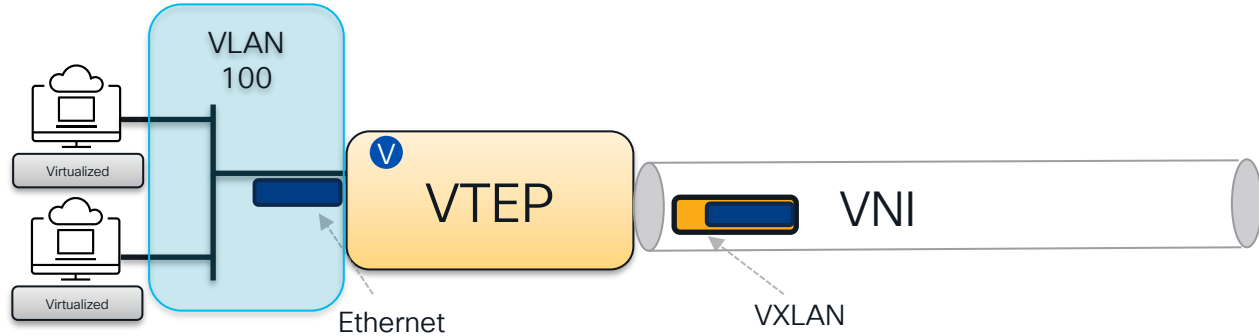
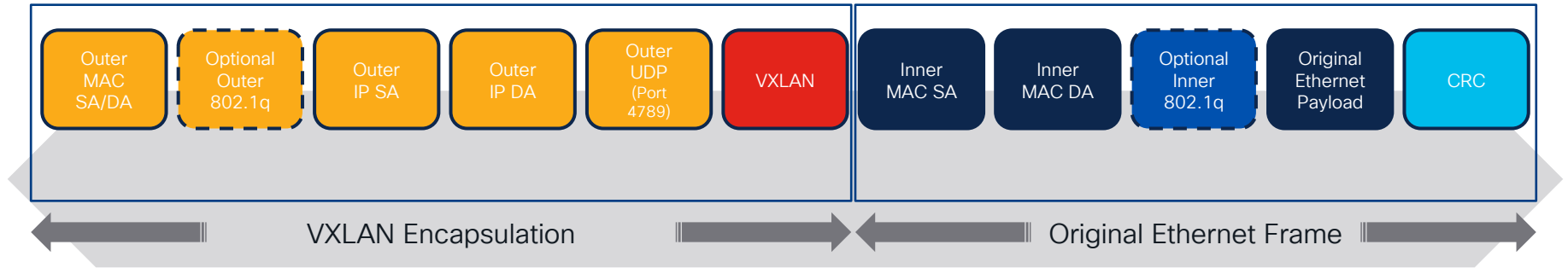


VXLAN (or Virtual) Tunnel Endpoint VTEP



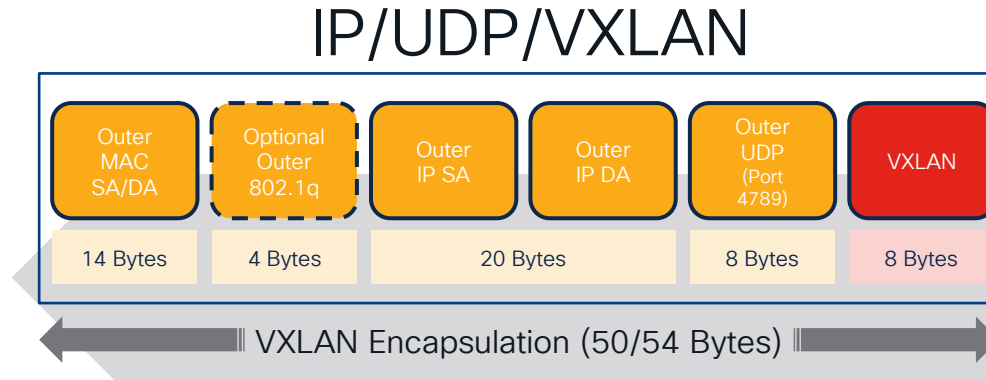
- VXLAN Tunnel Endpoint - Network Virtualisation Edge
- Each VTEP is uniquely identified by an IP Address
- VTEP Discovers or learns remote VTEPs, and end hosts attached to them
- VTEP bridges when forwarding packets within the same VNI and Routes for Inter-VNI traffic

VXLAN Encapsulation / Packet Format



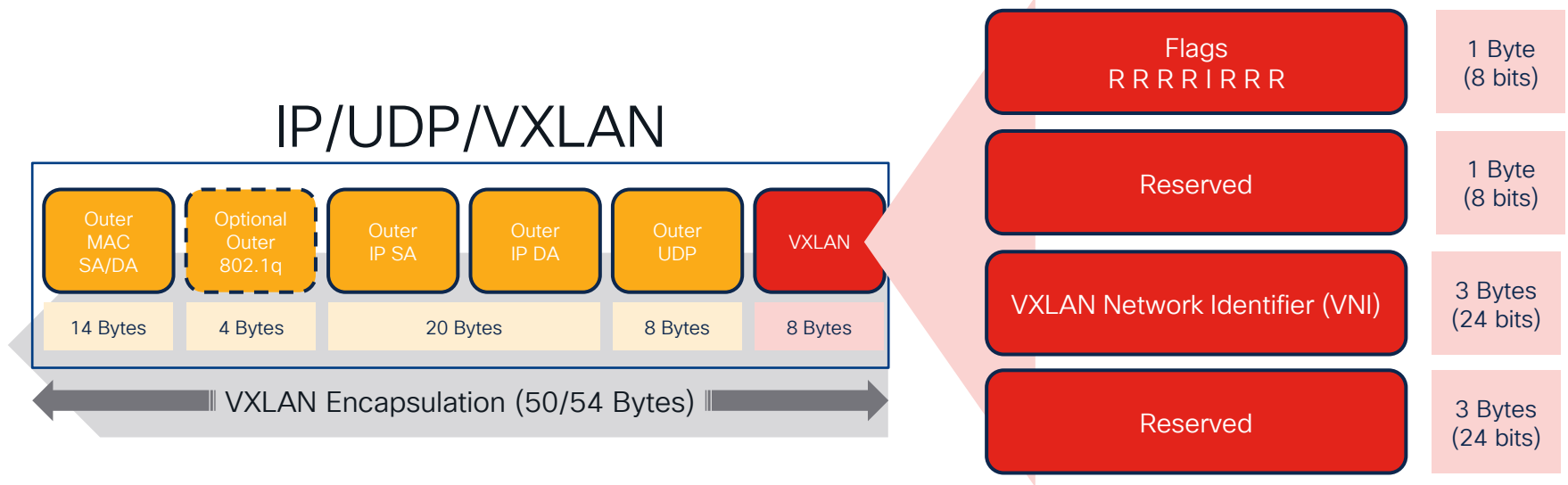
VXLAN Packet Format

- VXLAN uses MAC in UDP encapsulation

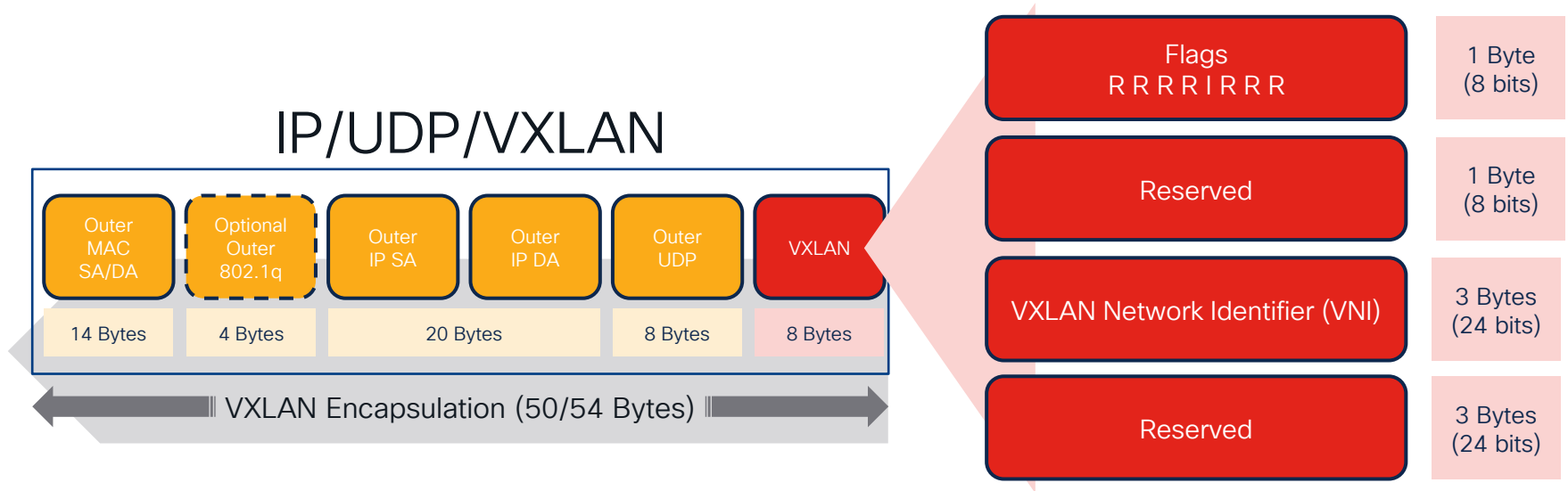


- Adds UDP and VXLAN Header before original Ethernet Frame

VXLAN Header Details

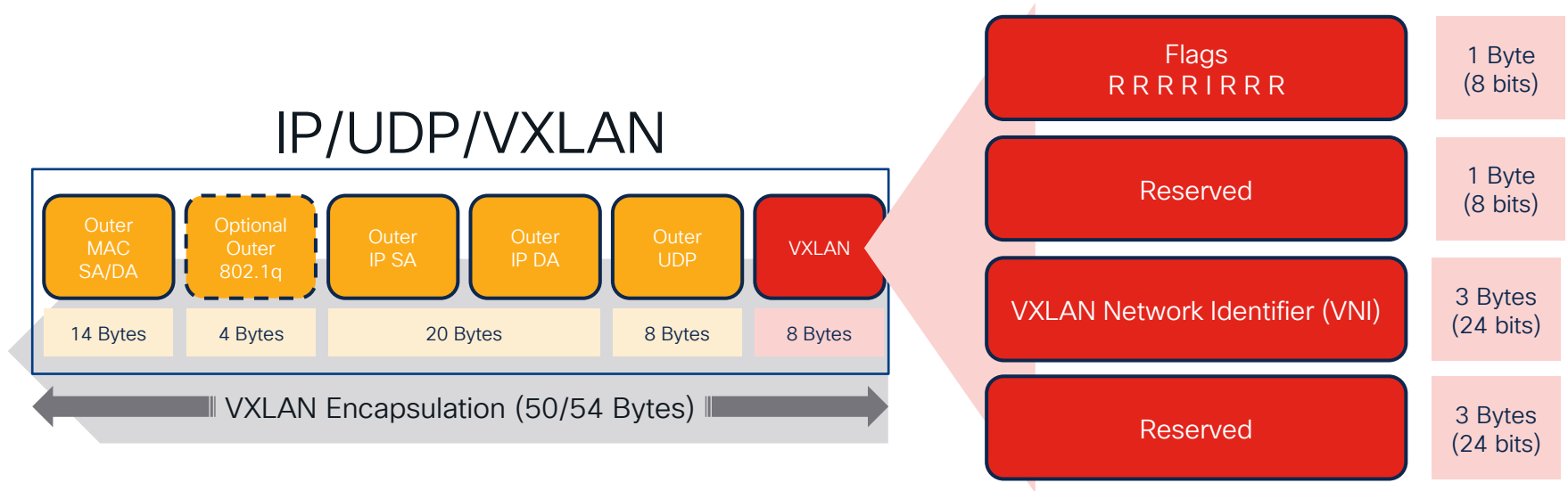


VXLAN Header Details



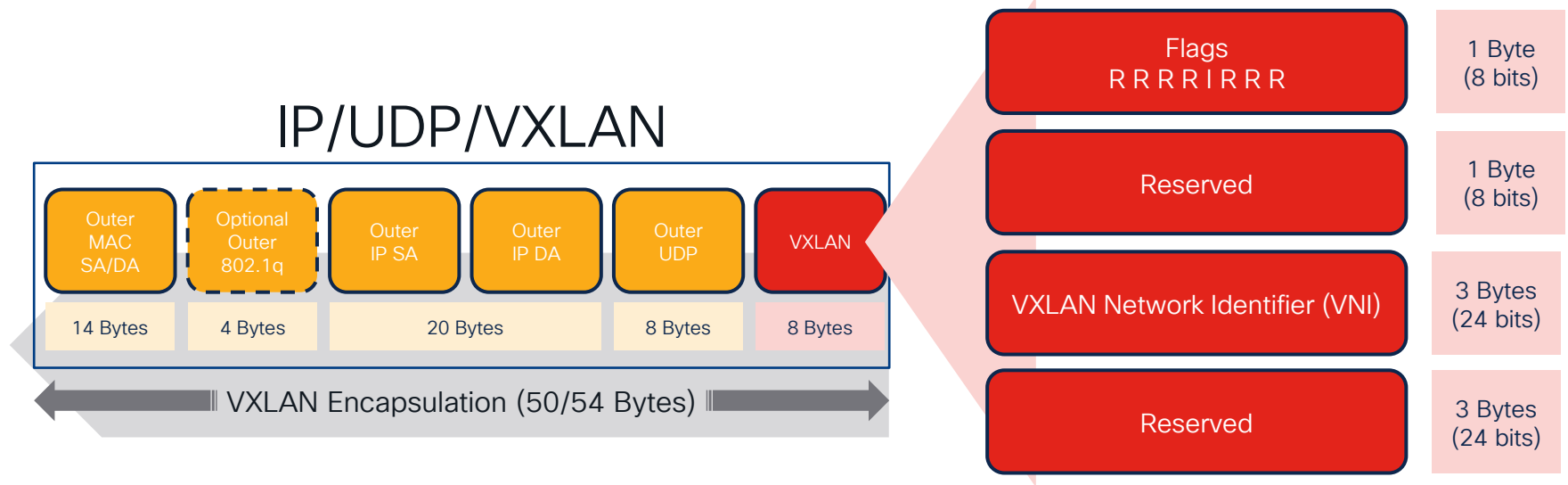
- Flags Field: I-flag (set to 1) for valid VNI. Other flags remain as R (set to 0)

VXLAN Header Details



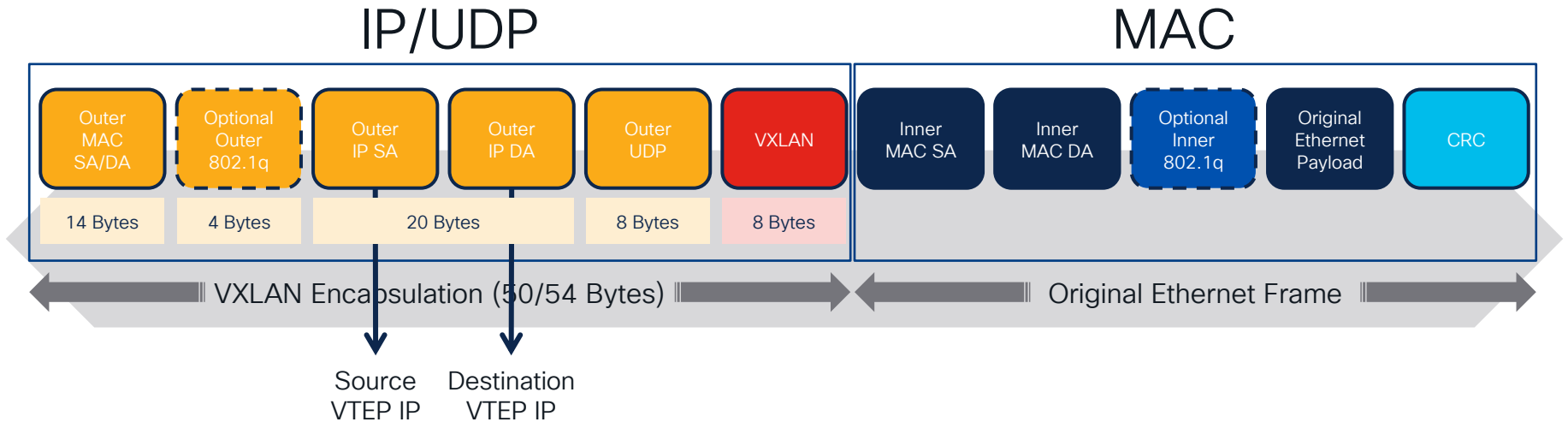
- Flags Field: I-flag (set to 1) for valid VNI. Other flags remain as R (set to 0)
- VNI Field: Allows VNI 1-16,777,215 (some implementation only 4096-16,777,215)

VXLAN Header Details



```
vlan 10
vn-segment 3001
```

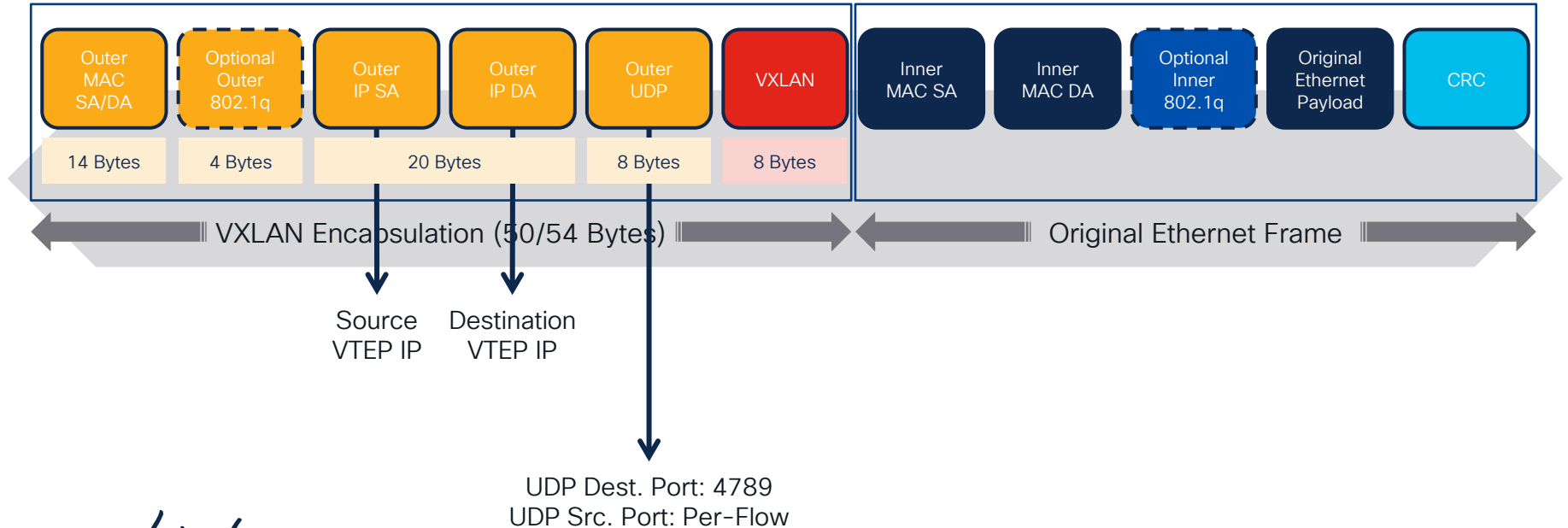
Transport Independence



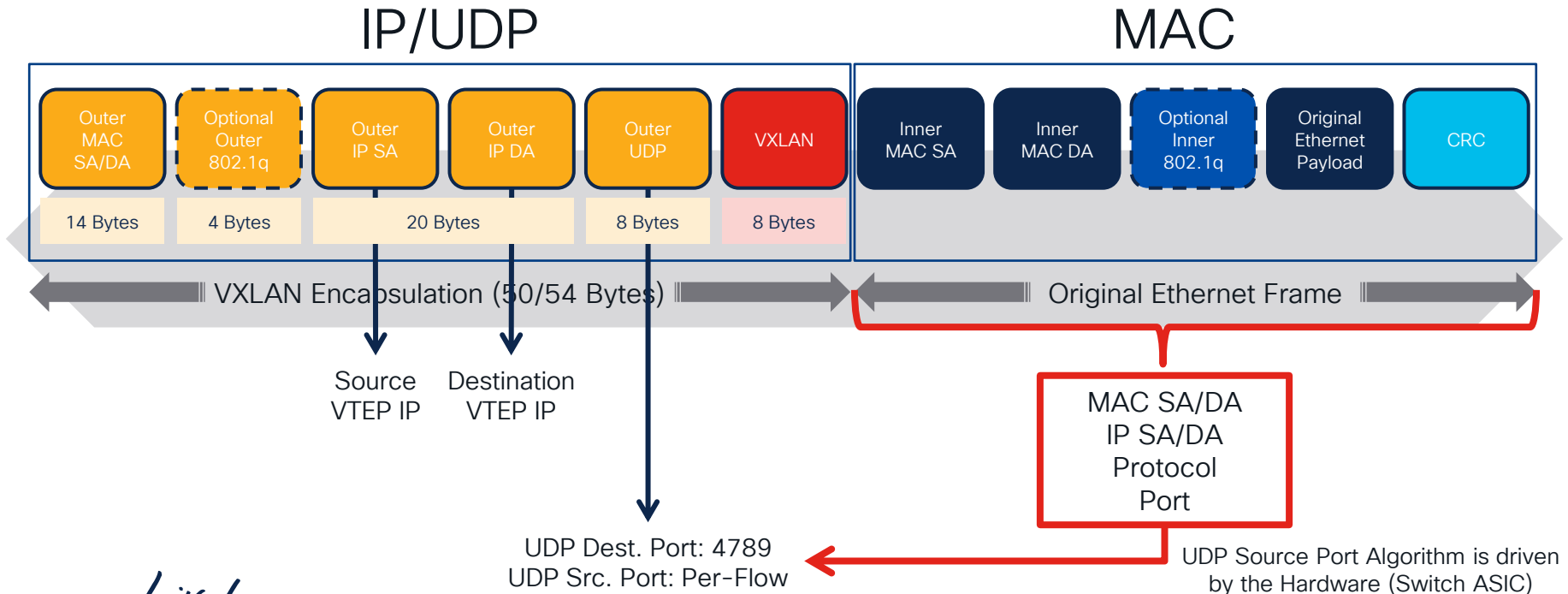
Multipath Capable

IP/UDP

MAC



Multipath Capable

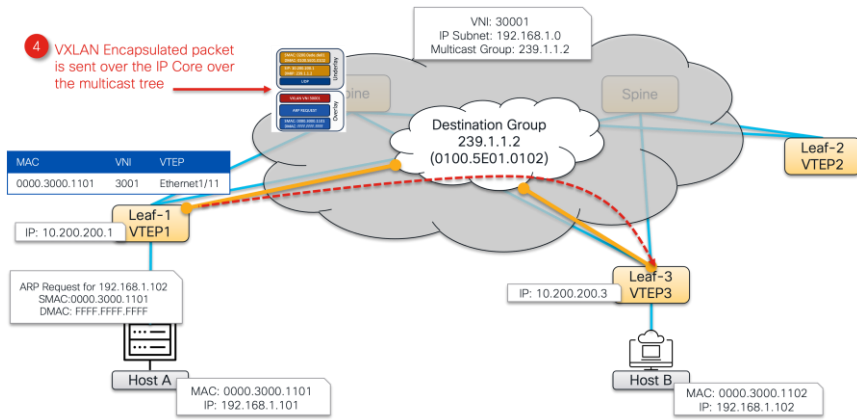
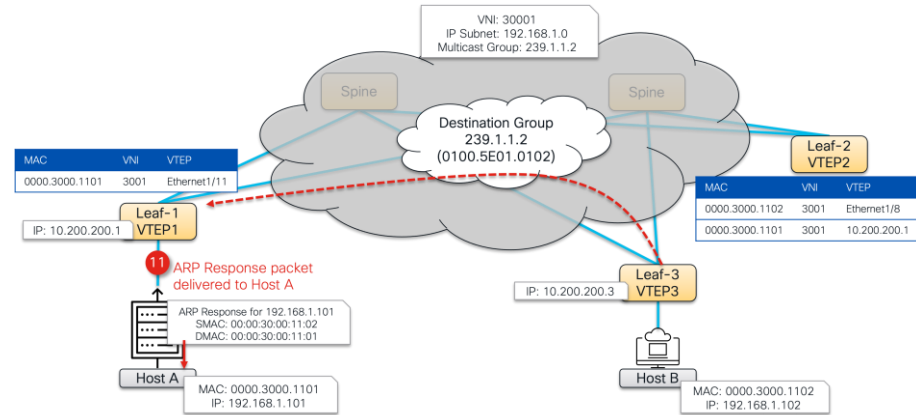


Data Plane Flood and Learn



Flood & Learn Mechanism

Multidestination Traffic is “Flooded” over the VXLAN Tunnel between VTEPs to “Learn” about the Host MAC addresses located behind the VTEPs so subsequent communication is delivered via Unicast



Flood & Learn Mechanism

Ingress Replication or Head-End Replication

- One unicast replica per remote VTEP
- Increases traffic load throughout the network
- Example: 10MB of BUM traffic for 20 remote VTEPs = 200MB of BUM traffic

Ingress Replication
to each VTEP

Leaf-1
VTEP1

ARP Request for 192.168.1.102
SMAC:0000.3000.1101
DMAC: FFFF.FFFF.FFFF



Host A

Leaf-2
VTEP2



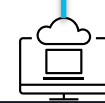
Host B

Leaf-3
VTEP3



Host C

Leaf-4
VTEP4



Host D

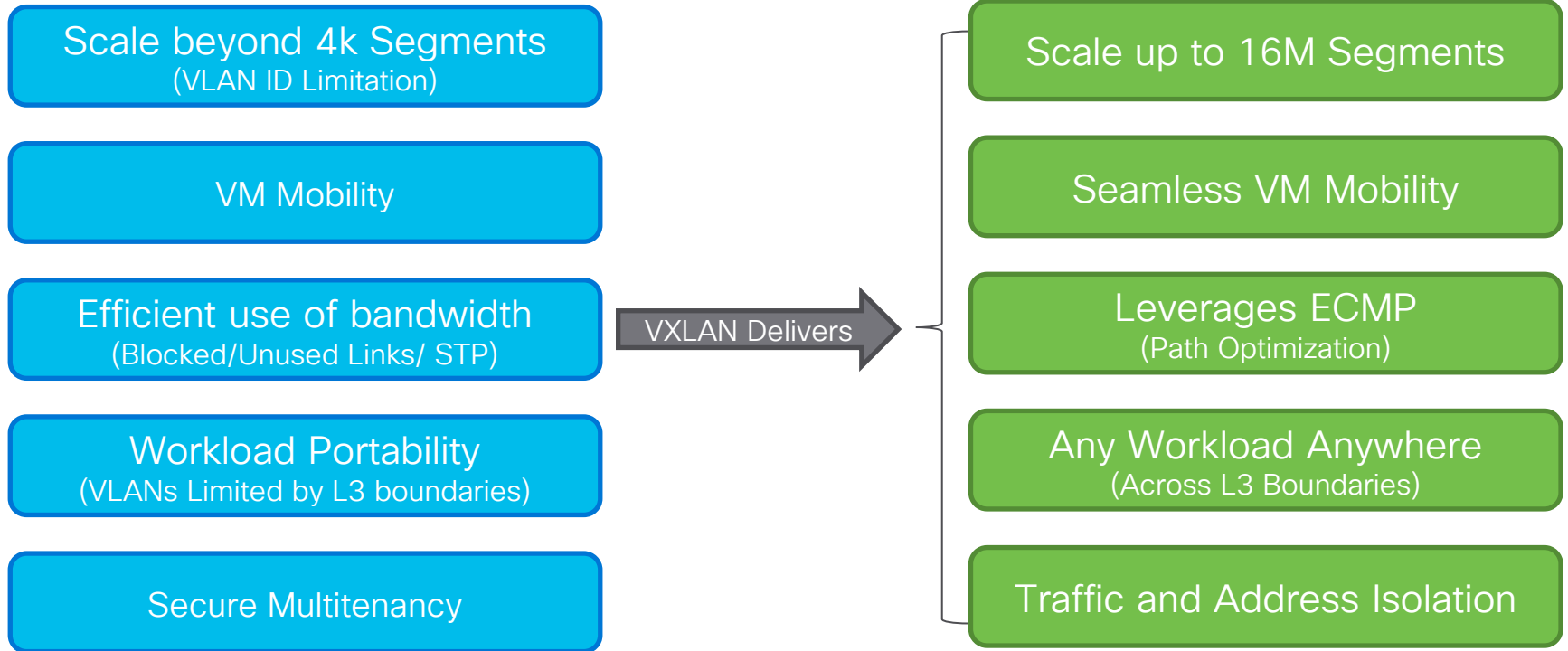
...

Leaf-20
VTEP20

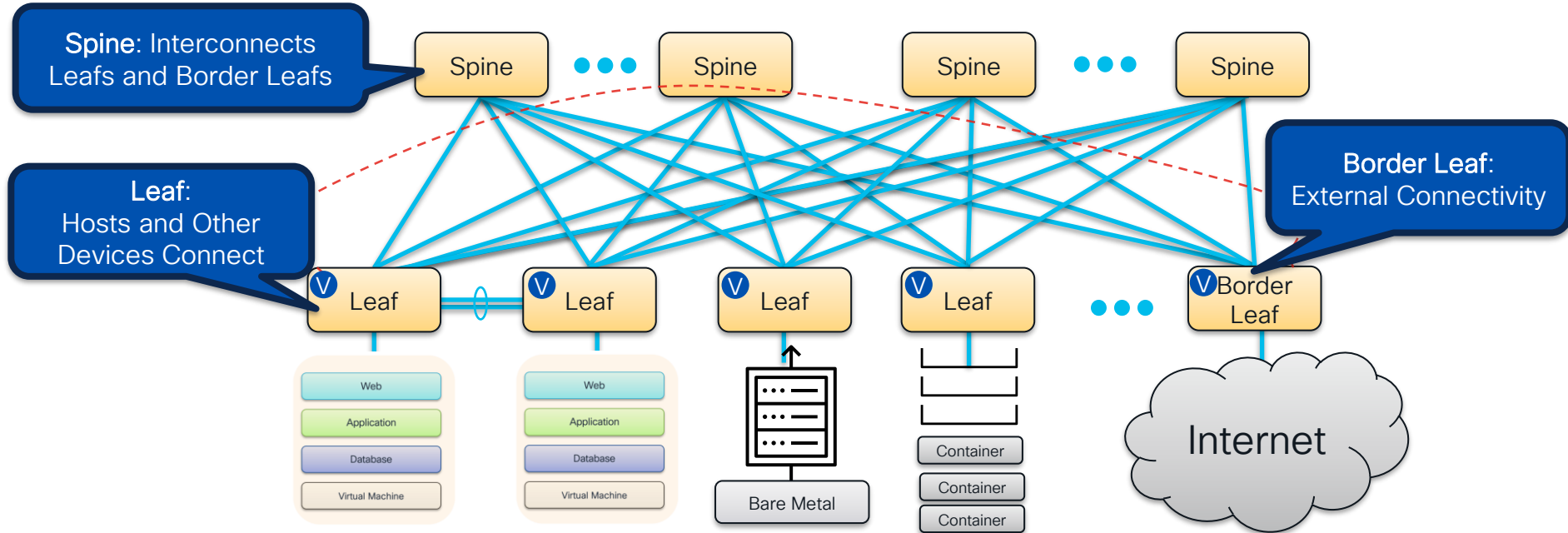


Host T

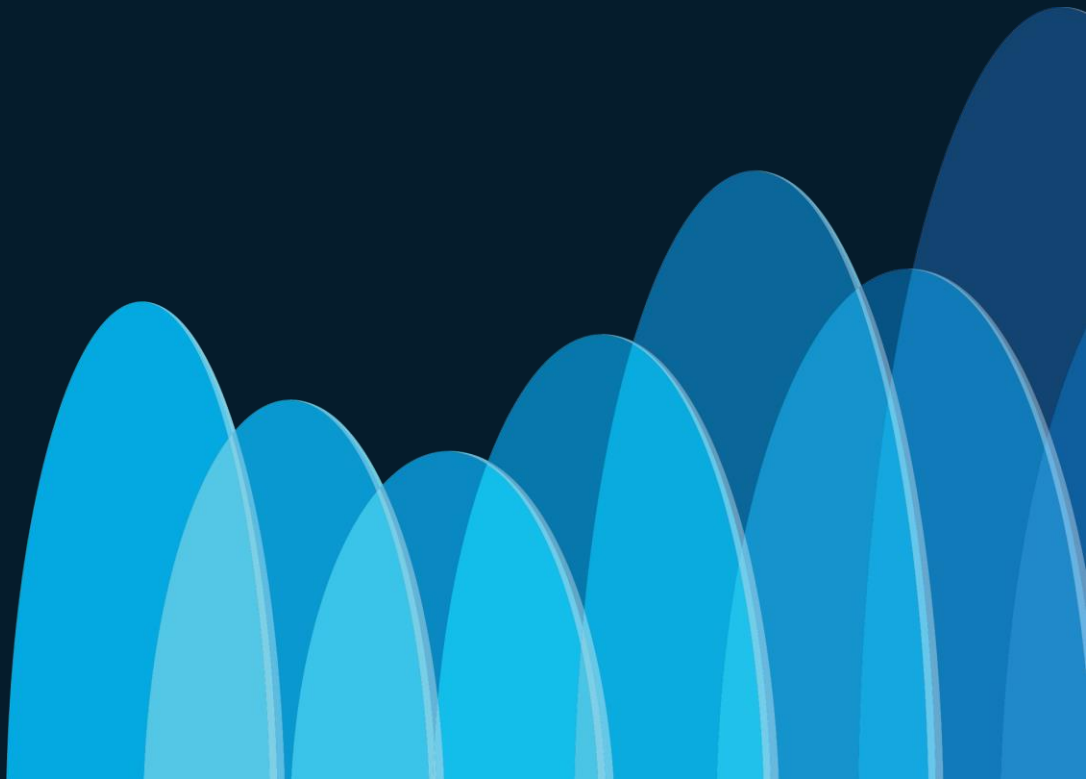
Why VXLAN? – How did we get here?



VTEP Device Roles



Control Plane EVPN



EVPN in the Data Center

IETF RFC/Draft for Control and Data Plane

Control-
Plane

EVPN MP-BGP RFC 7432

Data-
Plane

Multiprotocol Label
Switching (MPLS)
draft-ietf-l2vpn-evpn

Provider Backbone Bridges
(PBB)
draft-ietf-l2vpn-pbb-evpn

Network Virtualisation Overlay
(NVO)
draft-ietf-bess-evpn-overlay

- EVPN over NVO Tunnels (i.e., VXLAN) for Data Center Fabric Encapsulations
- Provides Layer-2 and Layer-3 Overlays over IP Networks



BGP EVPN Route Types – RFC 7432

RFC/Draft	Route Type	Description
RFC 7432	1	Ethernet Auto-Discovery EAD Route
	2	MAC/IP Advertisement Route
	3	Inclusive Multicast Ethernet Tag Route
	4	Ethernet Segment Route
draft-ietf-bess-evpn-prefix-advertisement	5	IP Prefix Route



BGP EVPN *Route Types - RFC 7432

RFC/Draft	Route Type	Description	
RFC 7432	1	Ethernet Auto-Discovery EAD Route	BGP Based Multi-Homing Mass Withdrawal/Aliasing
	2	MAC/IP Advertisement Route	L2 VNI MAC or MAC-IP from L2 MAC Learning or ARP
	3	Inclusive Multicast Ethernet Tag Route	Dynamic Peer Discovery for EVPN Ingress Replication
	4	Ethernet Segment Route	BGP Based Multi-Homing BUM DF Election/Split-Horizon
draft-ietf-bess-evpn- prefix-advertisement	5	IP Prefix Route	IETF Draft, Advertise IP Prefixes

*Type 6 - Selective Multicast Ethernet Tag Route

*Type 7/8 - IGMP Sync Routes

MP-BGP VPN Terminology

VPN Address Family

A Multi-Protocol BGP Extension to Distribute VPN Routes

Virtual Routing & Forwarding (VRF)

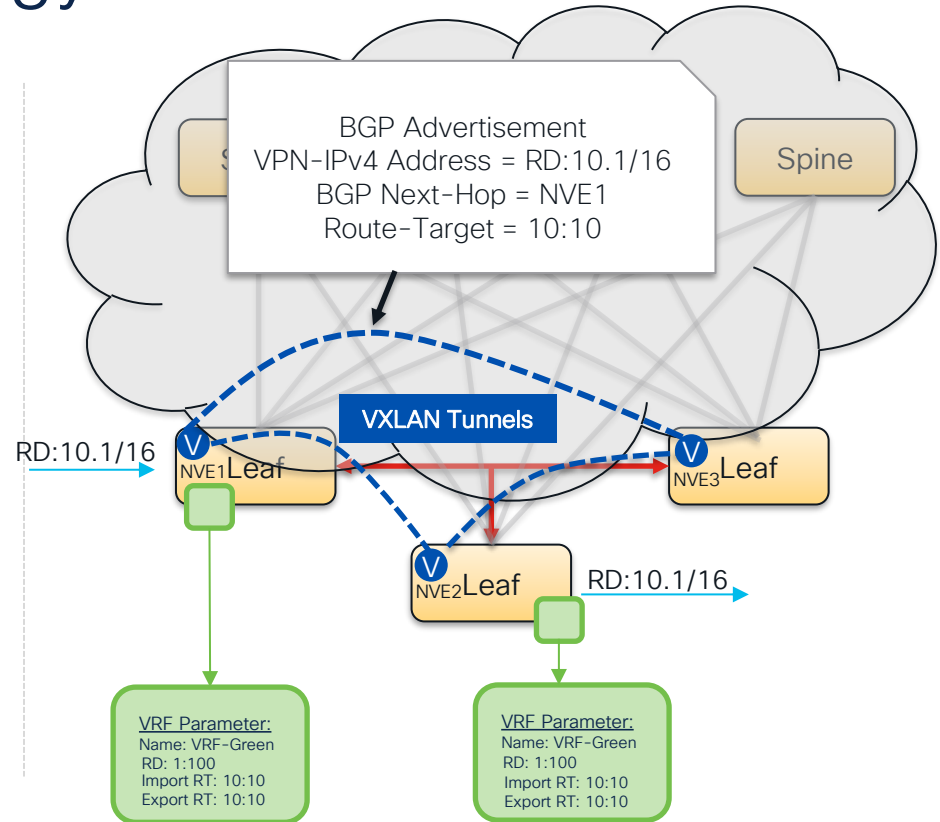
Overlay MAC or IP Routes are stored in dedicated VRF tables (MAC or IP VRF)

Route Distinguisher (RD)

8-byte Field, VRF Parameter
Value to make the MAC or IP Routes unique
(RD + VPN prefix)

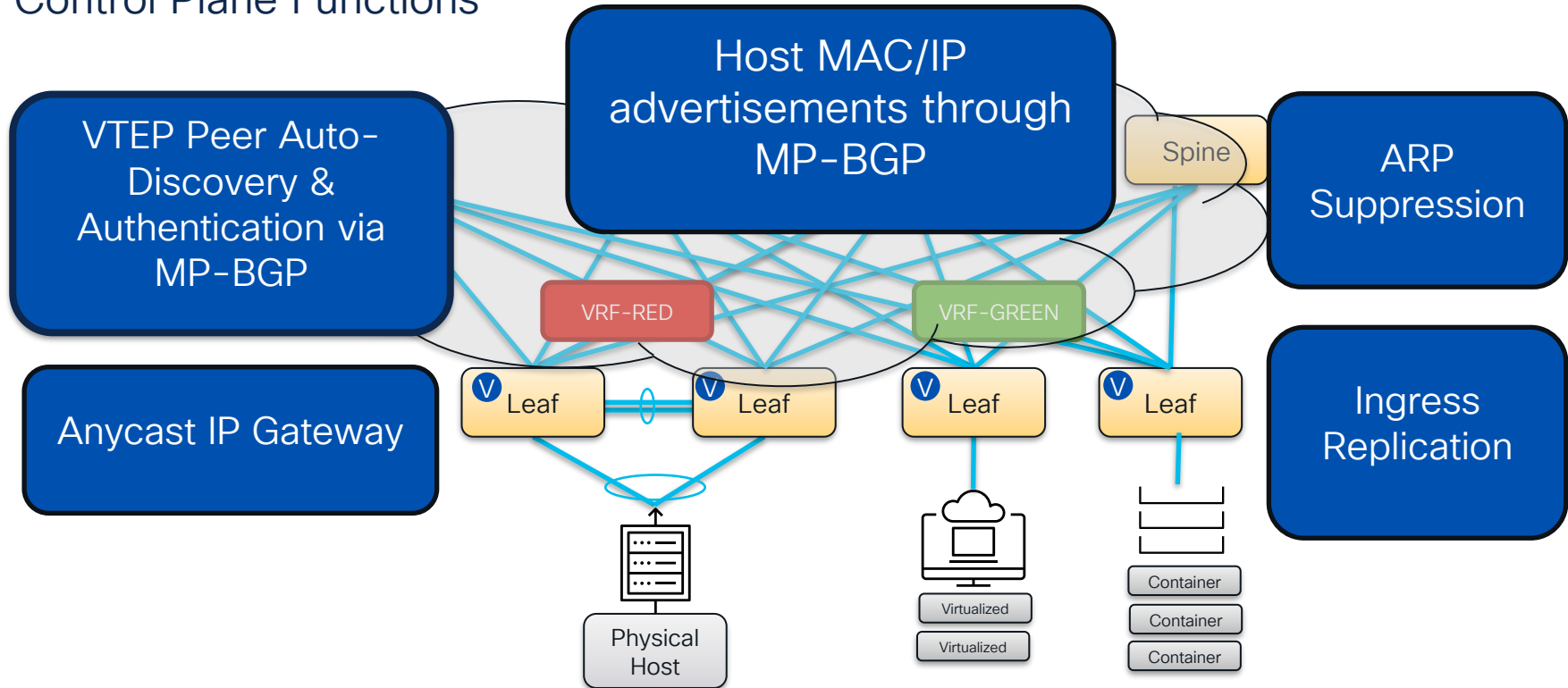
Route Target (RT)

8-byte Field, VRF Parameter
Value for the Import/Export Rules of VPN Routes (MAC or IP) into the VRF

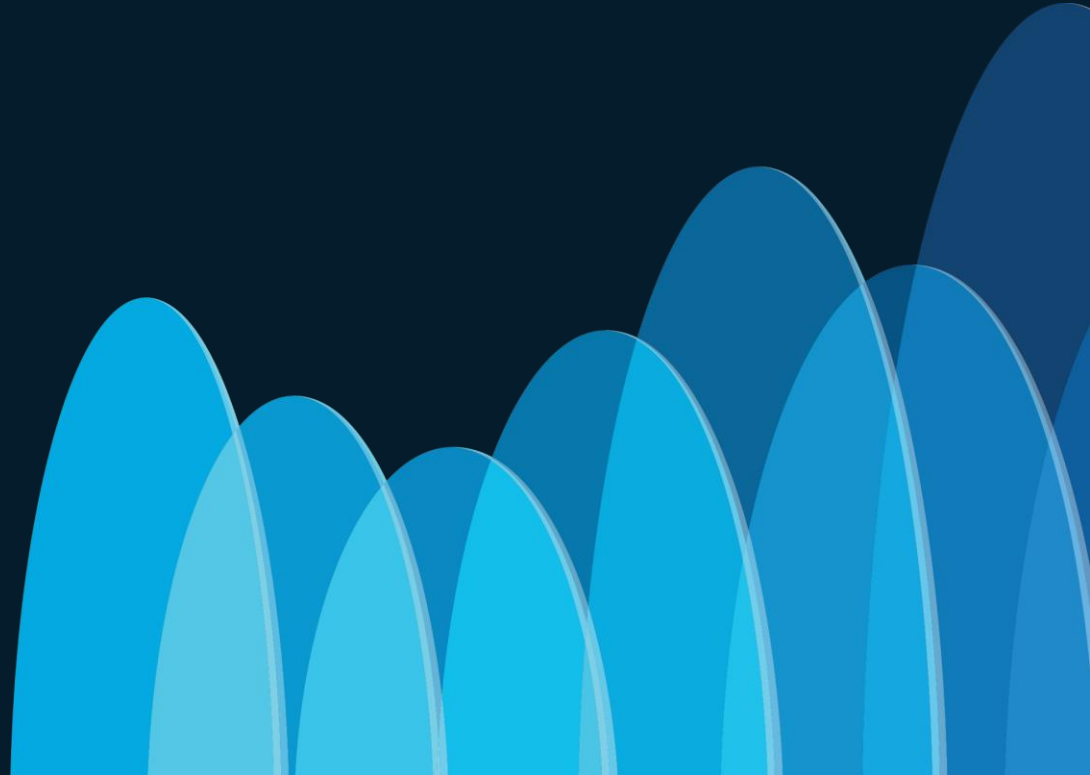


EVPN in the Data Center

Control Plane Functions



Configuring the EVPN VXLAN Fabric



VXLAN BGP EVPN

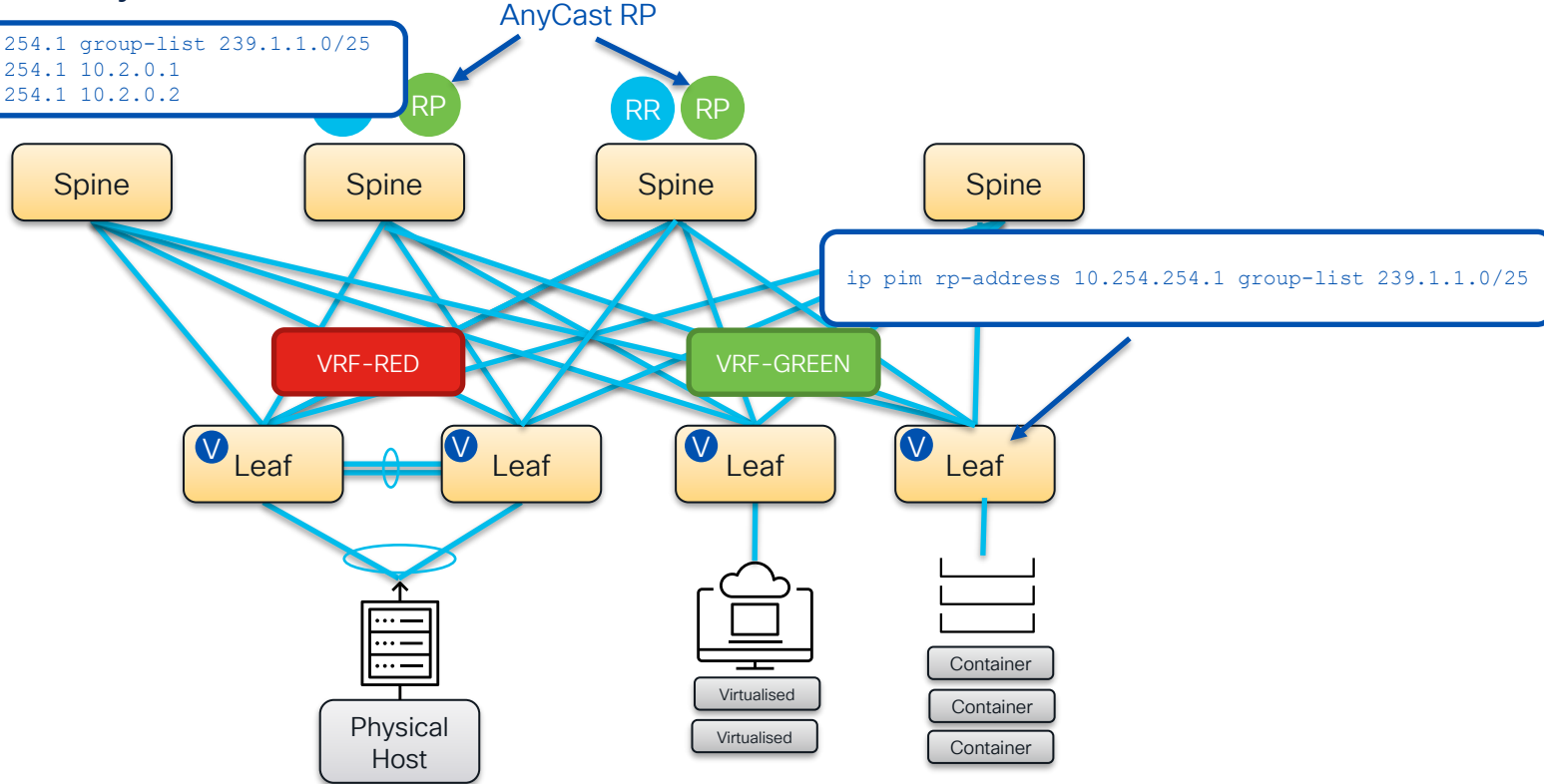
Network Underlay

RR Route Reflector

RP Rendezvous Point

```
ip pim rp-address 10.254.254.1 group-list 239.1.1.0/25
ip pim anycast-rp 10.254.254.1 10.2.0.1
ip pim anycast-rp 10.254.254.1 10.2.0.2
```

AnyCast RP



VXLAN BGP EVPN

Network Underlay

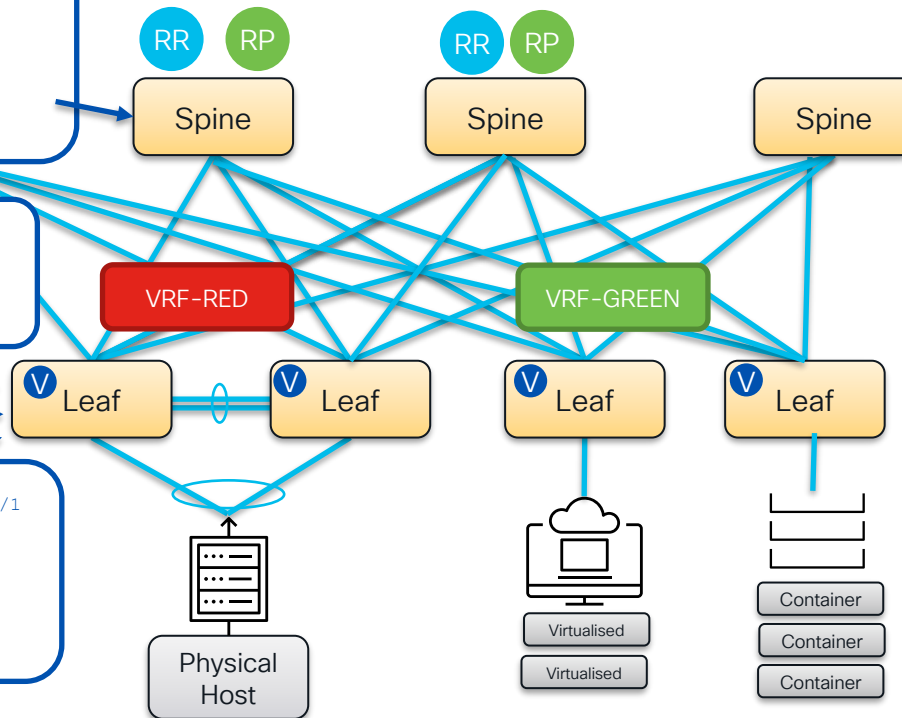
RR Route Reflector
RP Rendezvous Point

```
interface Ethernet1/2
description connected-to-leaf-1-Ethernet1/1
no switchport
mtu 9216
ip address 10.4.0.5/30
ip ospf network point-to-point
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
no shutdown
```

```
interface loopback0
description Routing loopback interface
ip address 10.2.0.7/32
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
```

```
router ospf UNDERLAY
router-id 10.2.0.7
```

```
interface Ethernet1/2
description connected-to-spine-1-Ethernet1/1
no switchport
mtu 9216
ip address 10.4.0.6/30
ip ospf network point-to-point
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
no shutdown
```



Virtual Networks and Overlay Interface

RP Rendezvous Point

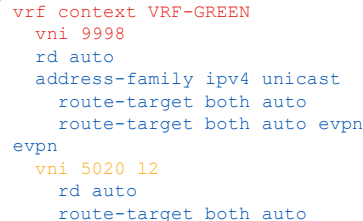
L3 VNI

```
vlan 1000
  vn-segment 9999
vlan 2000
  vn-segment 9998
```

```
interface nve1
  no shutdown
  host-reachability protocol bgp
  advertise virtual-rmac
  source-interface loopback1
  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



VXLAN BGP EVPN

BGP on Leaf and Spine



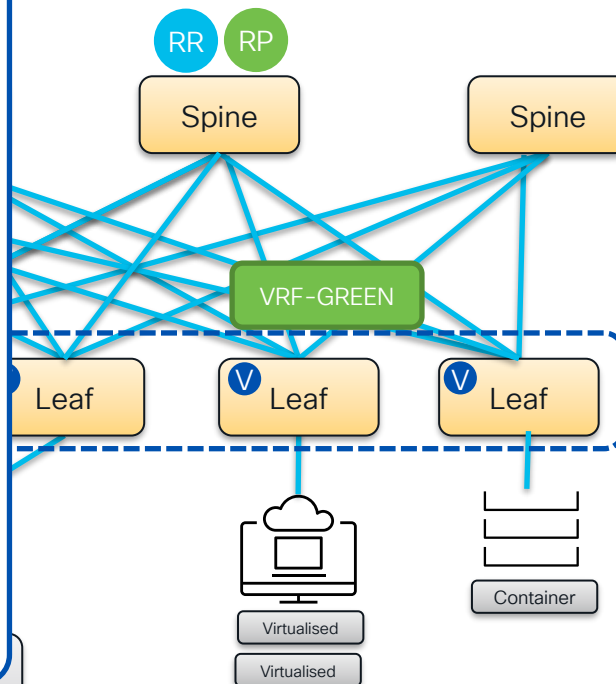
Route Reflector



Rendezvous Point

```
router bgp 65010
  router-id 10.2.0.7
  address-family l2vpn evpn
    advertise-pip
  neighbor 10.2.0.1
    remote-as 65010
    update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
  neighbor 10.2.0.2
    remote-as 65010
    update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
vrf VRF-RED
  address-family ipv4 unicast
    advertise l2vpn evpn
    redistribute direct route-map fabric-rmap-redist-subnet
    maximum-paths ibgp 2
  address-family ipv6 unicast
    advertise l2vpn evpn
    redistribute direct route-map fabric-rmap-redist-subnet
    maximum-paths ibgp 2
vrf VRF-GREEN
  address-family ipv4 unicast
    advertise l2vpn evpn
    redistribute direct route-map fabric-rmap-redist-subnet
    maximum-paths ibgp 2
  address-family ipv6 unicast
    advertise l2vpn evpn
    redistribute direct route-map fabric-rmap-redist-subnet
```

Leaf BGP Config



Spine BGP Config

```
router bgp 65010
  router-id 10.2.0.2
  neighbor 10.2.0.3
    remote-as 65010
    update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
    route-reflector-client
  neighbor 10.2.0.4
    remote-as 65010
    update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
    route-reflector-client
  neighbor 10.2.0.5
    remote-as 65010
    update-source loopback0
  address-family l2vpn evpn
    send-community
    send-community extended
    route-reflector-client
```

VXLAN BGP EVPN

Distributed Anycast Gateway

RR Route Reflector

RP Rendezvous Point

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

VLAN to VNI Mapping

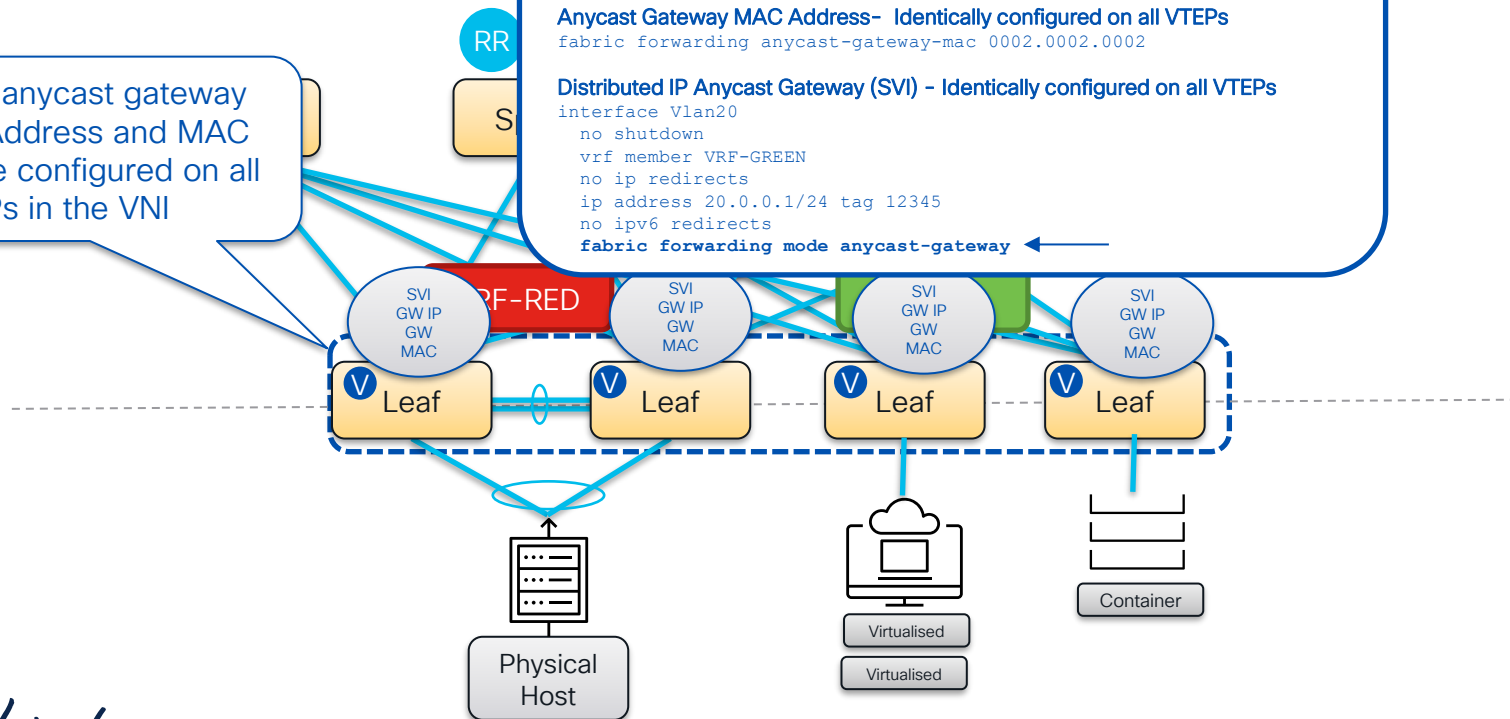
```
vlan 20  
vn-segment 5020
```

Anycast Gateway MAC Address- Identically configured on all VTEPs

```
fabric forwarding anycast-gateway-mac 0002.0002.0002
```

Distributed IP Anycast Gateway (SVI) - Identically configured on all VTEPs

```
interface Vlan20  
no shutdown  
vrf member VRF-GREEN  
no ip redirects  
ip address 20.0.0.1/24 tag 12345  
no ipv6 redirects  
fabric forwarding mode anycast-gateway
```



Conclusion



Control-Plane Comparison

Flood&Learn (RFC7348)

- Ethernet over IP
 - No Spanning-Tree (terminates at NVE)
 - Endpoint Learning is based on Flood and Learn (it's in the name)
 - Requires Extra Work for Routing
 - FHRP for Default Gateway
 - Over-the-Top VRF-lite for Prefix Routing (or use the Underlay?!)

EVPN – Ethernet VPN (RFC8365)

- A Better Ethernet over IP
 - No Spanning-Tree (terminates at NVE)
 - Endpoint Learning is based on BGP exchange (EVPN uses BGP)
 - Provides Integrated Routing & Bridging (IRB)
 - Distributed Anycast Gateway for Default Gateway
 - Uses a Layer-3 VPN approach like MPLS L3VPN
- And there is much more in EVPN!

Recommended Reading

- [Building Data Centers with VXLAN BGP EVPN: A Cisco NX-OS Perspective](#)
- [Cisco Nexus 9000 Series NX-OS VXLAN Configuration Guide, Release 10.4\(x\)](#)
- [Migrating Classic Ethernet Environments to VXLAN BGP EVPN](#)
- [Migrating Cisco FabricPath Environments to VXLAN BGP EVPN White Paper](#)
- [VXLAN EVPN Multi-Site Design and Deployment White Paper](#)
- [VXLAN Design with Cisco Nexus 9300 Platform Switches](#)
- [VXLAN Network with MP-BGP EVPN Control Plane Design Guide](#)

Recommended CiscoLive Sessions

- VXLAN BGP EVPN Multi-Site [BRKDCN-2913]
- VXLAN EVPN: A Deep Dive into Packet Forwarding [BRKDCN-3966]
- Next Generation VXLAN Data Centre Architecture, Deployment, and Operation [TECDCN-2101]
- Architecting L4-L7 Network Services in a Multi-tenant Data Centre with VXLAN EVPN [BRKDCN-2974]
- Design, Automate, and Manage VXLAN BGP EVPN Multi-Site with NDFC [BRKDCN-2128]

Keynote Deep Dives

Wednesday
10:30am – 11:30am



Experiences Amplified:
How AI Can Fuel Better Employee and Customer Experiences

Level 1
Room 106



Smart, Secure, Seamless:
Transforming Experiences with Next-Generation Networking

Level 2
Room 204



Harness a Bold New Era:
Transform Data Centre and Service Provider Connectivity

Level 2
Room 203



Securing User to Application and Everything in Between

Level 2
Melbourne Room 2



Unlocking Digital Resilience through Unified Observability

The HUB
Centre Stage

Complete Your Session Evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to claim a **Cisco Live T-Shirt**.



Complete your surveys in the **Cisco Live mobile app**.



Continue your education



- Visit the Cisco Stand for related demos
- Book your one-on-one Meet the Expert meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand



Thank you

CISCO *Live!*

#CiscoLiveAPJC

CISCO *Live!*

GO BEYOND

#CiscoLiveAPJC