
cisco
You make possible

Overlay Management and Visibility with VXLAN

Shyam Kapadia, Principal Engineer
Tweet: @shyamkapadia

BRKDCN-2125

cisco Live!
Barcelona | January 27-31, 2020

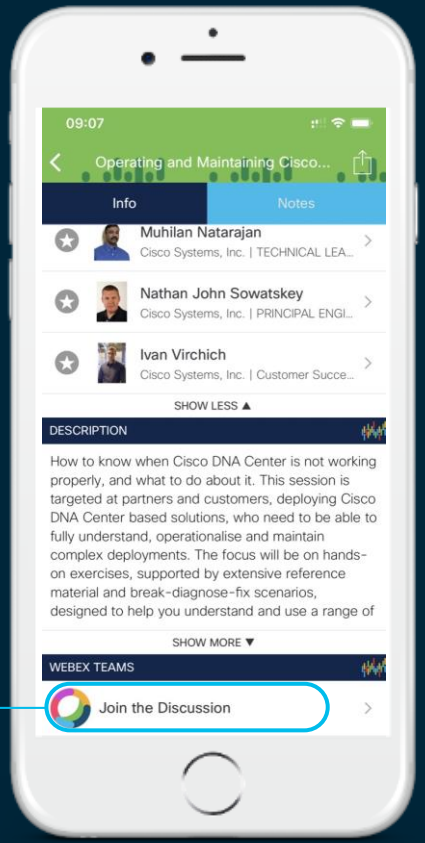
Cisco Webex Teams

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



Who am I?

- Born & brought-up in India
- PhD in Computer Science from USC (Fight on!)
- 12 years at Cisco
- Developer, Author, Inventor…
- Hobbies: Movies, Sports (Cricket, NFL, NBA)

Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

Overlays & Network Abstraction

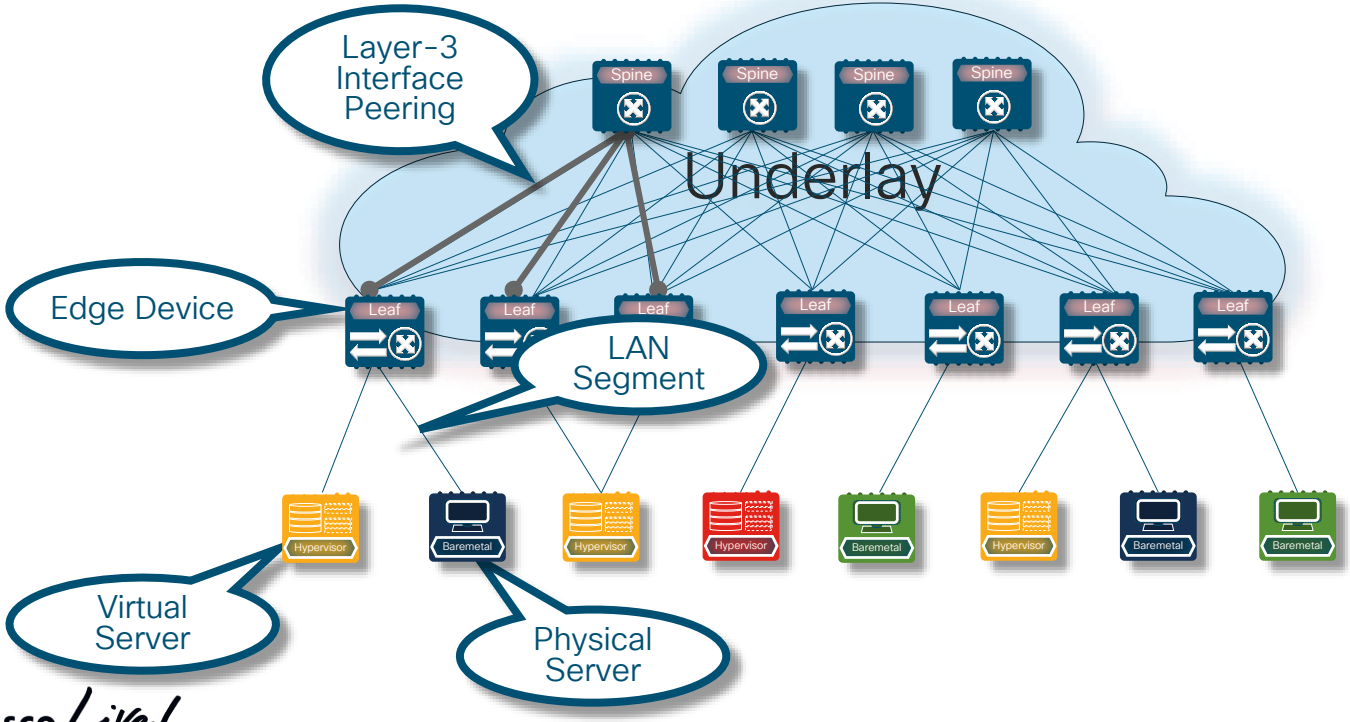
"All problems in computer science can be *solved* by another level of indirection"

David Wheeler

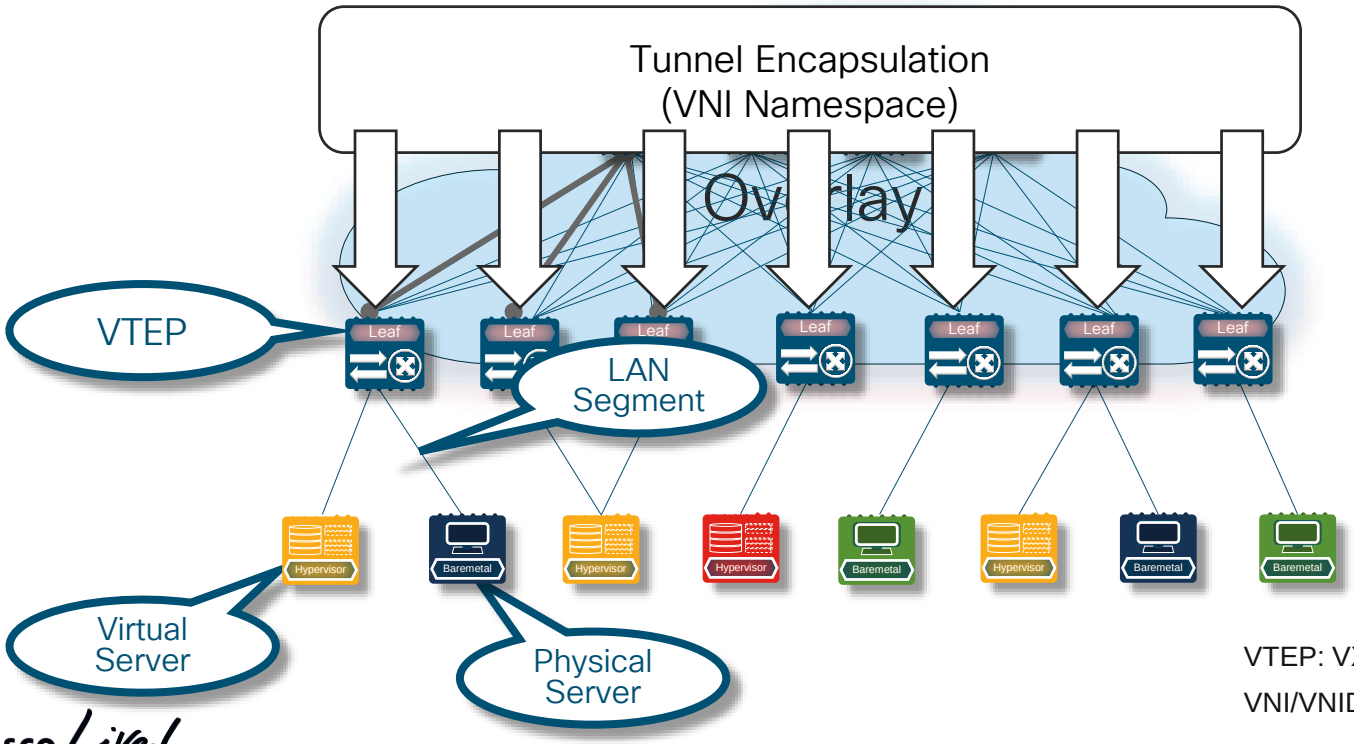
Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

Taxonomy – Underlay



Taxonomy – Overlay



VTEP: VXLAN Tunnel Endpoint

VNI/VNID: VXLAN Network Identifier

Understanding Overlay Technologies

Overlay Services
- Layer-2
- Layer-3
- Layer-2 and Layer-3

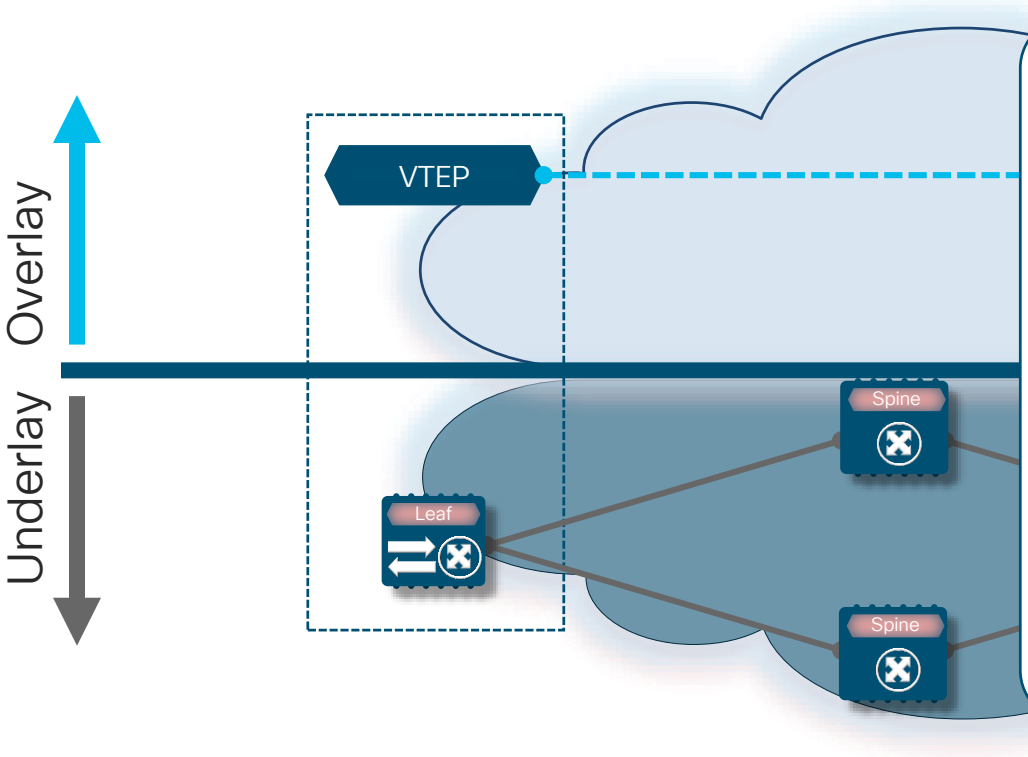
Tunnel Encapsulation

Underlay Transport Network

Control-Plane
- Peer-Discovery
- Route Learning and Distribution
  - Local Learning
  - Remote Learning

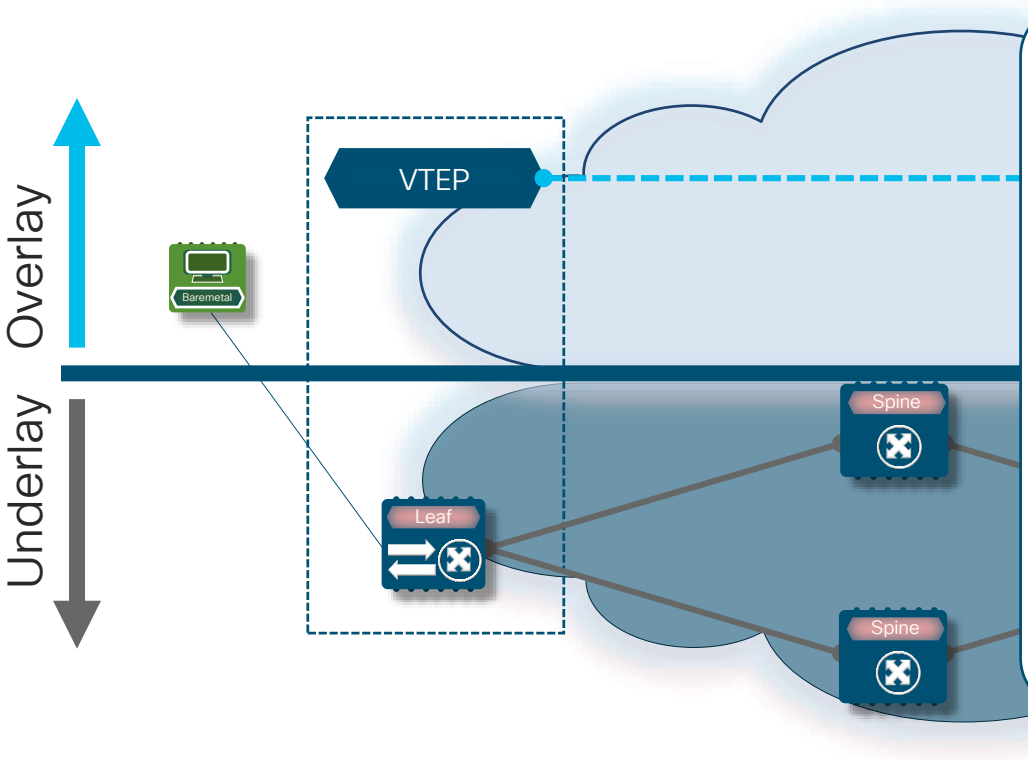
Data-Plane
- Overlay Layer-2/Layer-3 Unicast Traffic
- Overlay Broadcast, Unknown Unicast, Multicast (BUM) traffic forwarding
  - Ingress Replication (Unicast)
  - Multicast

Plumbing



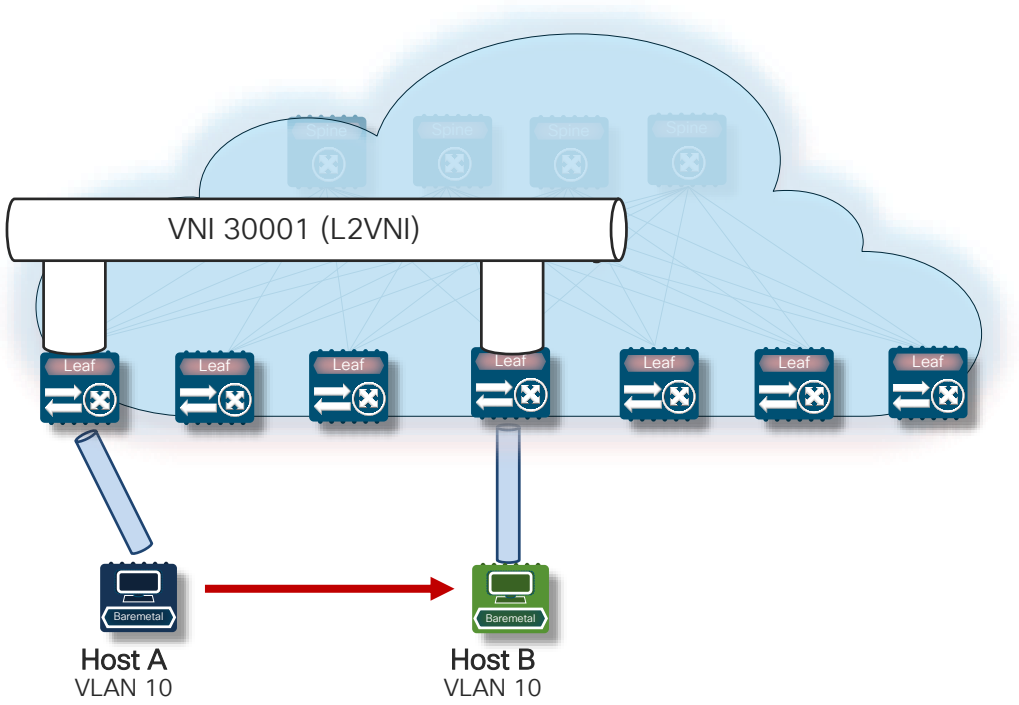
- Underlay – The Plumbing
  - Physical Connectivity
  - What is an Overlay?
- Overlay – Tunnels
  - Logical Transport Service
  - Virtual Private Network (VPN)
  - What is an Underlay?

Endpoint Connectivity vs. Endpoint Presence



- Connectivity – The Plumbing
  - Endpoint physically connected
  - ”I need a Interface”
- Presence – Tunnels
  - Endpoint logically present
  - ”I need other Endpoints to Communicate”

Overlay Forwarding – Bridging

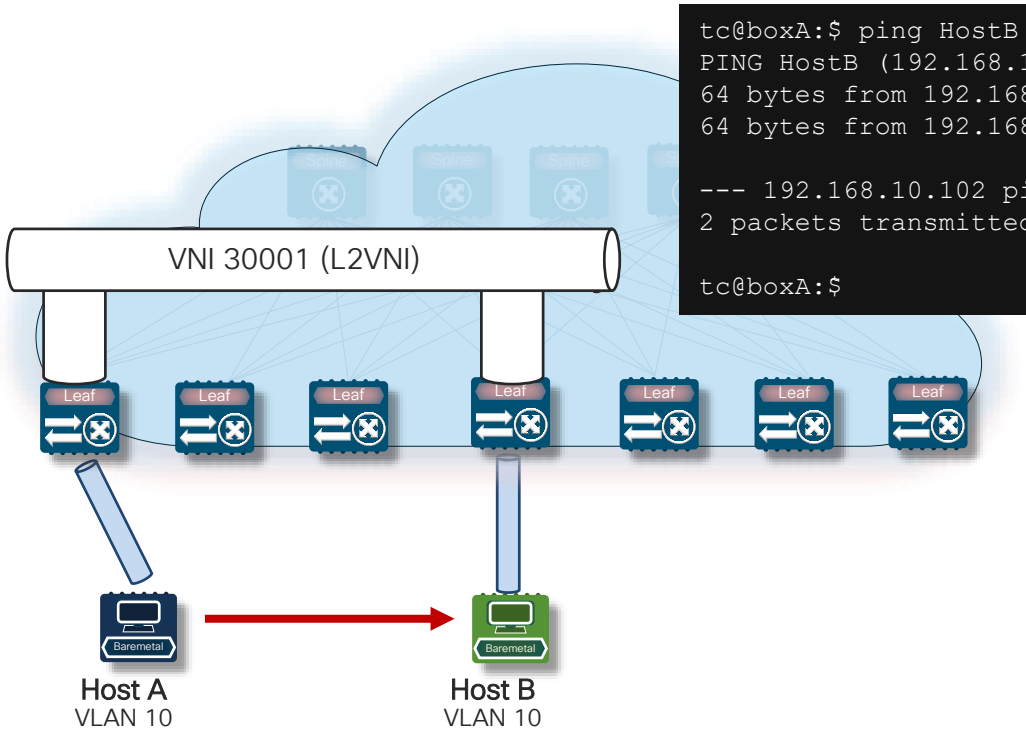


- Endpoint to Endpoint Communication
  - Layer-2 Service (VPN) for Bridging
- Leaf to Leaf Communication
  - The VTEP on the Leaf originates the Encapsulation
- The Spine sees only encapsulated packets
  - No Knowledge about Endpoints
  - Everything is IP/UDP*

cisco Live!

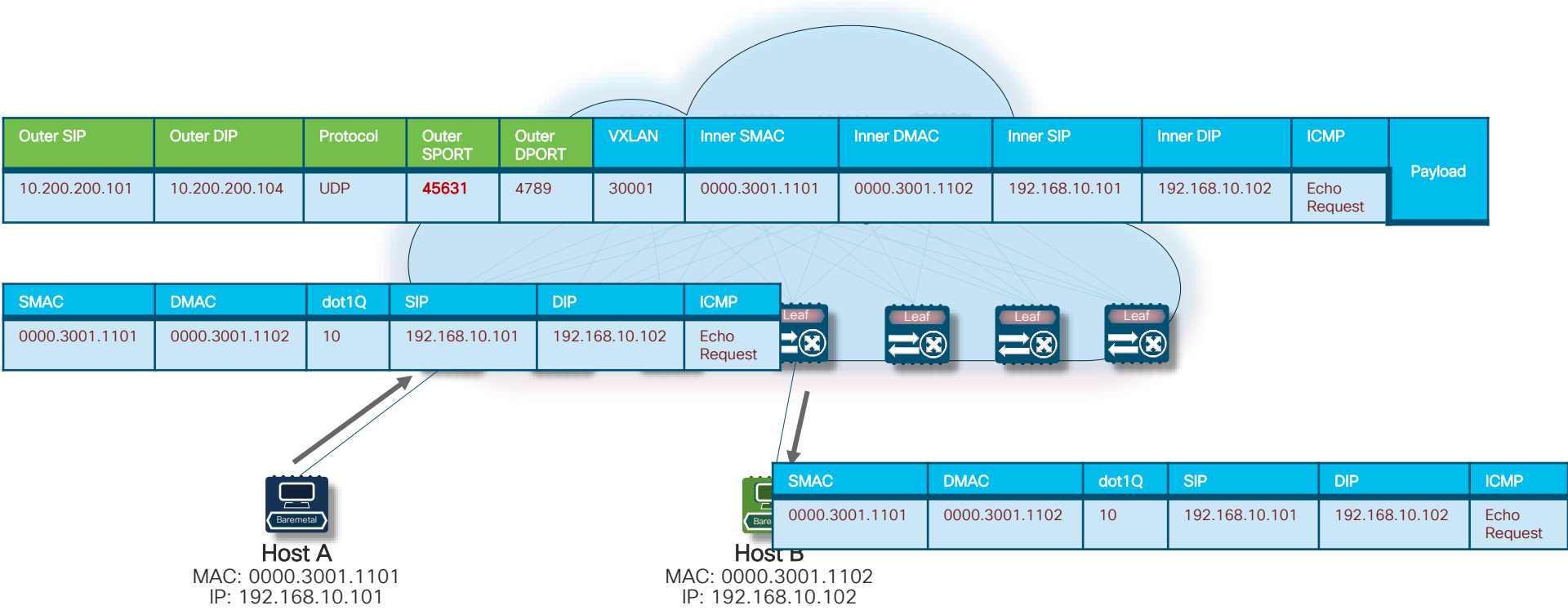
*VXLAN, LISP, OTV use IP/UDP Encapsulation

Overlay Forwarding – Pinging Host B

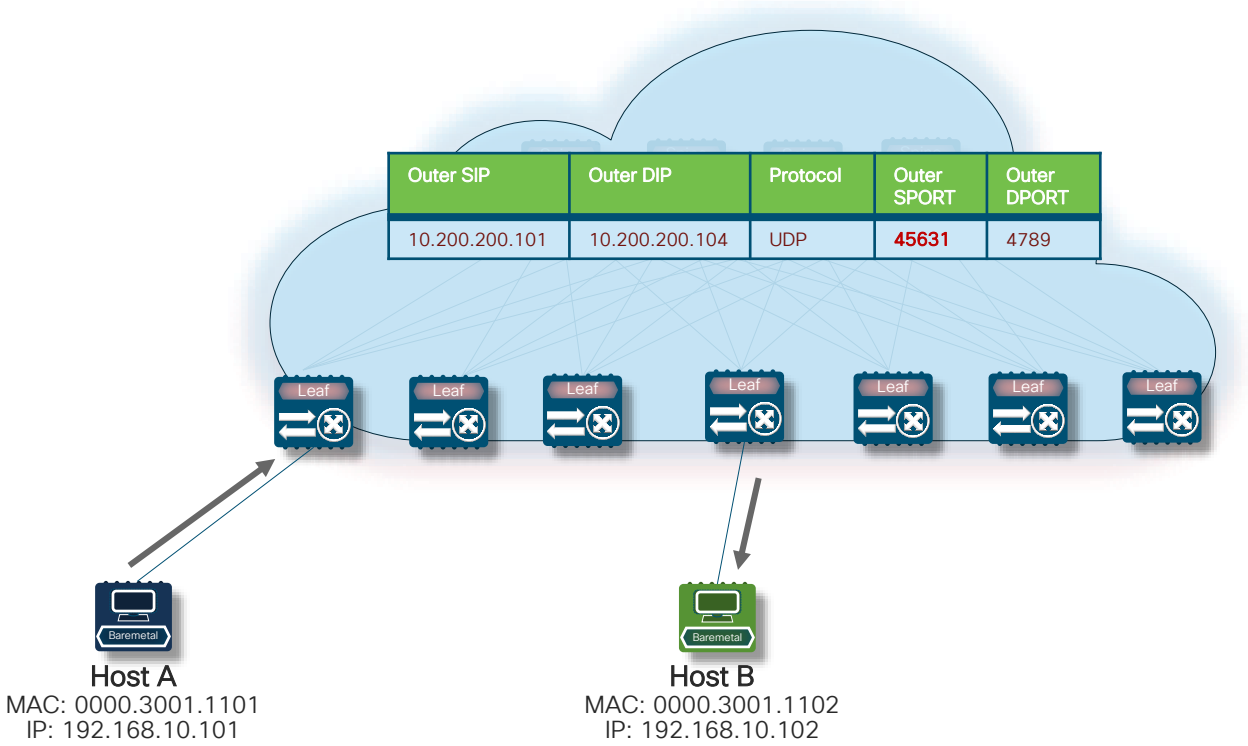



```
tc@boxA:$ ping HostB
PING HostB (192.168.10.102): 56 data bytes
64 bytes from 192.168.10.102 : icmp_seq=0 ttl=64 time=0.653 ms
64 bytes from 192.168.10.102 : icmp_seq=1 ttl=64 time=0.631 ms

--- 192.168.10.102 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss

tc@boxA:$
```

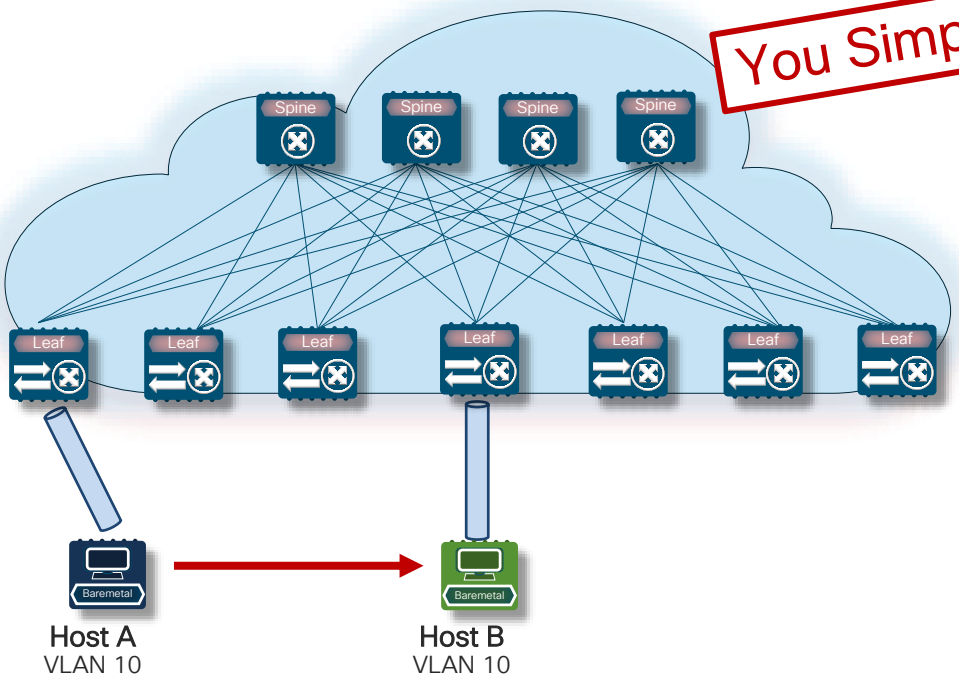
Overlay Forwarding – A Bridged Ping to Host B



Overlay Forwarding to B – Seen by the Underlay



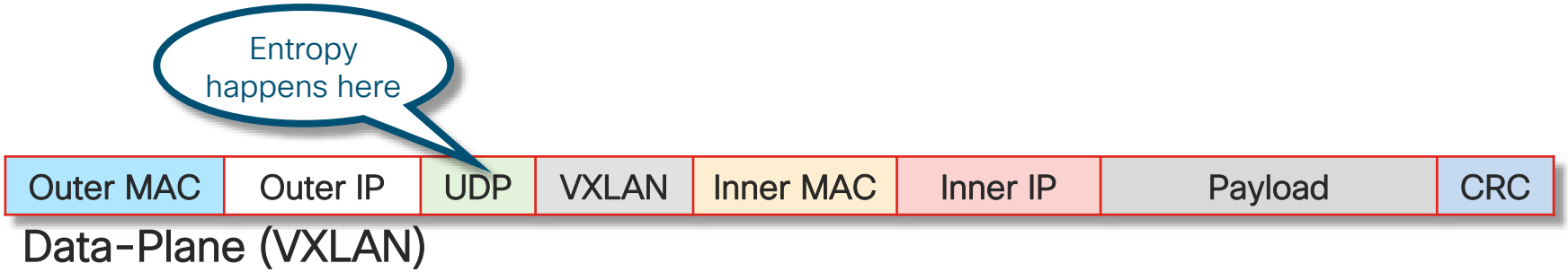
Overlay Forwarding – Which Underlay Path?



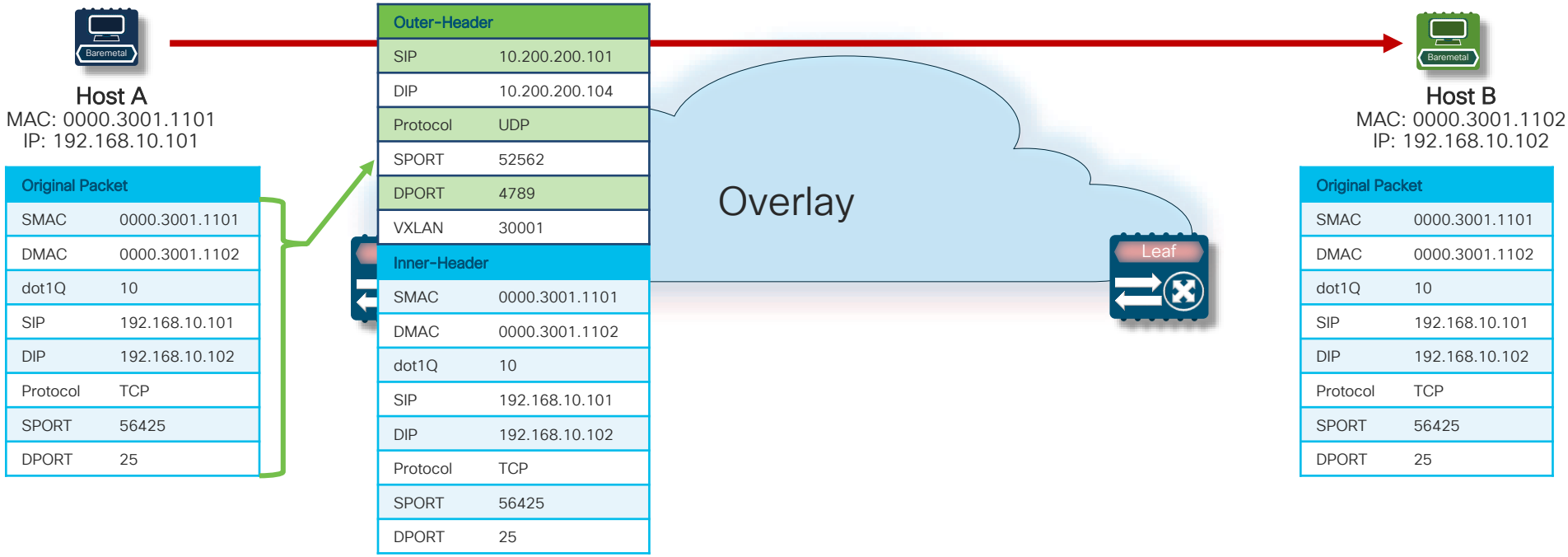
You Simply Don't Know

- Network Virtualization Overlay (NVO) – Network Abstraction from Physical Connectivity
- **Entropy** – Overlays provide Path Diversity across Equal Cost Multipath (ECMP)
- **Correlation** – Overlays don't present or represent the Underlay

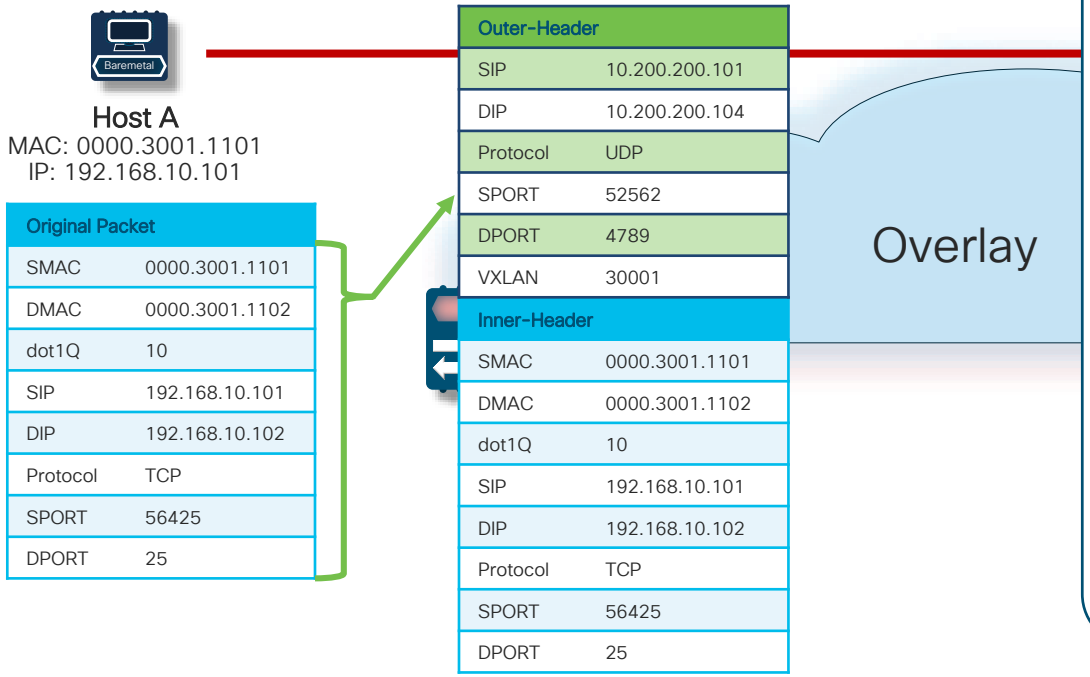
Entropy – the VXLAN Example



Path Diversity – Overlay Example (VXLAN)

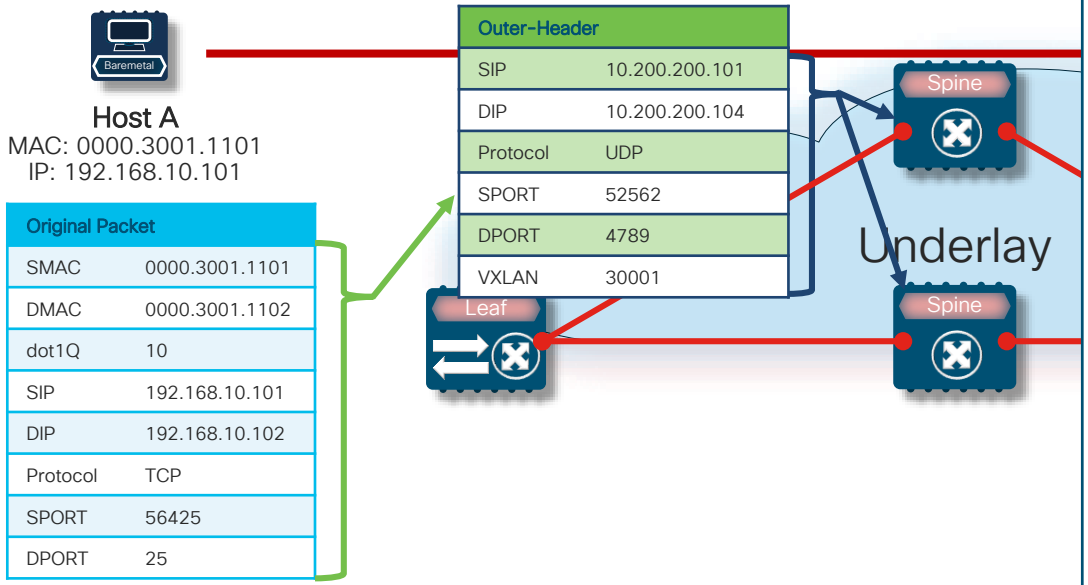


Path Diversity – Overlay Example (VXLAN)



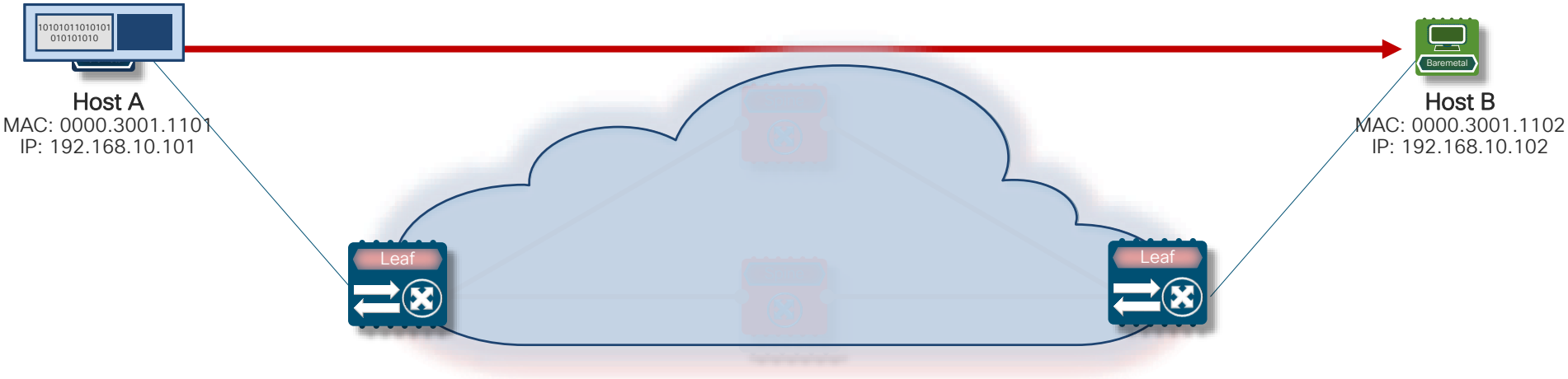
- VXLAN Entropy is based on 5-Tuple plus MAC addresses
  - Layer-2 Source MAC (SMAC)
  - Layer-2 Destination MAC (DMAC)
  - Layer-3 Source IP (SIP)
  - Layer-3 Destination IP (DIP)
  - IP Protocol
  - Protocol Source Port (SPORT)
  - Protocol Destination Port (DPORT)
- Some Platforms include IEEE 802.1q information (VLAN ID)
- Various Hashing Algorithms

Path Diversity – Underlay Example (Layer-3 ECMP)

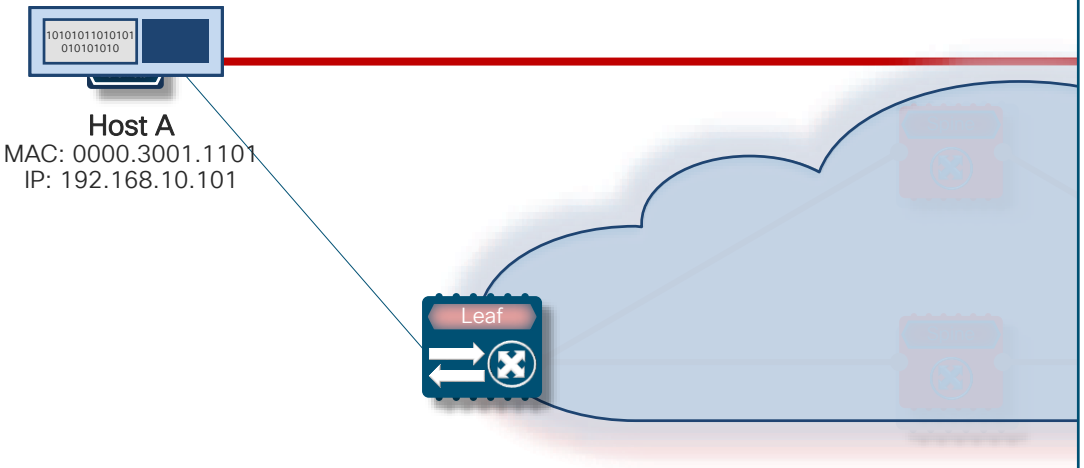


- Layer-3 ECMP Hashing is typically based on 5-Tuple
  - Layer-3 Source IP (SIP)
  - Layer-3 Destination IP (DIP)
  - IP Protocol
  - Protocol Source Port (SPORT)
  - Protocol Destination Port (DPORT)
- Platform specific

Entropy – the VXLAN Example



Entropy – the VXLAN Example

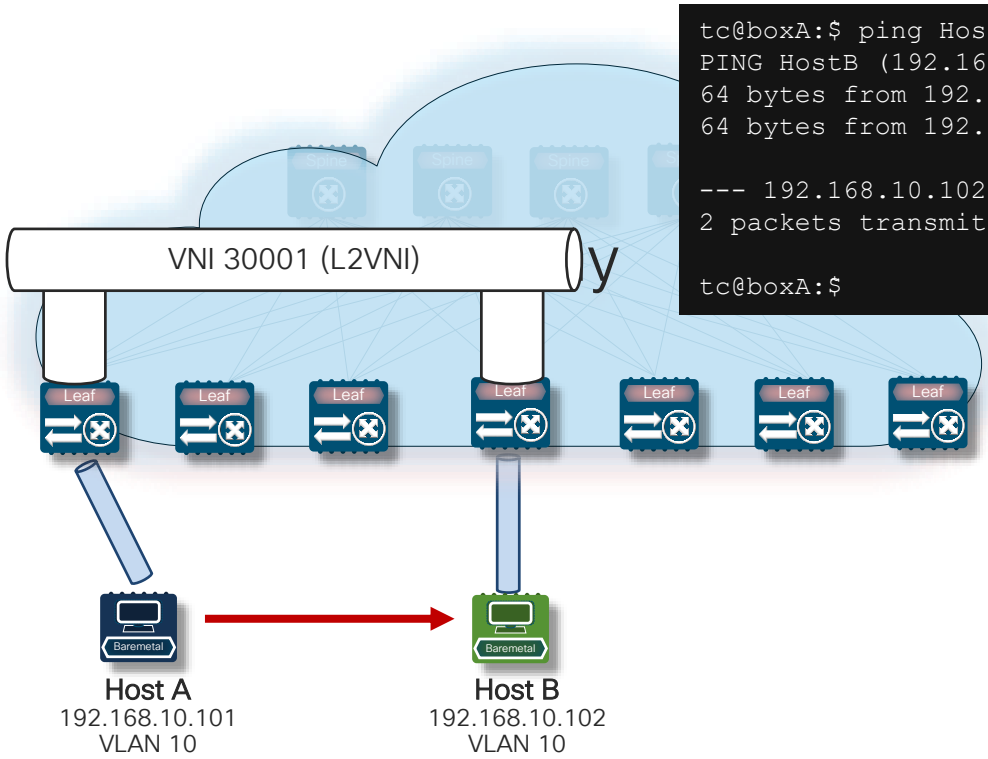


- VXLAN uses variable UDP Source Port in Outer Header
- Hash of the inner Layer-2/Layer-3/Layer-4 Headers of the original Ethernet Frame
- Enables entropy for ECMP Load balancing in the Network

Agenda

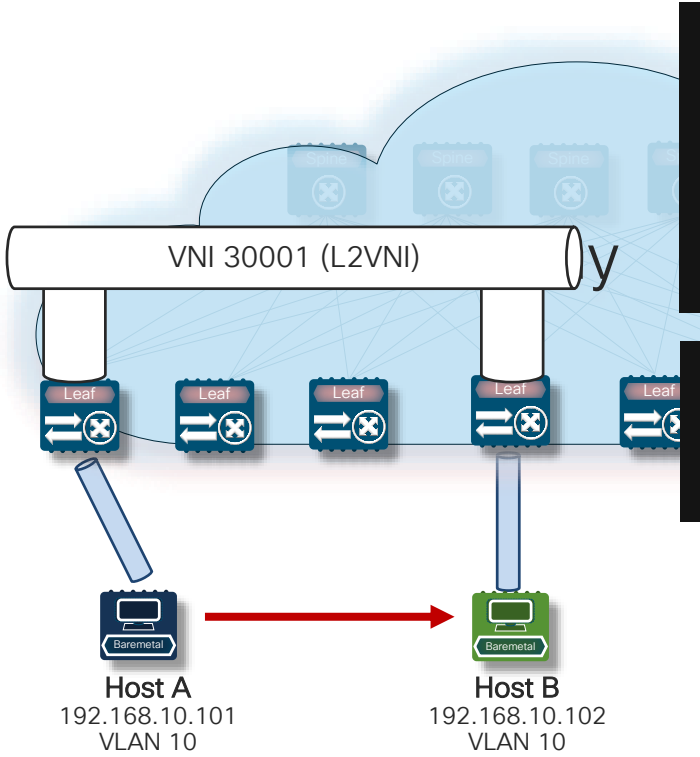
- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

Ping/Traceroute in the Overlay – Bridged




```
tc@boxA:$ ping HostB
PING HostB (192.168.10.102): 56 data bytes
64 bytes from 192.168.10.102 : icmp_seq=0 ttl=64 time=0.653 ms
64 bytes from 192.168.10.102 : icmp_seq=1 ttl=64 time=0.631 ms

--- 192.168.10.102 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss

tc@boxA:$
```

Ping/Traceroute in the Overlay – Bridged

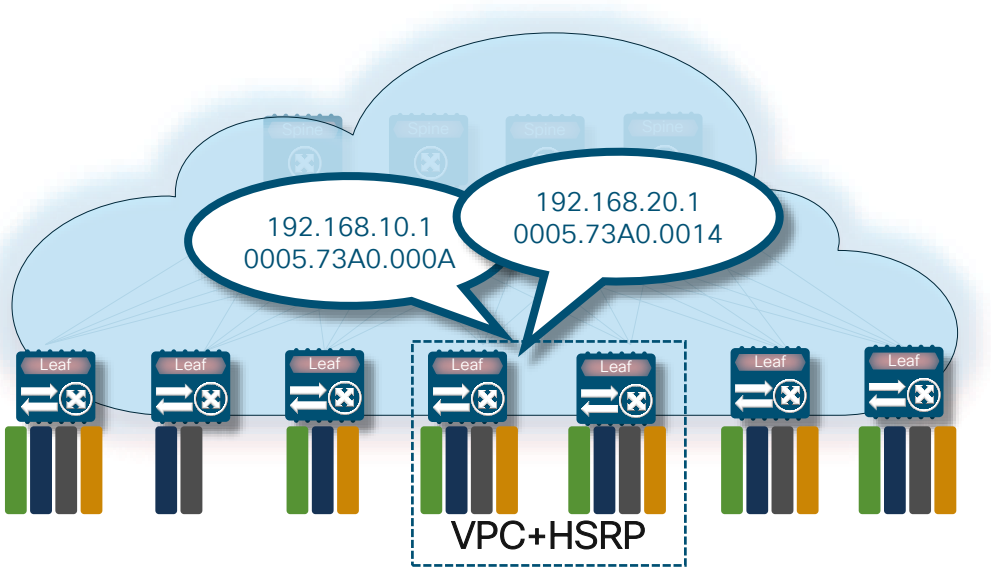



```
tc@boxA:$ ping HostB
PING HostB (192.168.10.102): 56 data bytes
64 bytes from 192.168.10.102 : icmp_seq=0 ttl=64 time=0.653 ms
64 bytes from 192.168.10.102 : icmp_seq=1 ttl=64 time=0.631 ms

--- 192.168.10.102 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss

tc@boxA:$
```

```
tc@boxA:$ traceroute HostB
traceroute to HostB (192.168.10.102), 30 hops max, 38 byte packets
1 192.168.10.102 (192.168.10.102)  0.302 ms  0.251 ms  0.259 ms

tc@boxA:$
```

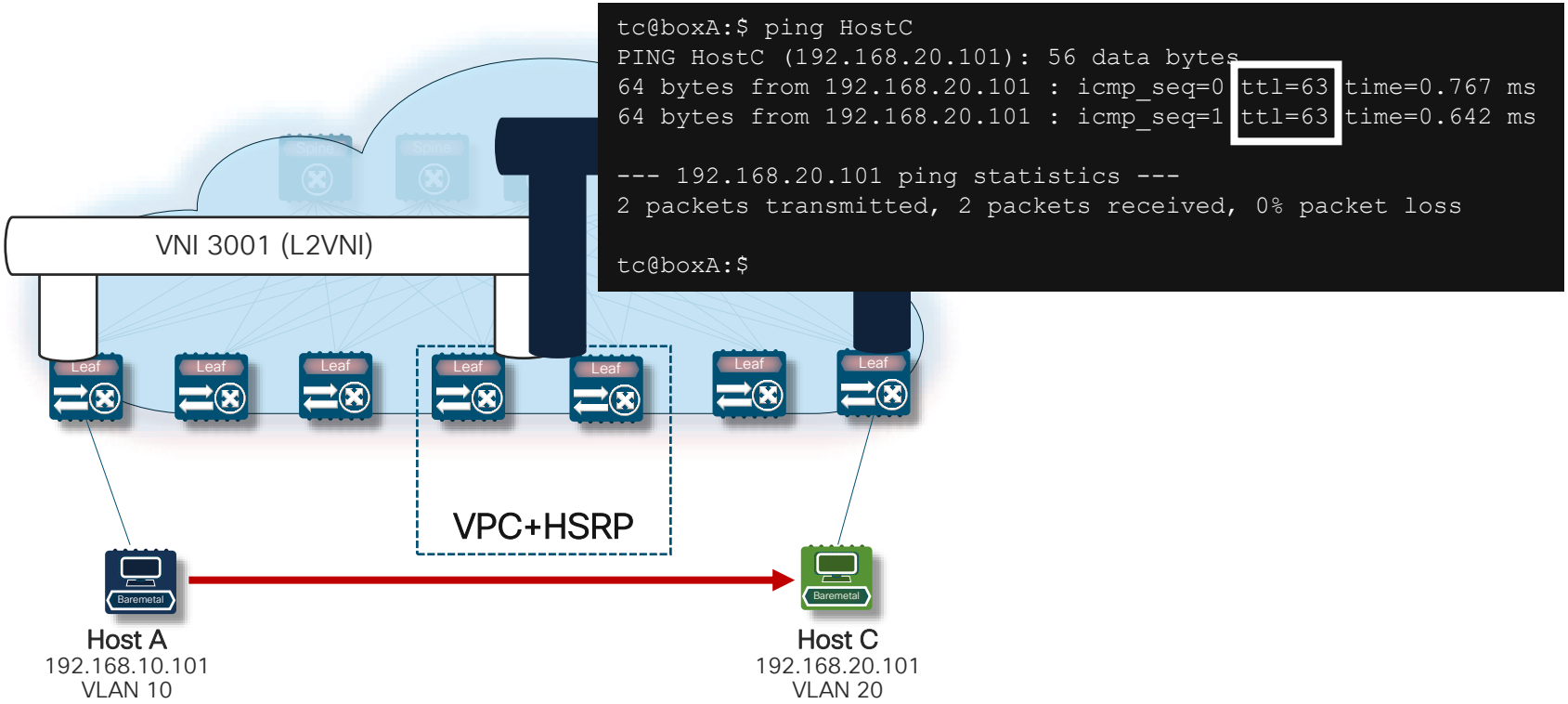
Centralized IP Gateway – VXLAN Flood & Learn

FYI

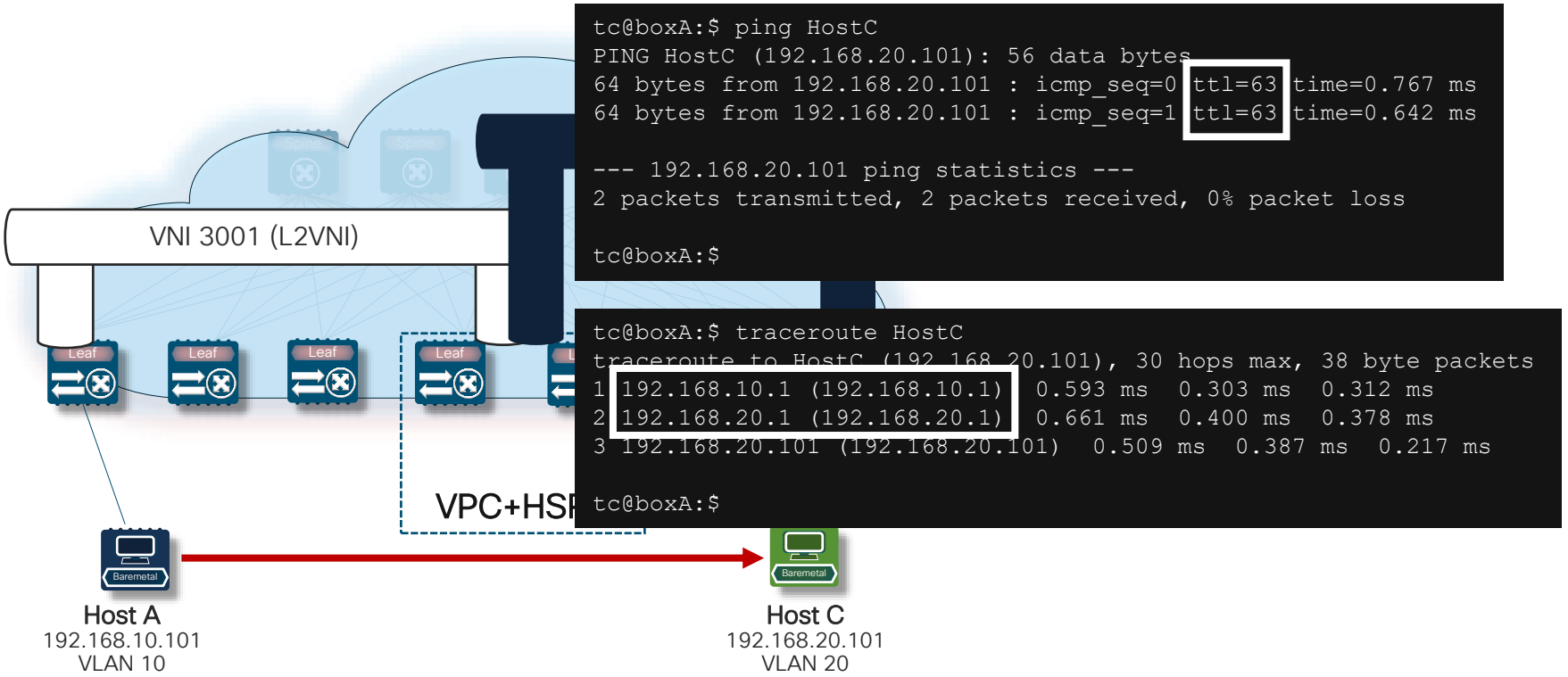


- Centralized First-Hop Routing on a Set of Devices
  - First-Hop Redundancy Protocol (FHRP) Approach – i.e. HSRP, VRRP
- Gateway is always active?!
  - Depends on VPC
- Centralized, Inter-VLAN/VNI Routing
  - Centralized MAC & ARP State
  - Large Configuration State at GW

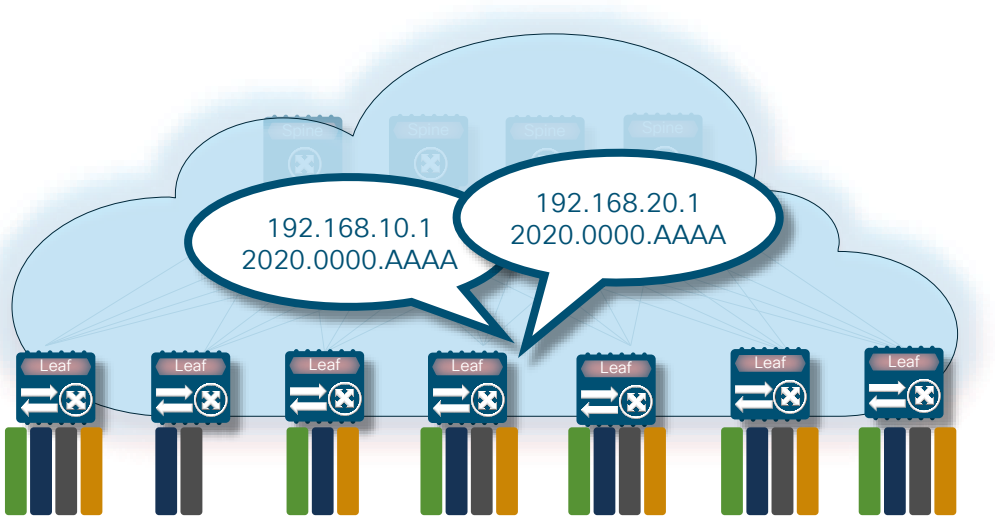
Ping/Traceroute in the Overlay – Routed with HSRP



Ping/Traceroute in the Overlay – Routed with HSRP

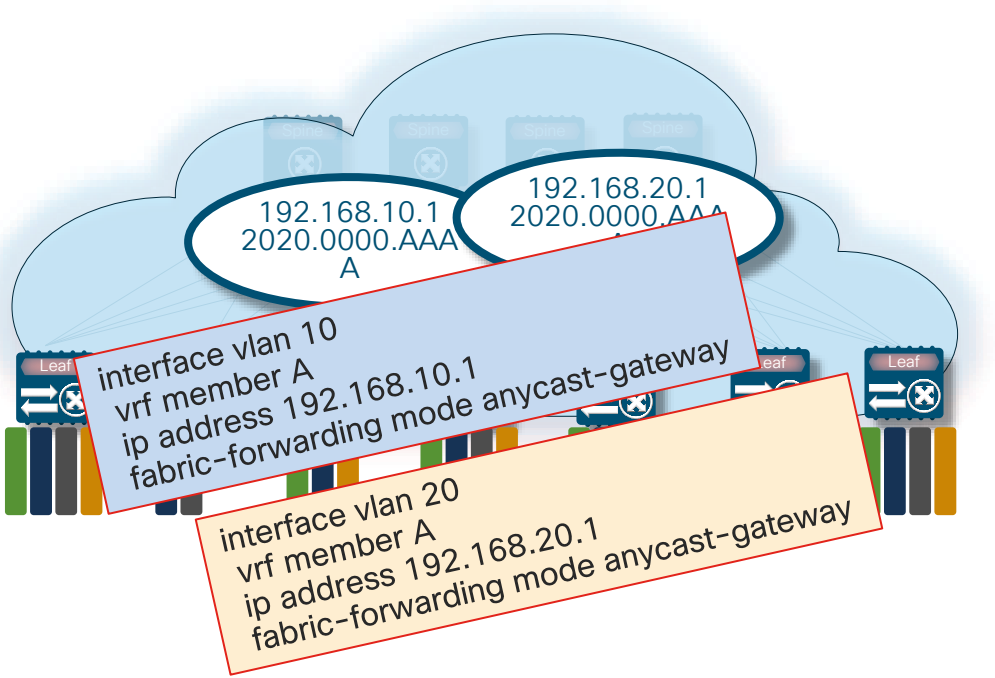


Distributed IP Anycast Gateway – VXLAN EVPN



- Distributed First-Hop Routing on Edge Device
  - All Edge Device share same Gateway IP and MAC address
  - Pervasive Gateway approach
- Gateway is always active
  - No redundancy protocol for hello or state exchange
- Distributed and smaller state
  - Only local Endpoints ARP entries

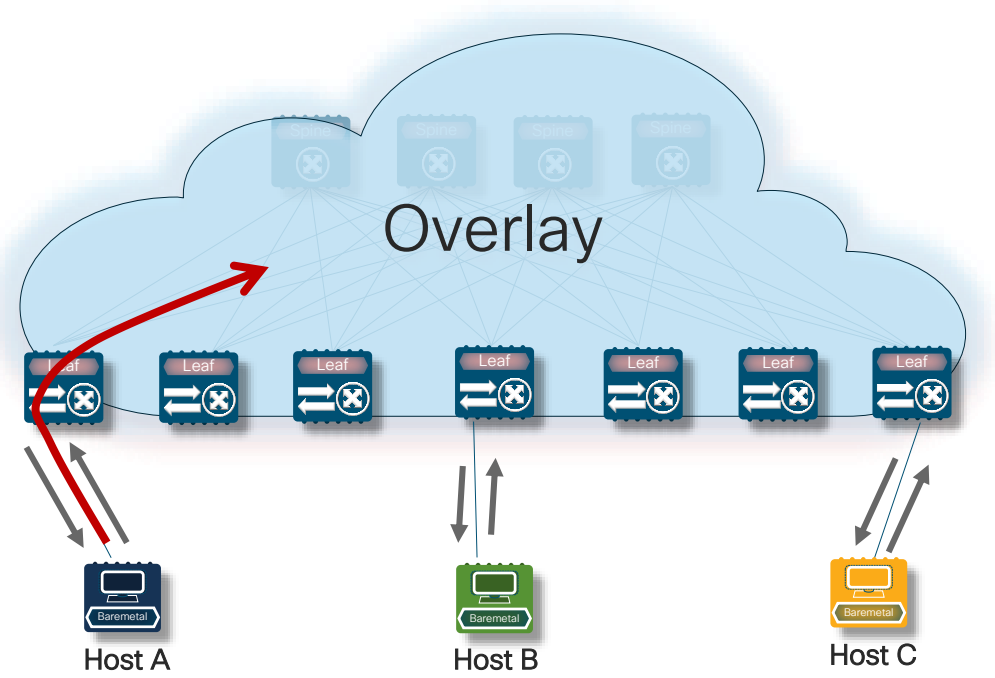
Distributed IP Anycast Gateway – VXLAN EVPN



- Distributed First-Hop Routing on Edge Device
  - All Edge Device share same Gateway IP and MAC address
  - Pervasive Gateway approach
- Gateway is always active
  - No redundancy protocol for hello or state exchange
- Distributed and smaller state
  - Only local Endpoints ARP entries

Anycast – One-to-Nearest Association

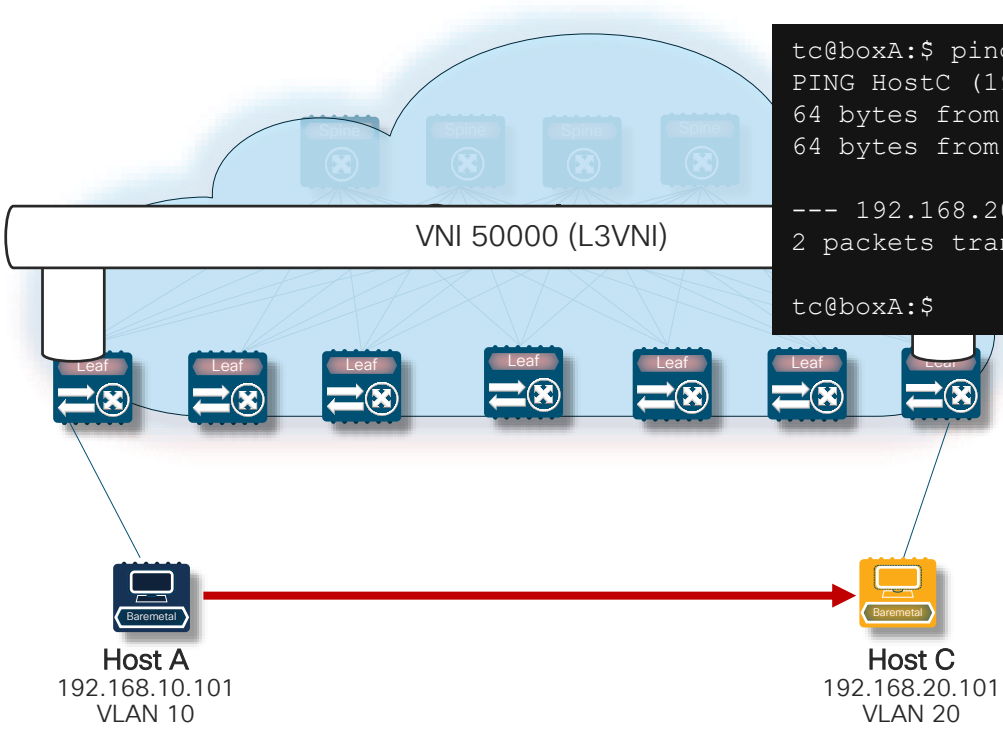
FYI



- Network Addressing and Routing Methodology
- Datagrams sent from a single Sender to the Topologically Nearest Node
- Group of potential Receivers, all identified by the same Destination Address

*L3VNI: VNI for all Routing operation (”VRF-VNI”)

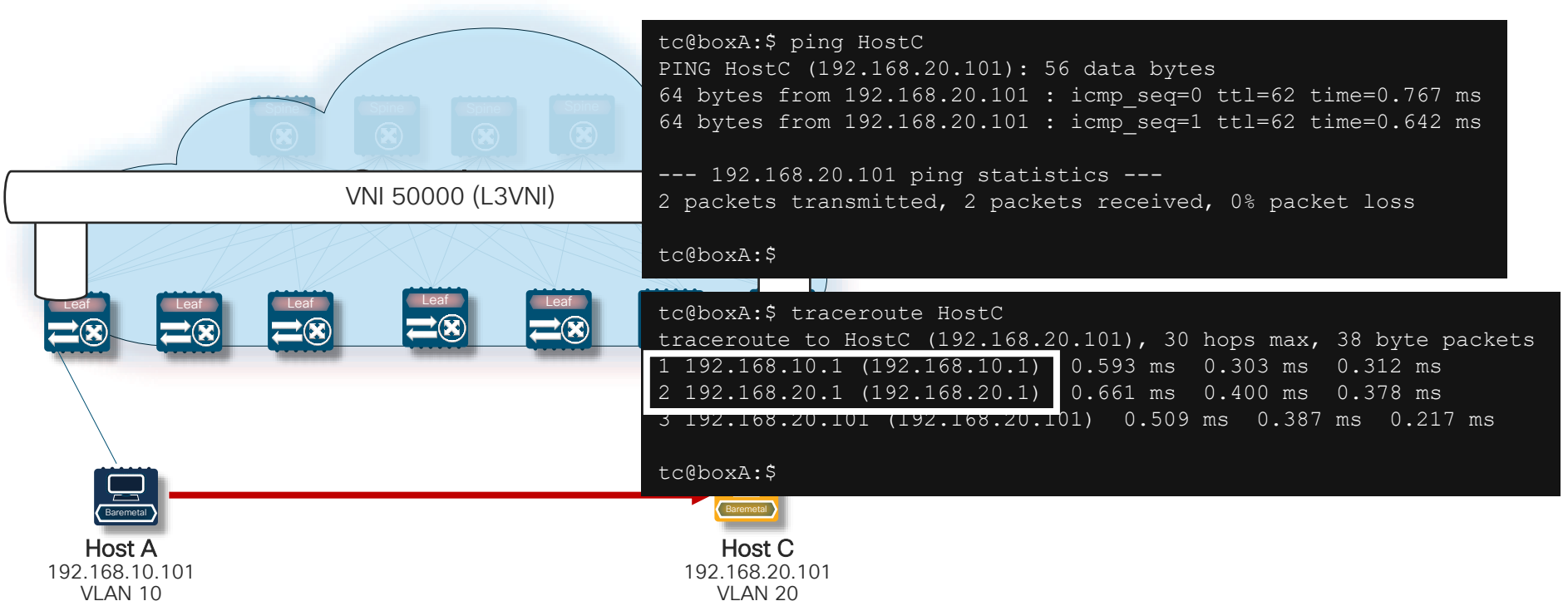
Ping/Traceroute in the Overlay – Routed with EVPN




```
tc@boxA:$ ping HostC
PING HostC (192.168.20.101): 56 data bytes
64 bytes from 192.168.20.101 : icmp_seq=0 ttl=62 time=0.767 ms
64 bytes from 192.168.20.101 : icmp_seq=1 ttl=62 time=0.642 ms

--- 192.168.20.101 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss

tc@boxA:$
```

*L3VNI: VNI for all Routing operation ("VRF-VNI")

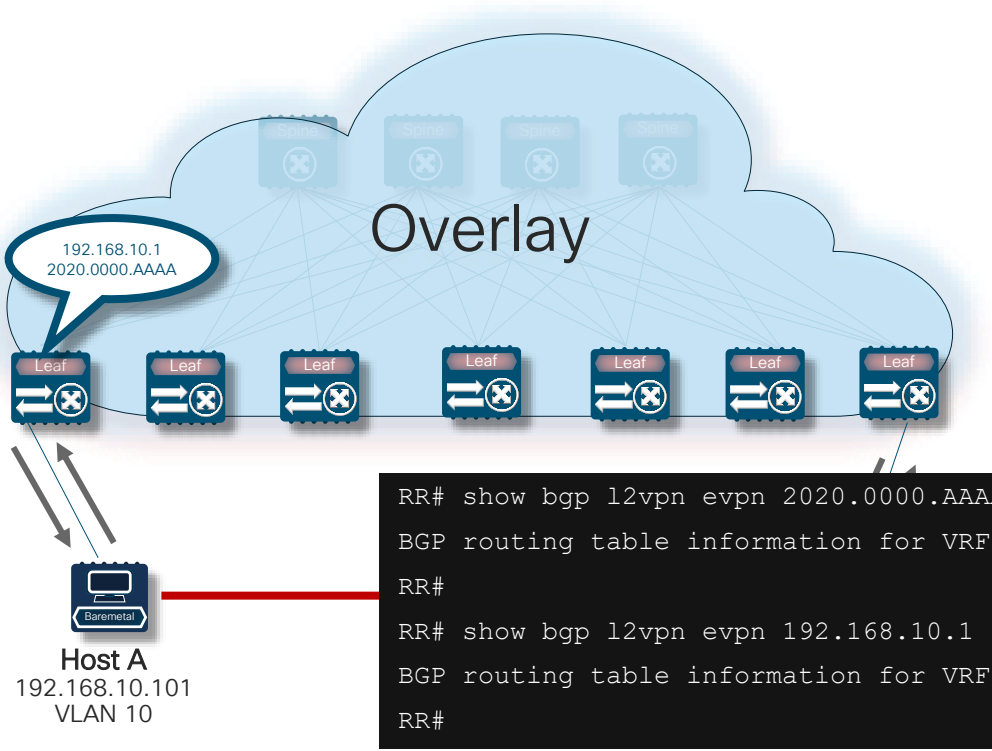
Ping/Traceroute in the Overlay – Routed with EVPN




```
tc@boxA:$ ping HostC
PING HostC (192.168.20.101): 56 data bytes
64 bytes from 192.168.20.101 : icmp_seq=0 ttl=62 time=0.767 ms
64 bytes from 192.168.20.101 : icmp_seq=1 ttl=62 time=0.642 ms

--- 192.168.20.101 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss

tc@boxA:$
```

```
tc@boxA:$ traceroute HostC
traceroute to HostC (192.168.20.101), 30 hops max, 38 byte packets
1 192.168.10.1 (192.168.10.1)  0.593 ms  0.303 ms  0.312 ms
2 192.168.20.1 (192.168.20.1)  0.661 ms  0.400 ms  0.378 ms
3 192.168.20.101 (192.168.20.101)  0.509 ms  0.387 ms  0.217 ms

tc@boxA:$
```

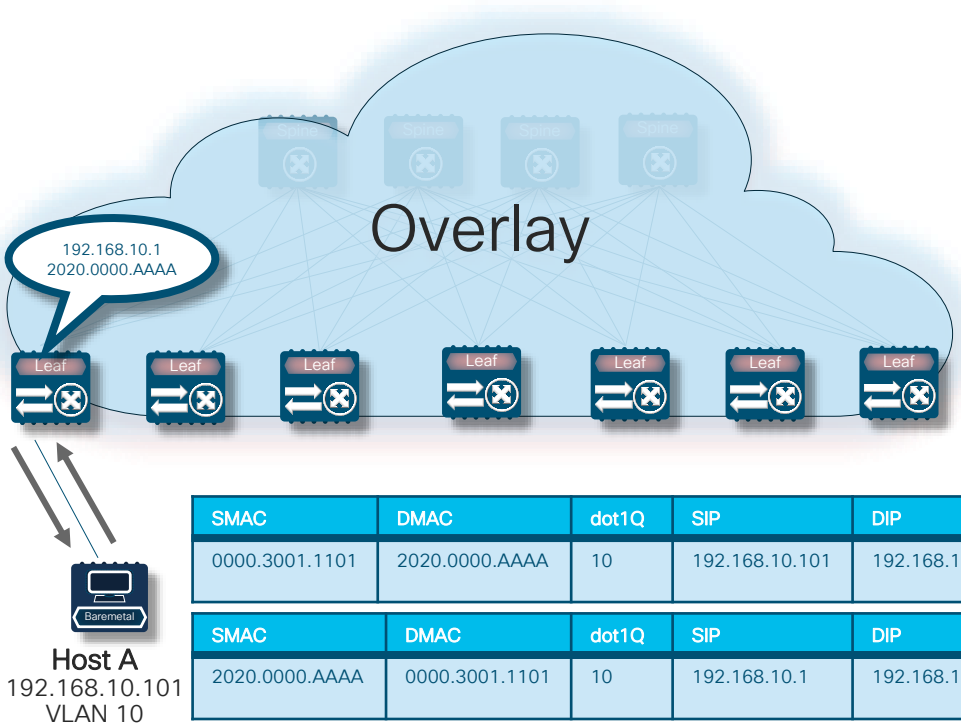
Distributed Anycast Gateway and Ping



- Distributed Anycast Gateway only exists towards LAN Segment
  - The Distributed Anycast Gateway MAC and IP is not advertised over BGP EVPN

```
RR# show bgp l2vpn evpn 2020.0000.AAAA
BGP routing table information for VRF default, address family L2VPN EVPN
RR#
RR# show bgp l2vpn evpn 192.168.10.1
BGP routing table information for VRF default, address family L2VPN EVPN
RR#
```

Ping To Distributed Anycast Gateway – Local



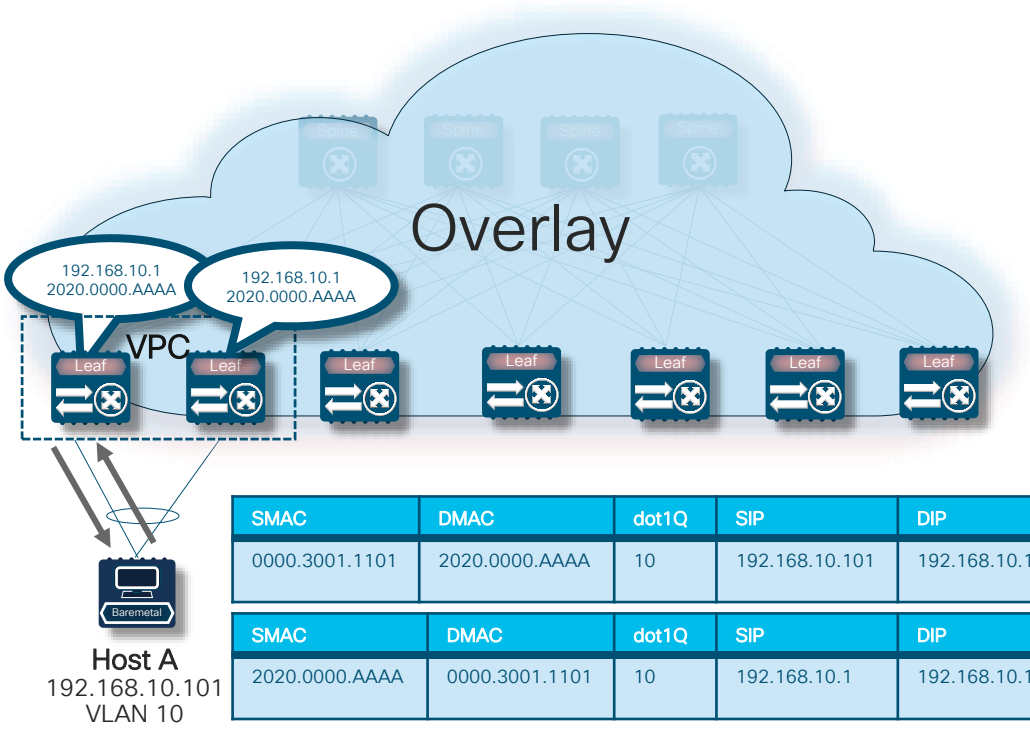
SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3001.1101	2020.0000.AAAA	10	192.168.10.101	192.168.10.1	Echo Request

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3001.1101	10	192.168.10.1	192.168.10.101	Echo Reply

OK

- Simple Ping on LAN Segment
- Distributed Anycast Gateway is represented by an SVI (Switch Virtual Interface)

Ping To Distributed Anycast Gateway – VPC



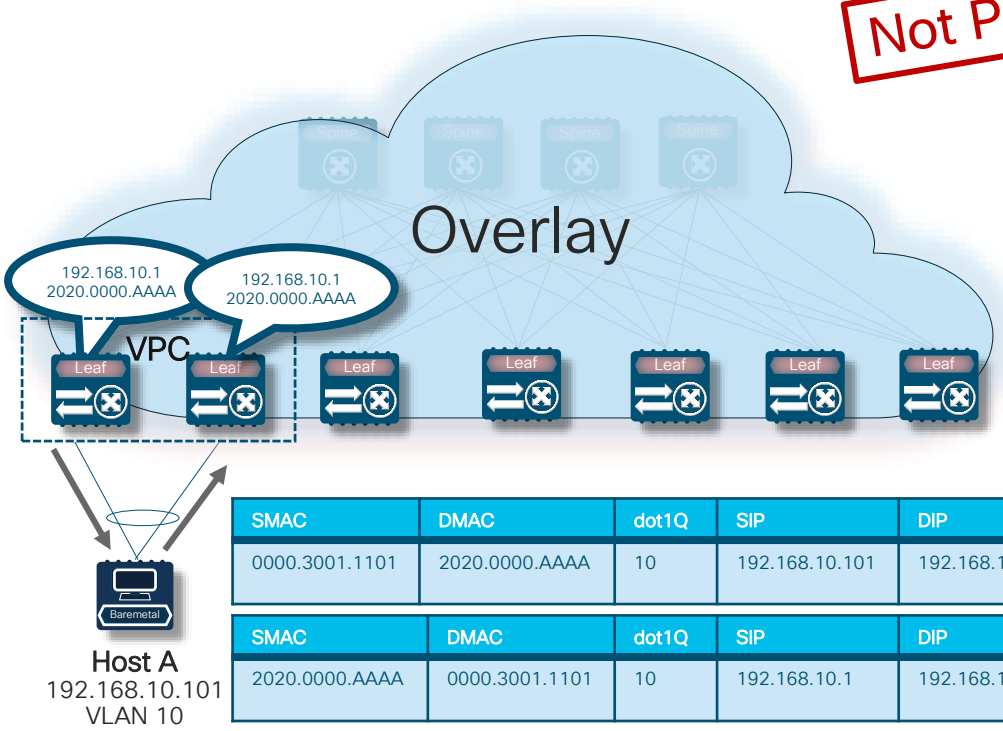
SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3001.1101	2020.0000.AAAA	10	192.168.10.101	192.168.10.1	Echo Request

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3001.1101	10	192.168.10.1	192.168.10.101	Echo Reply

OK

- Simple Ping on LAN Segment
- Distributed Anycast Gateway is represented by an SVI (Switch Virtual Interface)
- Port-Channel hashing decides to which Switch to go

Ping From Distributed Anycast Gateway – VPC



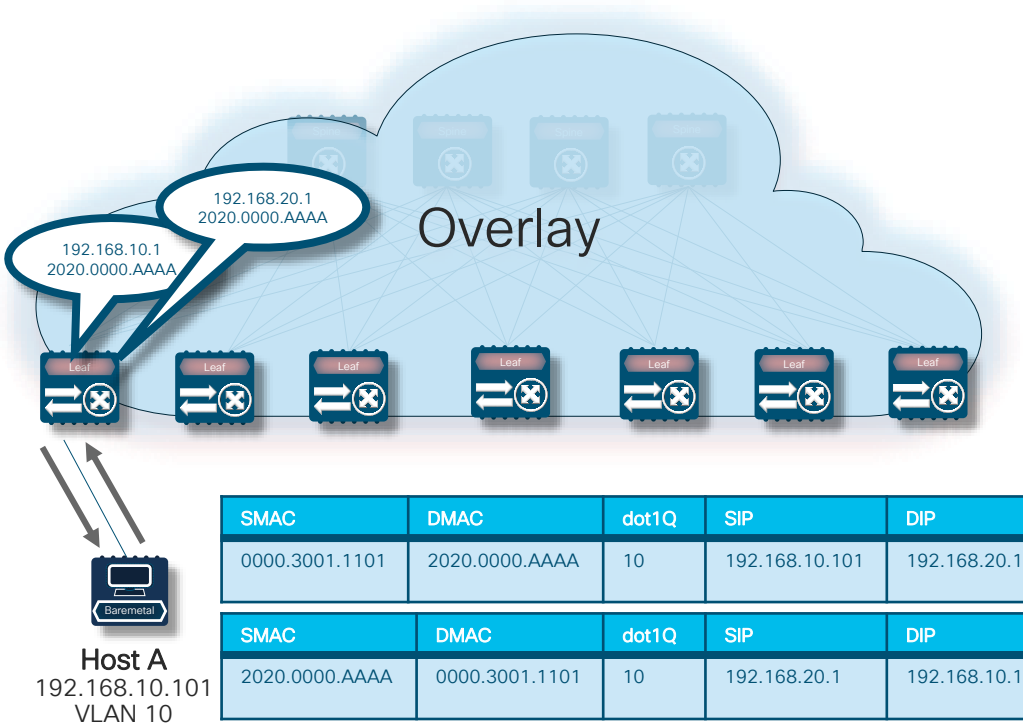
Not Predictable

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3001.1101	2020.0000.AAAA	10	192.168.10.101	192.168.10.1	Echo Reply

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3001.1101	10	192.168.10.1	192.168.10.101	Echo Request

- Ping from Distributed Anycast Gateway (SVI) to Local Host
- From Switch to Endpoint; local virtual Port-Channel (VPC) interface is used
- From Endpoint to Leaf; Port-Channel hashing decides what Interface to use
  - No response if Source and Destination Switch is different

Ping To Distributed Anycast Gateway – Local



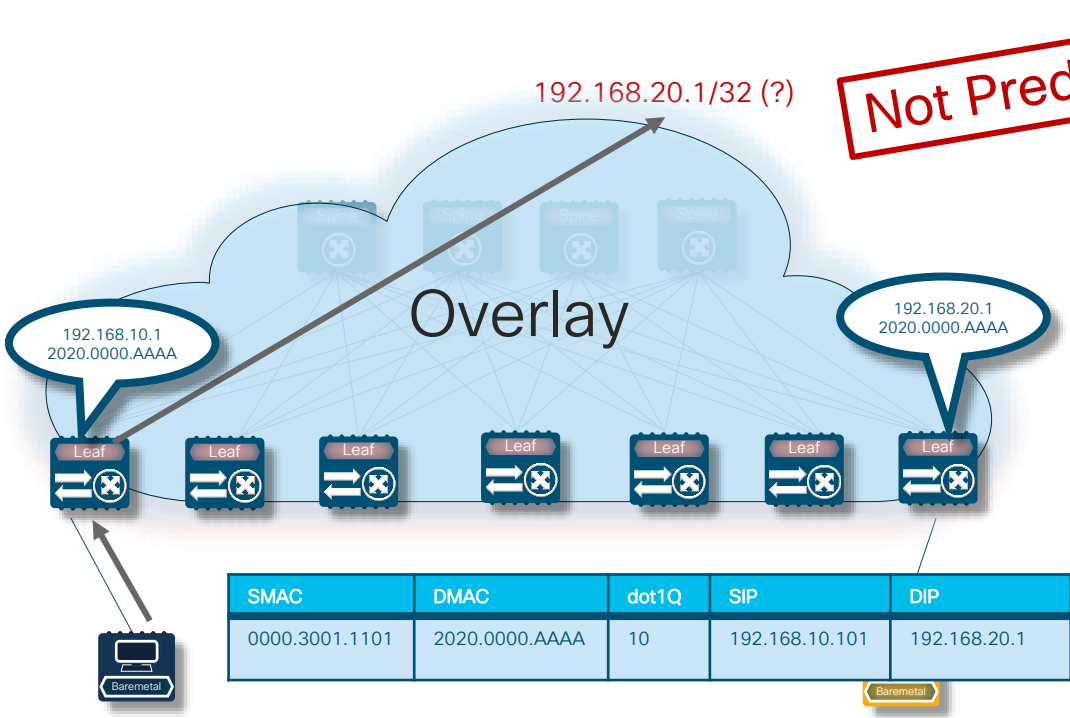
SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3001.1101	2020.0000.AAAA	10	192.168.10.101	192.168.20.1	Echo Request

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3001.1101	10	192.168.20.1	192.168.10.101	Echo Reply

OK

- Simple Ping on LAN Segment across IP Subnet
- Distributed Anycast Gateway is represented by an SVI (Switch Virtual Interface)
- Local Inter-VLAN Routing to reach Local SVI

Ping To Distributed Anycast Gateway – Remote

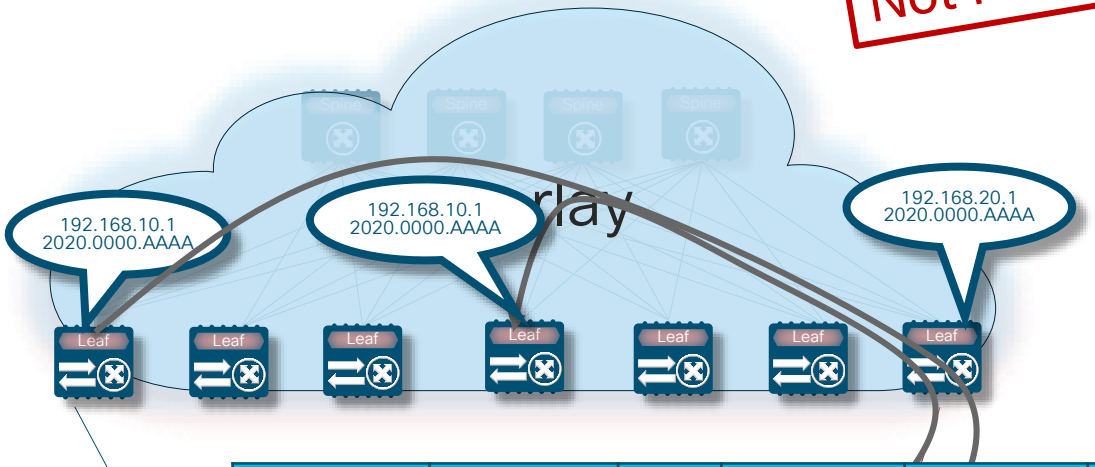


SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3001.1101	2020.0000.AAAA	10	192.168.10.101	192.168.20.1	Echo Request

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3001.1101	10	192.168.20.1	192.168.10.101	Echo Reply

- Ping into the Blue
- Distributed Anycast Gateway can be on any Switch
- Distributed Anycast Gateway is only advertised as IP Subnet (direct route)
- VXLAN Entropy and ECMP Hashing defines Destination

Ping From Distributed Anycast Gateway – Remote



SMAC	DMAC	dot1Q	SIP	DIP	ICMP
2020.0000.AAAA	0000.3002.2101	10	192.168.10.1	192.168.20.101	Echo Request

Baremetal

SMAC	DMAC	dot1Q	SIP	DIP	ICMP
0000.3002.2101	2020.0000.AAAA	10	192.168.20.101	192.168.10.1	Echo Reply

Not Predictable

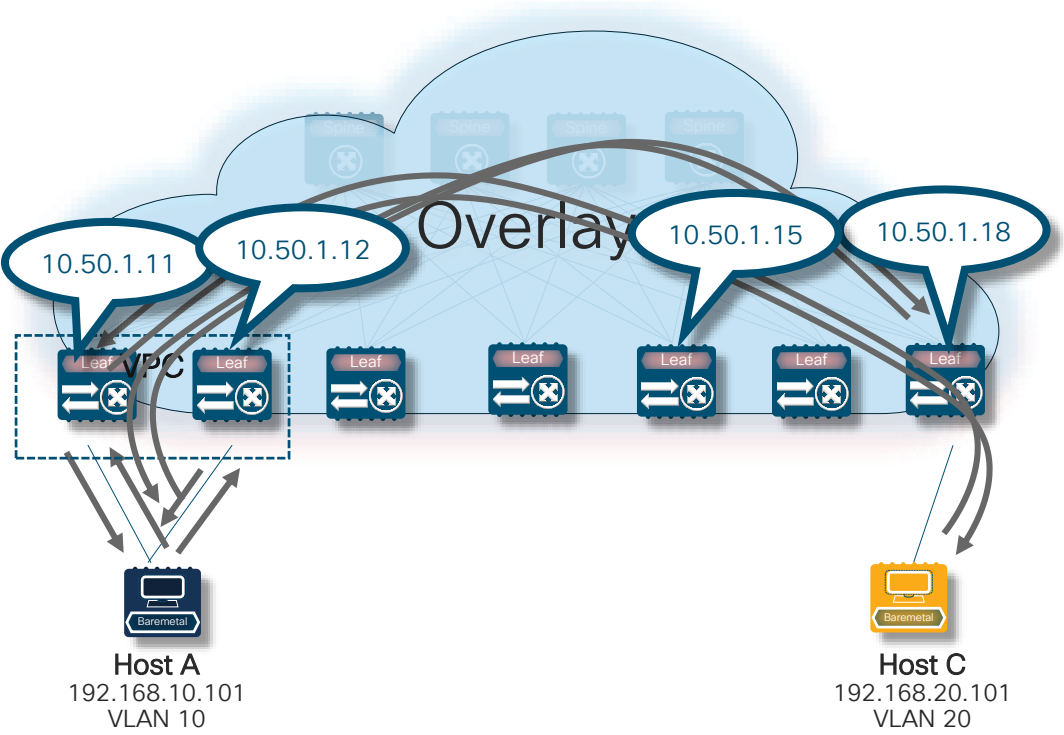
- Ping from Distributed Anycast Gateway (SVI) to Remote Host
- From Switch to Endpoint; shortest path
- From Endpoint to Leaf; response could go to any Switch with Distributed Anycast Gateway IP Address
  - No response if Source and Destination Switch is different

Problem – Predictable Overlay Connectivity Test

- Endpoint to SVI Connectivity Tests are not consistent with ”traditional” Networking
  - Nature of Anycast IP Addressing
  - Overlay Entropy
  - ECMP Hashing
  - Port-Channel Hashing

From	To	Comment	Result	Predictability
Endpoint	SVI	Local Switch Same Subnet	OK	100%
SVI	Endpoint	Local Switch Same Subnet	OK	100%
Endpoint	SVI	VPC Port-Channel Same Subnet	OK	100%
SVI	Endpoint	VPC Port-Channel Same Subnet	Not OK	<50%
Endpoint	SVI	Local Switch Different Subnet	OK	100%
SVI	Endpoint	Local Switch Different Subnet	OK	100%
Endpoint	SVI	Remote Switch Different Subnet	OK	100%
SVI	Endpoint	Remote Switch Different Subnet	Not OK	<50%

Ping From/To Loopback – Local with/without VPC



- Simple Ping to a Routed IP Address
- Uses Distributed Anycast Gateway to Reach Local or Remote Loopbacks
- Allows most Flexible Connectivity Tests

```
interface loopback10
  vrf member BLUE
  ip address 10.50.1.L#/32 tag 12345
```

Solution – Predictable Overlay Connectivity Test

- Avoid False Positives
  - Create a per-VRF Loopback
  - Execute Connectivity Tests against the Loopback
- Loopback Connectivity Tests – 100% Predictable in any Case
  - Endpoint to Loopback
  - Loopback to Endpoint
  - Loopback to Loopback
- A Loopback can Save your Day!

From	To	Comment	Result	Predictability
Endpoint	Loopback	Local Switch Same Subnet	OK	100%
Loopback	Endpoint	Local Switch Same Subnet	OK	100%
Endpoint	Loopback	VPC Port-Channel Same Subnet	OK	100%
Loopback	Endpoint	VPC Port-Channel Same Subnet	OK	100%
Endpoint	Loopback	Local Switch Different Subnet	OK	100%
Loopback	Endpoint	Local Switch Different Subnet	OK	100%
Endpoint	Loopback	Remote Switch Different Subnet	OK	100%
Loopback	Endpoint	Remote Switch Different Subnet	OK	100%

Operations, Administration, and Management (OAM)

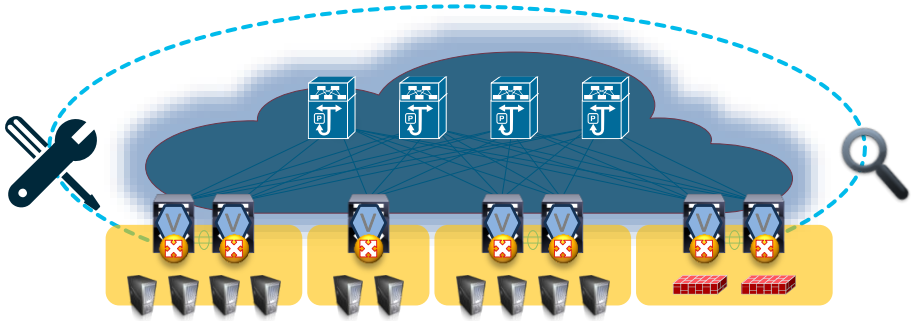
Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

Operations, Administration, and Management (OAM)

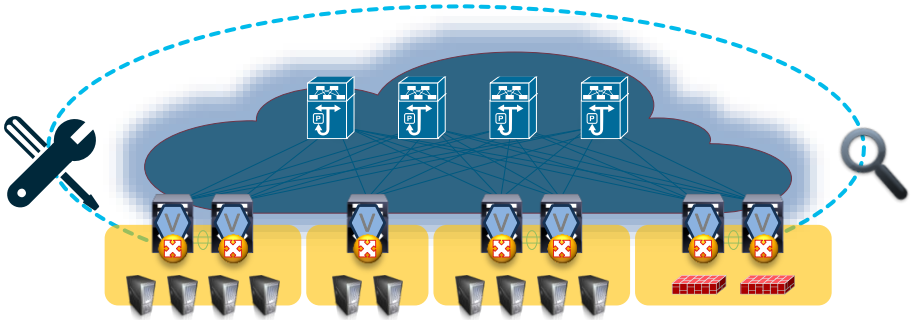
- OAM – processes, activities, tools and standards
- Various Modes of Operation
- Pro-Active
  - Controlling a Situation
- Re-Active
  - Responding to a Situation

VXLAN OAM – Re-Active



Ping / Path MTU	Pathtrace
• Check liveliness of End-Host • Option to specify Payload Parameters	• Trace paths to End-Host and Tunnel-Endpoint • Get Path, Interface and Error statistics along path • Specify Payload Parameters for Path Selection

VXLAN OAM – Pro-Active



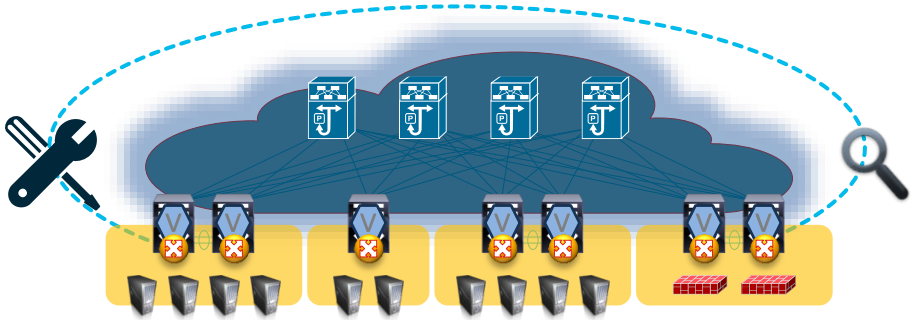
Endpoint Locator

- Locate End-Host and Segment Identifier
- Track History of End-Host
- Provide Fabric Host-Count and Activity

Pro-Active Monitoring

- Proactive Monitoring with Threshold and State Notifications

VXLAN OAM – OAM Model of Operation



Endpoint Locator	Ping / Path MTU	Pathtrace	Pro-Active Monitoring
• Locate End-Host and Segment Identifier • Track History of End-Host • Provide Fabric Host-Count and Activity	• Check liveliness of End-Host • Option to specify Payload Parameters	• Trace paths to End-Host and Tunnel-Endpoint • Get Path, Interface and Error statistics along path • Specify Payload Parameters for Path Selection	• Proactive Monitoring with Threshold and State Notifications

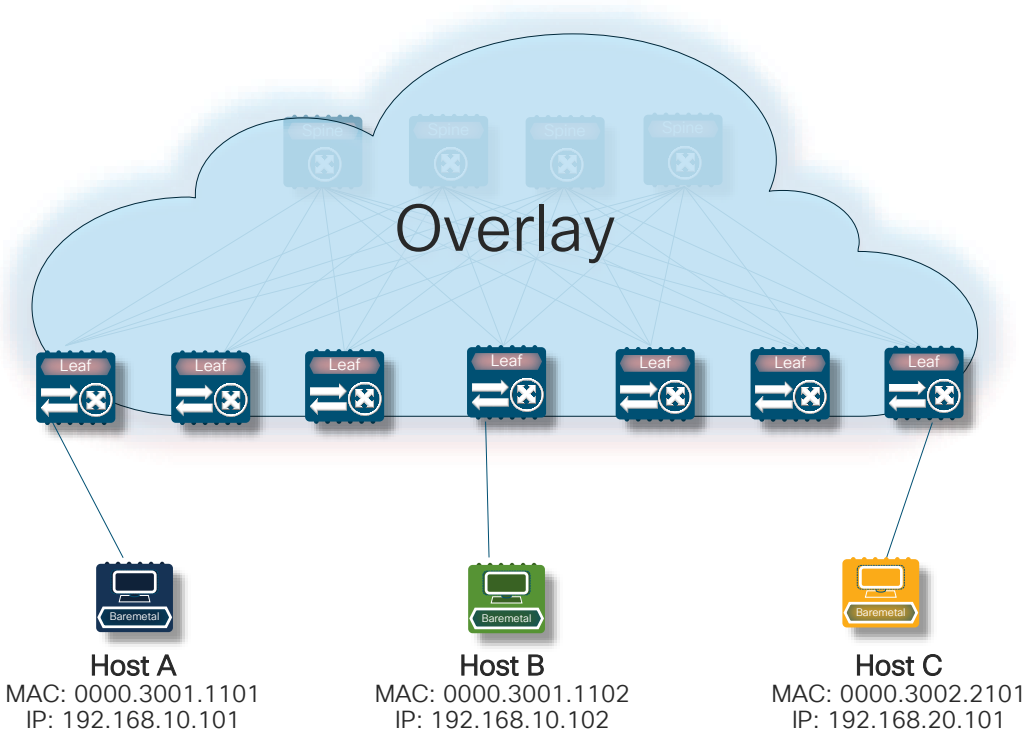
Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

NGOAM or VXLAN OAM

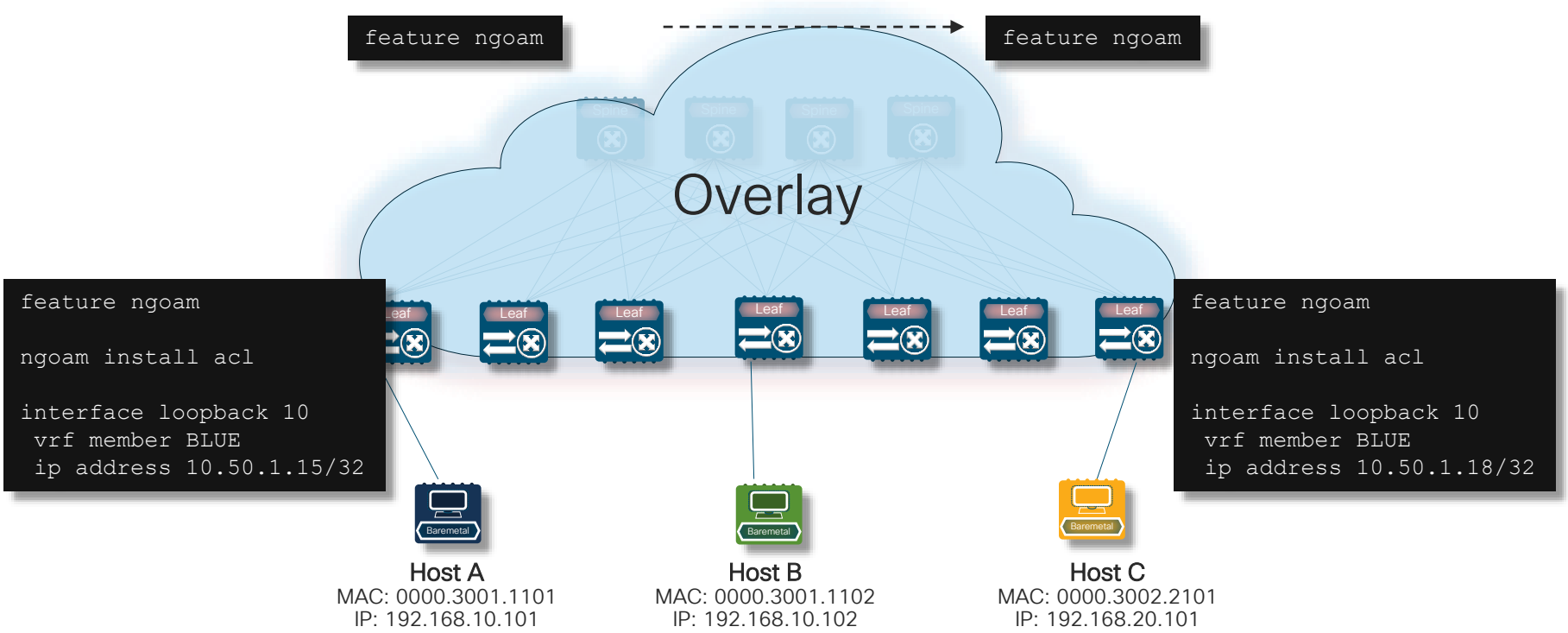
- Next Generation OAM for Data Center Fabrics
- Running on Nexus 9000, Nexus 7000, Nexus 5600, Nexus 3000
  - VXLAN Today
  - All IP Tomorrow
- Various Methods to Execute and Retrieve Data
  - Command Line Interface (CLI)
  - NX-API
  - DCNM (using NX-API)

VXLAN OAM – Pre-Requisites

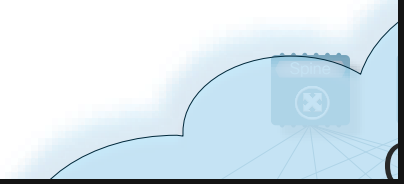


- Enable “feature ngoam”
  - Required on VTEPs and intermediate Devices (i.e. Spines)
- Activate OAM filters
  - Configure “ngoam install acl”
  - Required on all VTEPs
- Have a Loopback with Unique IP
  - Source for Loopback Messages
  - VRF-aware (per VRF Loopback)
  - IP must be reachable in Overlay

VXLAN OAM – Pre-Requisites (Command Line)



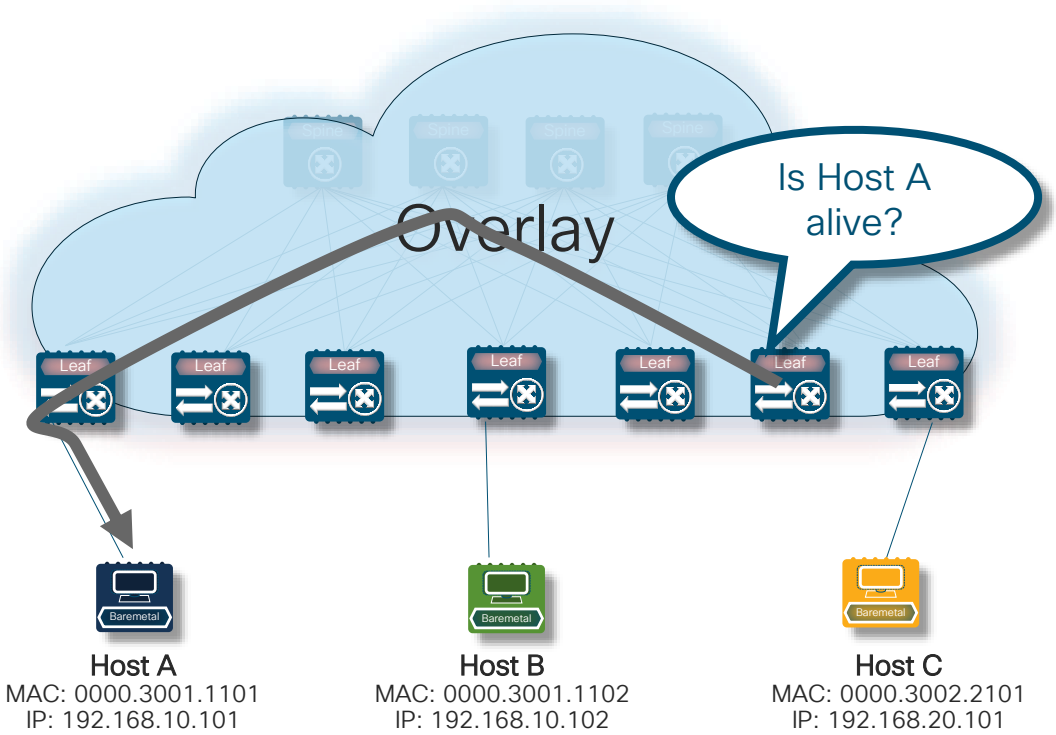
VXLAN OAM – Reachability Verification



```
L15# ping 10.50.1.18 source 10.50.1.15 vrf BlUE PING
 10.50.1.18 (10.50.1.18) from 10.50.1.15: 56 data bytes
64 bytes from 10.50.1.18: icmp_seq=0 ttl=254 time=1.041 ms
64 bytes from 10.50.1.18: icmp_seq=1 ttl=254 time=0.704 ms
64 bytes from 10.50.1.18: icmp_seq=2 ttl=254 time=1.303 ms
64 bytes from 10.50.1.18: icmp_seq=3 ttl=254 time=0.811 ms
64 bytes from 10.50.1.18: icmp_seq=4 ttl=254 time=0.745 ms
--- 10.50.1.18 ping statistics ---
5 packets transmitted, 5 packets received, 0.00% packet loss
round-trip min/avg/max = 0.704/0.92/1.303 ms
```

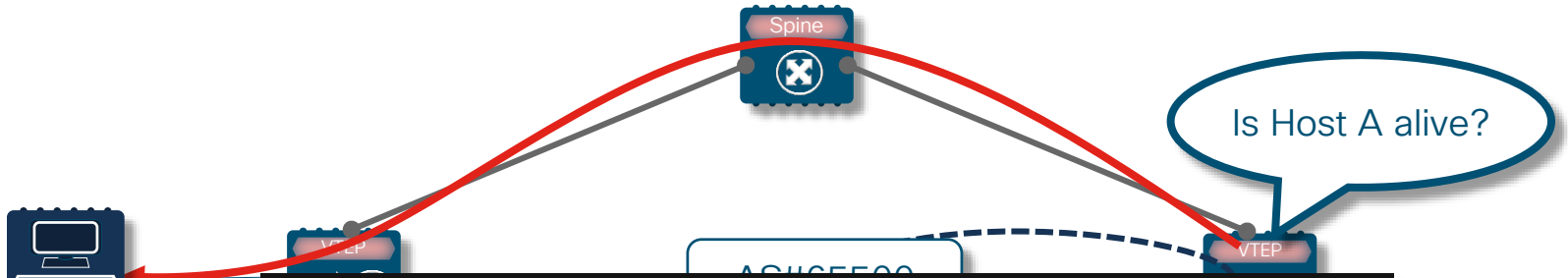
```
L15# show ip route 10.50.1.18 vrf BLUE
IP Route Table for VRF "BLUE”
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.50.1.18/32, ubest/mbest: 1/0
    *via 10.200.200.18%default, [200/0], 00:11:11, bgp-65501,
     internal, tag 65501 (evpn) segid: 50001 tunnelid: 0xac8c812 encap: VXLAN

          BGP-EVPN: VNI=50001 (EVPN)
        client-specific data: 4d
        recursive next hop: 10.200.200.18/32%default
        extended route information: BGP origin AS 65501 BGP peer AS 65501
```

Endpoint Reachability – VXLAN OAM



- Endpoint Reachability
  - Uses ICMP
  - VTEP to Endpoint reachability
  - VTEP to VTEP reachability
- Validates ECMP Path
  - Single Random Path
  - Multiple, Random/Specified Path
- Provides VXLAN Outer UDP Source Port (SPORT) as output

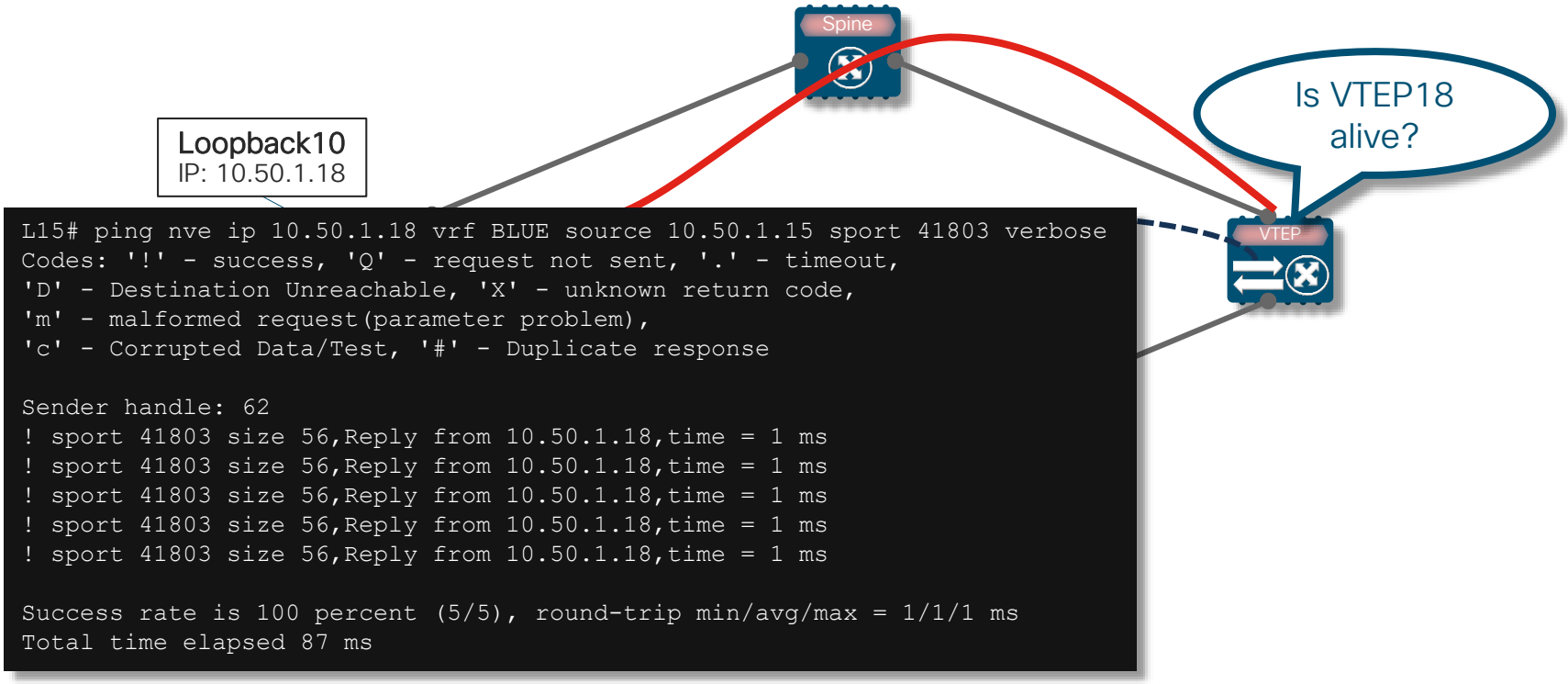
Endpoint Reachability – VXLAN OAM



Host A
MAC: 0000.3001.1101
IP: 192.168.10.101

```
L15# ping nve ip 192.168.10.101 vrf BLUE source 10.50.1.15 verbose
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Sender handle: 32
! sport 35977 size 56,Reply from 192.168.10.101,time = 1 ms
! sport 35977 size 56,Reply from 192.168.10.101,time = 2 ms
! sport 35977 size 56,Reply from 192.168.10.101,time = 1 ms
! sport 35977 size 56,Reply from 192.168.10.101,time = 1 ms
! sport 35977 size 56,Reply from 192.168.10.101,time = 1 ms

Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms
Total time elapsed 89 ms
```

VTEP Reachability – VXLAN OAM




```
L15# ping nve ip 10.50.1.18 vrf BLUE source 10.50.1.15 sport 41803 verbose
Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Sender handle: 62
! sport 41803 size 56,Reply from 10.50.1.18,time = 1 ms
! sport 41803 size 56,Reply from 10.50.1.18,time = 1 ms
! sport 41803 size 56,Reply from 10.50.1.18,time = 1 ms
! sport 41803 size 56,Reply from 10.50.1.18,time = 1 ms
! sport 41803 size 56,Reply from 10.50.1.18,time = 1 ms

Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms
Total time elapsed 87 ms
```

CLI Options – ICMP-based NVE Ping

```
ping nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback verbose
```

```
ping nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback sport Outer Source Port verbose
```

```
ping nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback egress Uplink Interface verbose
```

- Issues Ping to Host or Loopback IP address
- Specifies the VRF where Source and Destination Endpoint exists
- Choose the local Loopback IP as a Source IP address for the NVE Ping

- Use a specific VXLAN Outer Source Port
  - Otherwise Random Generated VXLAN Source Ports are used
- Use specific egress Interface
  - i.e. Uplink towards Spine
  - Otherwise ECMP hashing is used with Random or defined VXLAN Source Port

CLI Options – NVE Ping with MAC

```
ping nve mac Destination Host MAC Local-VLAN profile Profile # verbose
```

```
ping nve mac Destination Host MAC Local-VLAN profile Profile # sport Outer Source Port verbose
```

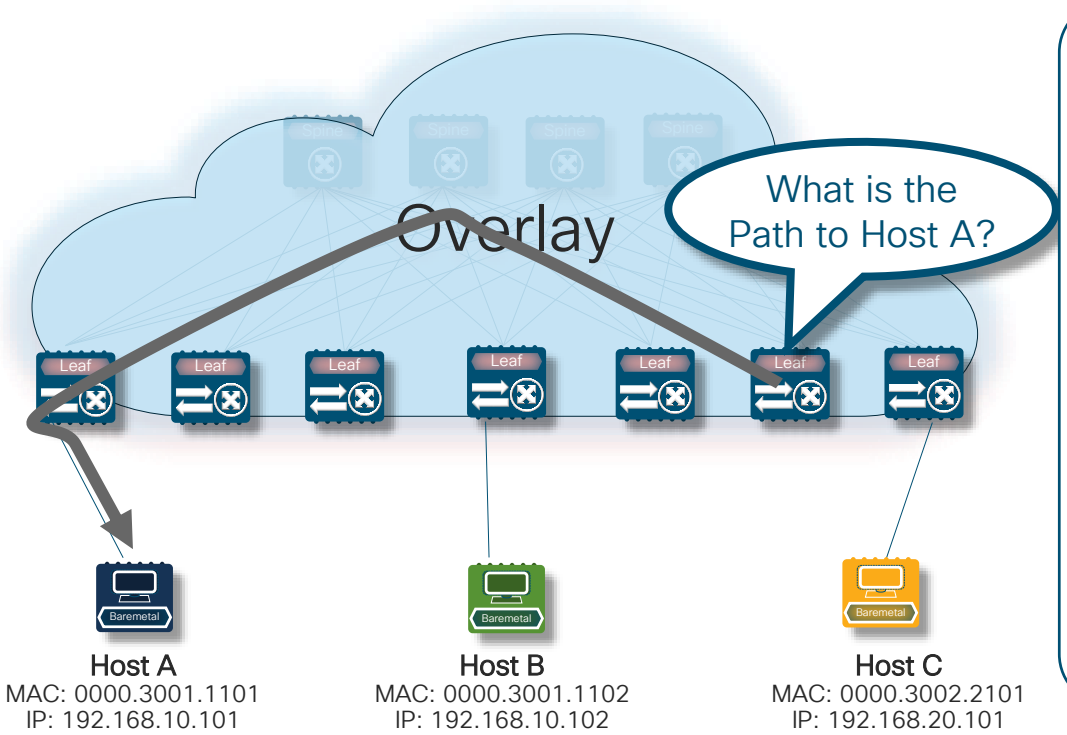
```
ping nve mac Destination Host MAC Local-VLAN profile Profile # egress Uplink Interface verbose
```

- Issues Ping to Destination MAC
- Input requires VLAN mapped with L2VNI where Destination MAC resides
- Uses nv03 tissa draft

```
ngoam profile 4
  oam-channel 2
```

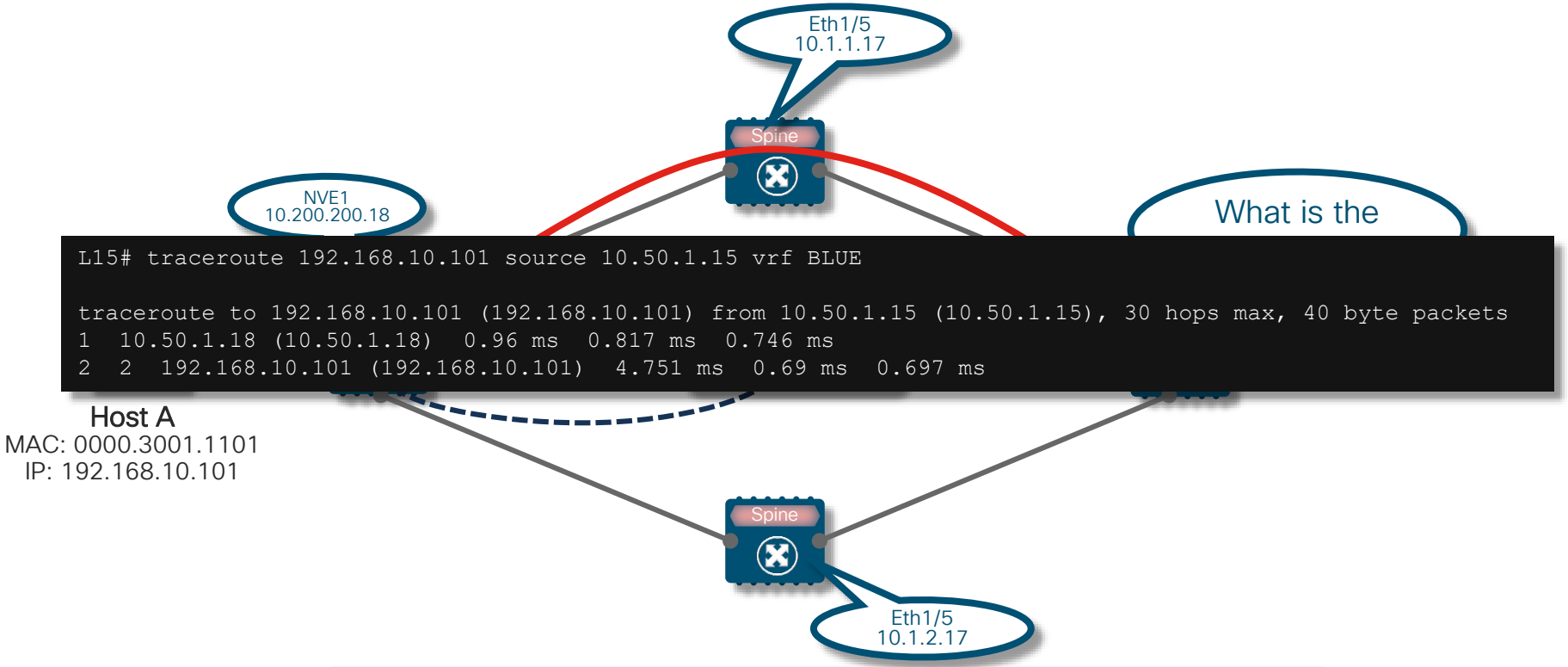
- OAM response returned from destination leaf
- Use specific egress Interface
  - Otherwise ECMP hashing is used with Random or defined VXLAN Source Port

Endpoint Traceroute – VXLAN OAM



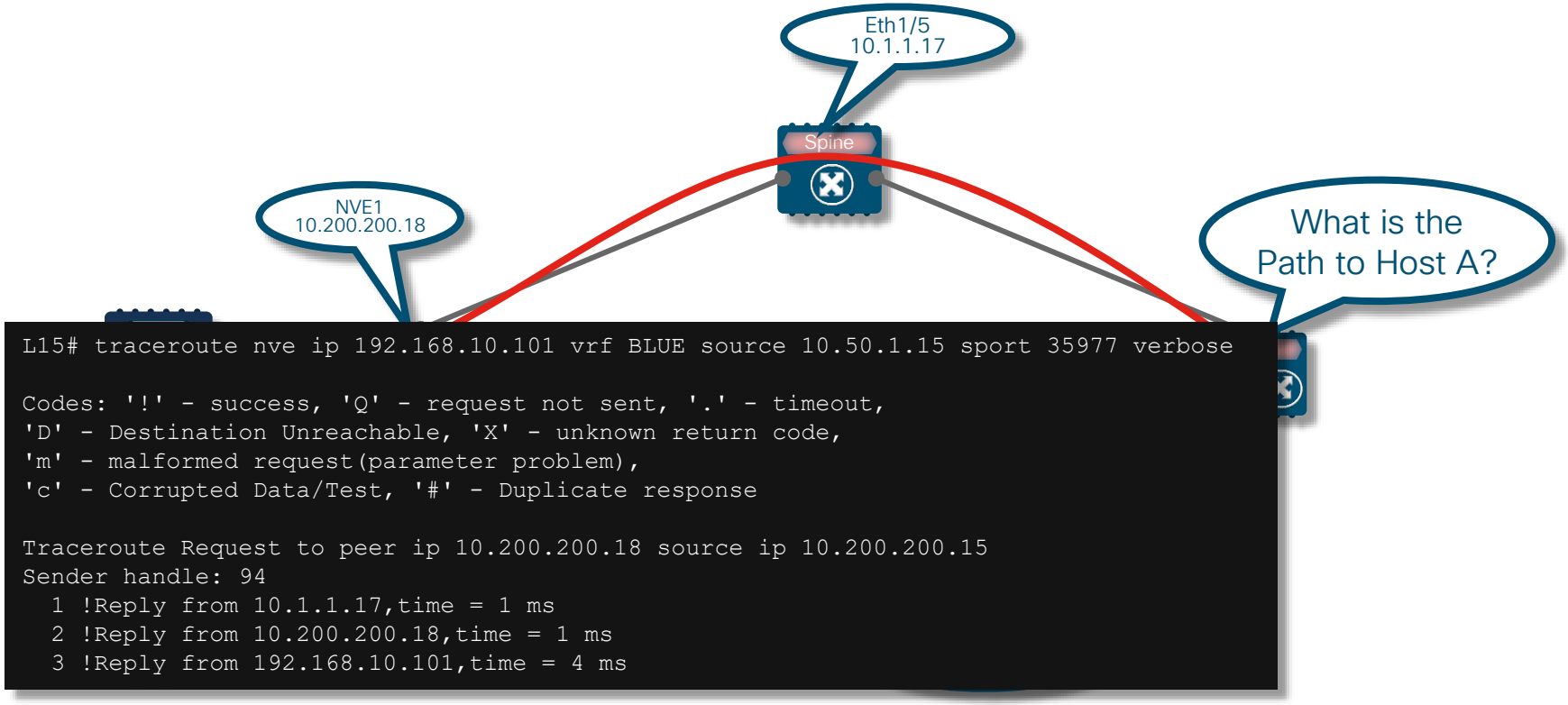
- Endpoint Traceroute
  - Uses ICMP
  - VTEP to Endpoint
  - VTEP to VTEP
- Validates Overlay Path
  - Single Specified Path
  - Multiple, Specified Path
- Provides Overlay to Underlay correlation

How would a normal Traceroute look like?



Which Path did my Traceroute take?

Endpoint Traceroute – VXLAN OAM




```
L15# traceroute nve ip 192.168.10.101 vrf BLUE source 10.50.1.15 sport 35977 verbose

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Traceroute Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 94
  1 !Reply from 10.1.1.17,time = 1 ms
  2 !Reply from 10.200.200.18,time = 1 ms
  3 !Reply from 192.168.10.101,time = 4 ms
```

Endpoint Traceroute – VXLAN OAM – Close-Up

```
L15# traceroute nve ip 192.168.10.101 vrf BLUE source 10.50.1.15 sport 35977 verbose

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Traceroute Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 94
  1 !Reply from 10.1.1.17,time = 1 ms
  2 !Reply from 10.200.200.18,time = 1 ms
  3 !Reply from 192.168.10.101,time = 4 ms
```

Spine Ingress Interface IP

Destination VTEP IP

Host A IP

Spine Ingress Interface and Destination VTEP IP Address are Underlay Information – additions vs. standard Traceroute

CLI Options – ICMP-based NVE Traceroute

```
traceroute nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback verbose
```

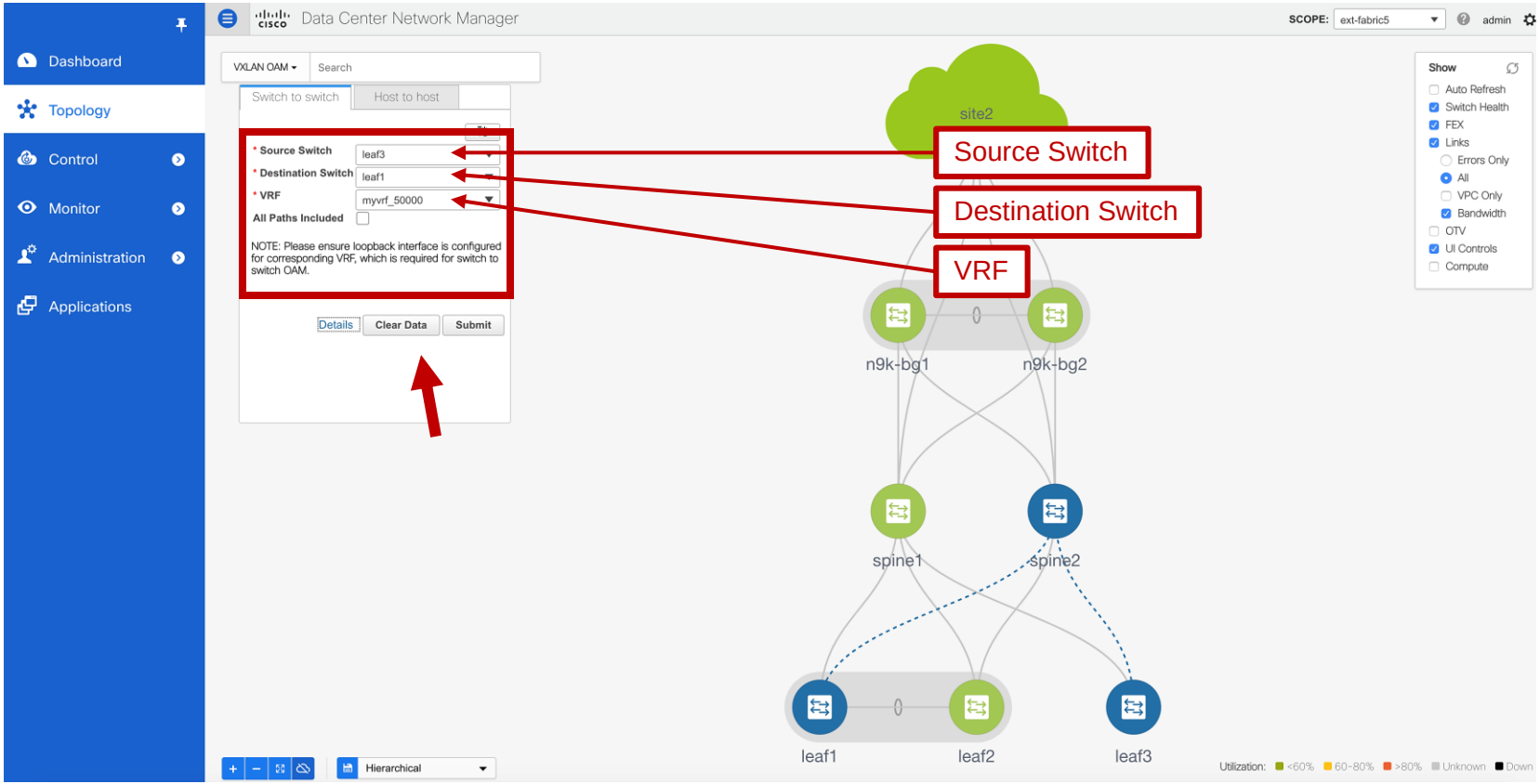
```
traceroute nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback sport Outer Source Port verbose
```

```
traceroute nve ip Destination Host/Loopback vrf VRF-Name source Source Loopback egress Uplink Interface verbose
```

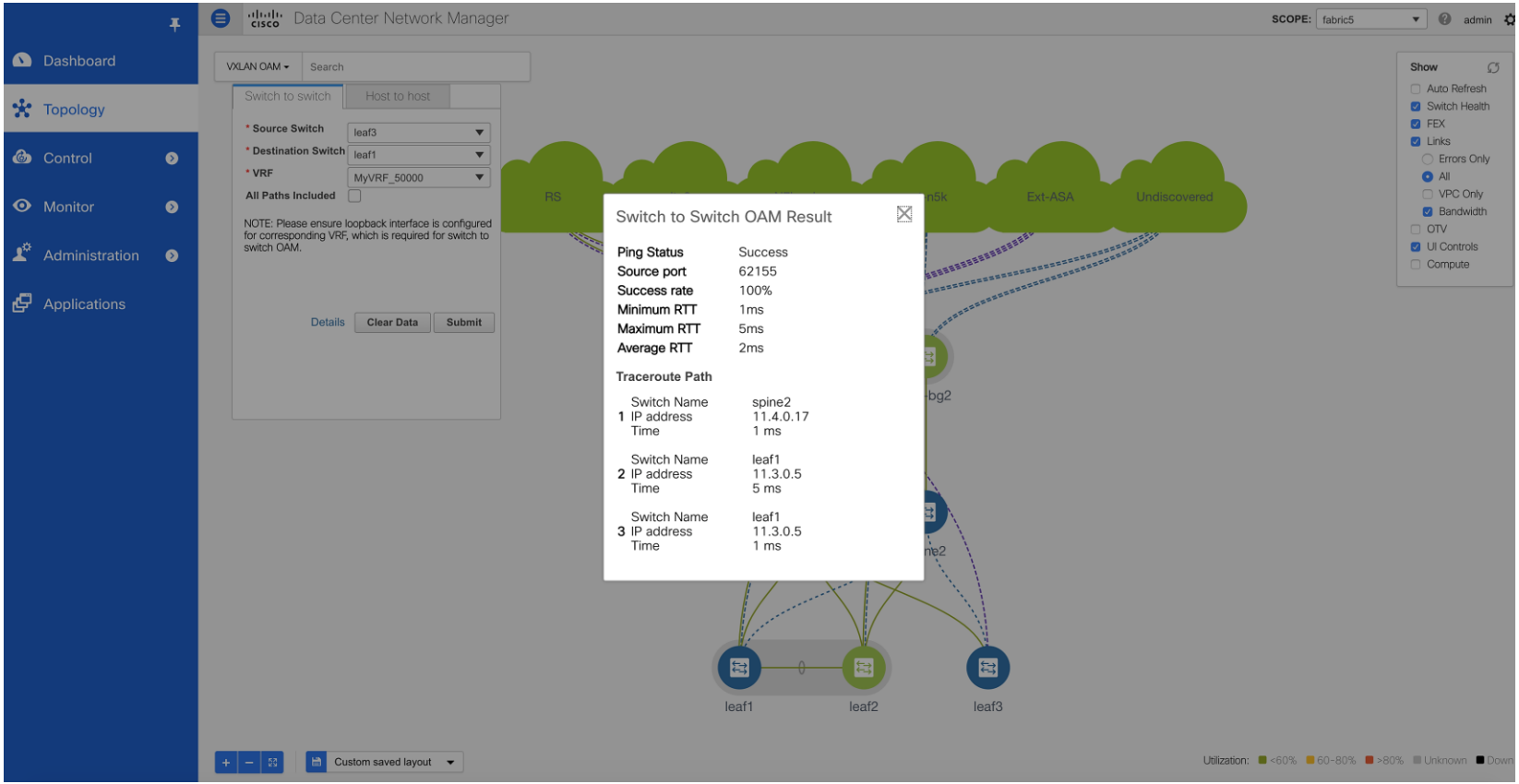
- Issues Traceroute to Host or Loopback IP address
- Specifies the VRF where Source and Destination Endpoint exists
- Chose the local Loopback IP as a Source IP address for the NVE Ping

- Use a specific VXLAN Outer Source Port
  - Otherwise Random Generated VXLAN Source Ports are used
- Use specific egress Interface
  - i.e. Uplink towards Spine
  - Otherwise ECMP hashing is used with Random or defined VXLAN Source Port

Ping/Traceroute with DCNM

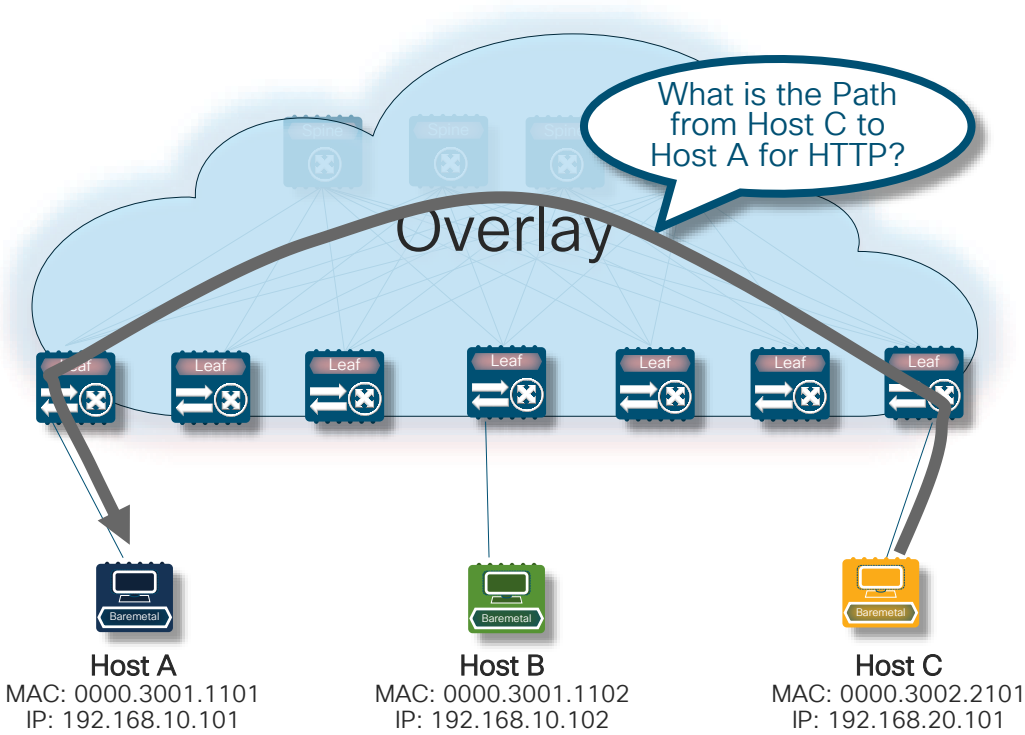


Ping/Traceroute with DCNM



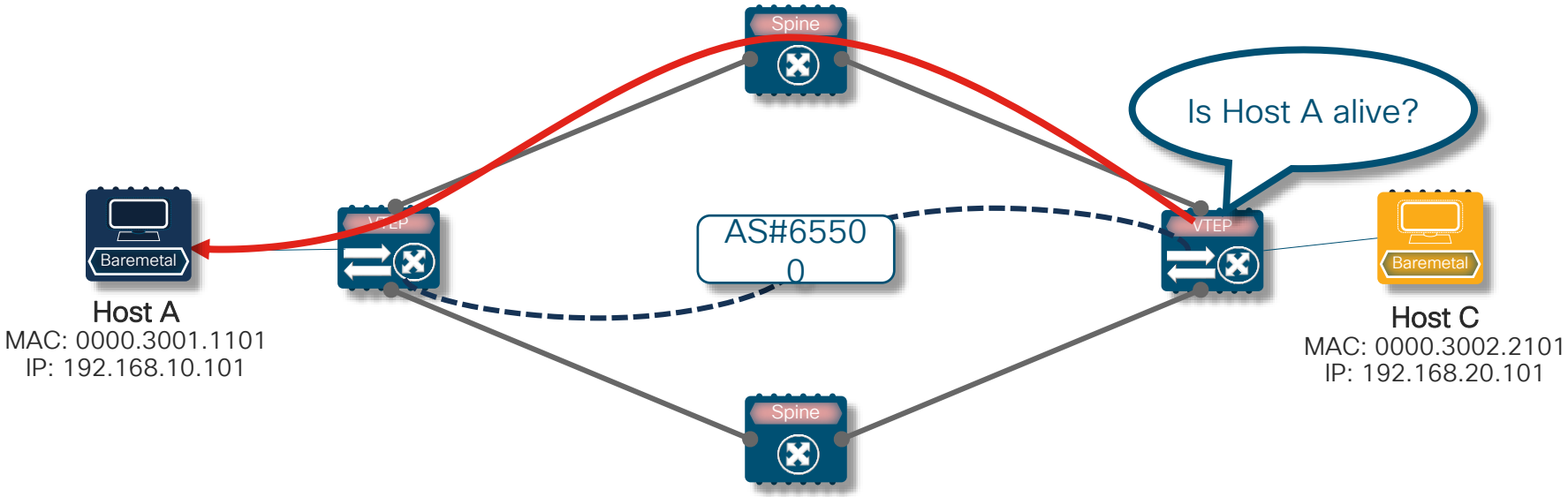
Demo – VXLAN OAM NVE Ping and Traceroute

Pathtrace for Enhanced Network Visibility

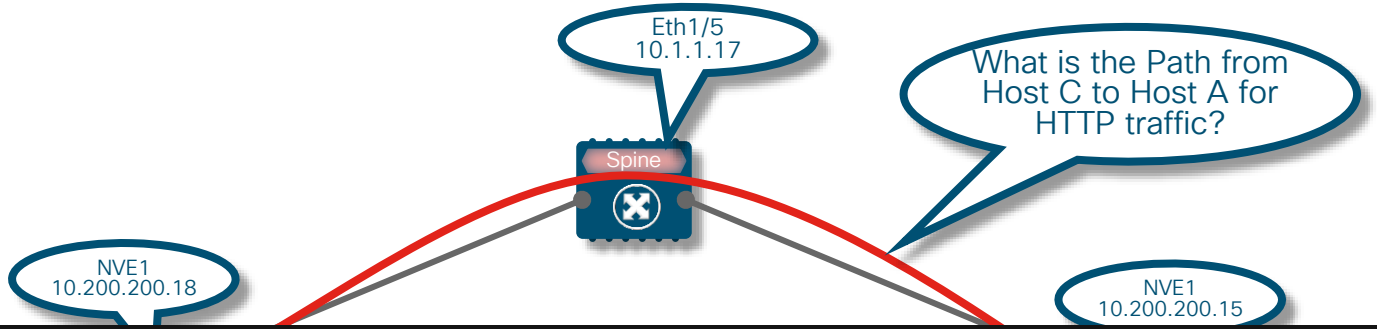


- Application Specific Pathtrace
  - Uses “draft-tissa-nvo3-oam-fm”
  - Endpoint to Endpoint Pathtrace
  - Adds Interface Load and Error Statistics of the Path
  - Uses Protocol Information
- Validates Specific or All Paths
- Provides Overlay to Underlay correlation
- Superset of NVE Ping/Traceroute

Endpoint Reachability – VXLAN OAM

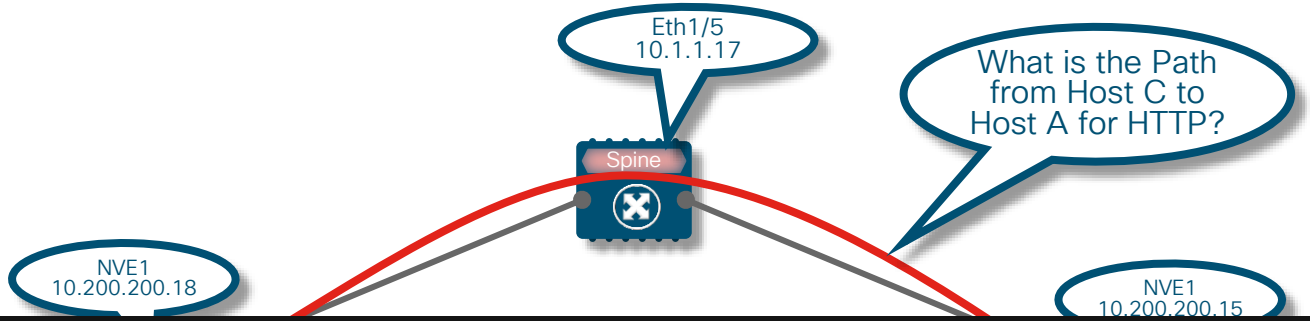


Pathtrace with Known VTEP




```
L15# pathtrace nve ip 10.200.200.18 vrf BLUE payload ip 192.168.10.101 192.168.20.101 port 54321 80 proto 6 payloac

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Path trace Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 142

Hop   Code   ReplyIP   IngressI/f  EgressI/f   State
=====================================================
  1 !Reply from 10.1.1.17, Eth1/5  Eth1/8  UP / UP
  2 !Reply from 10.200.200.18, Eth1/54  Unknown  UP / DOWN
```

Pathtrace with Unknown VTEP




```
L15# pathtrace nve ip unknown vrf BLUE payload ip 192.168.10.101 192.168.20.101 port 54321 80 proto 6 payload-end

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Path trace Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 142

Hop   Code   ReplyIP   IngressI/f  EgressI/f   State
=================================================
  1 !Reply from 10.1.1.17, Eth1/5  Eth1/8  UP / UP
  2 !Reply from 10.200.200.18, Eth1/54  Vlan10  UP / UP
```

Pathtrace – VXLAN OAM – Close-Up

```
L15# pathtrace nve ip unknown vrf BLUE
     payload
     ip 192.168.10.101 192.168.20.101
     port 54321 80
     proto 6
     payload-end

Codes: '!' - success, 'Q' - request not sent, '.' - timeout,
'D' - Destination Unreachable, 'X' - unknown return code,
'm' - malformed request(parameter problem),
'c' - Corrupted Data/Test, '#' - Duplicate response

Path trace Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 142

Hop   Code   ReplyIP   IngressI/f  EgressI/f   State
=================================================
  1 !Reply from 10.1.1.17, Eth1/5  Eth1/8  UP / UP
  2 !Reply from 10.200.200.18, Eth1/54  Vlan10  UP / UP
```

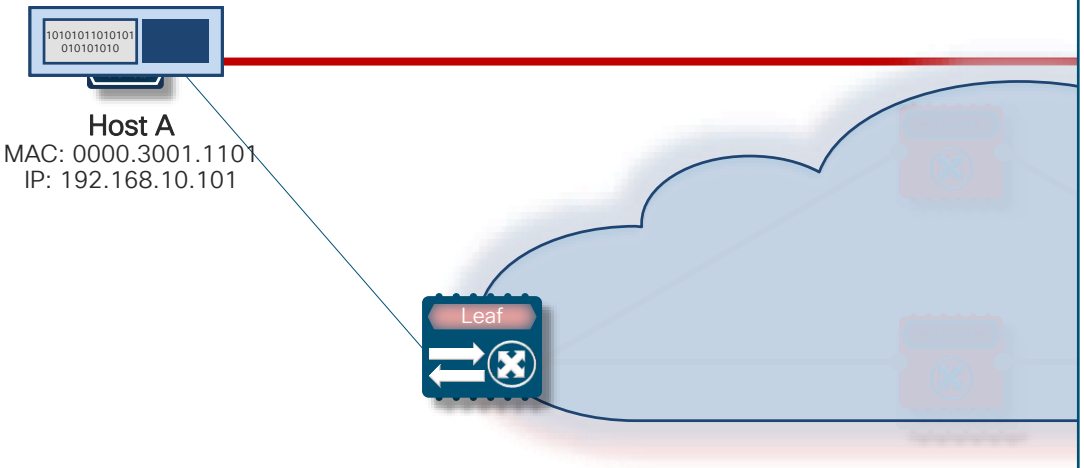
Known or Unknown VTEP IP Address

Destination Endpoint IP / Source Endpoint IP

Source Port / Destination Port

TCP (IANA Protocol Number 6)

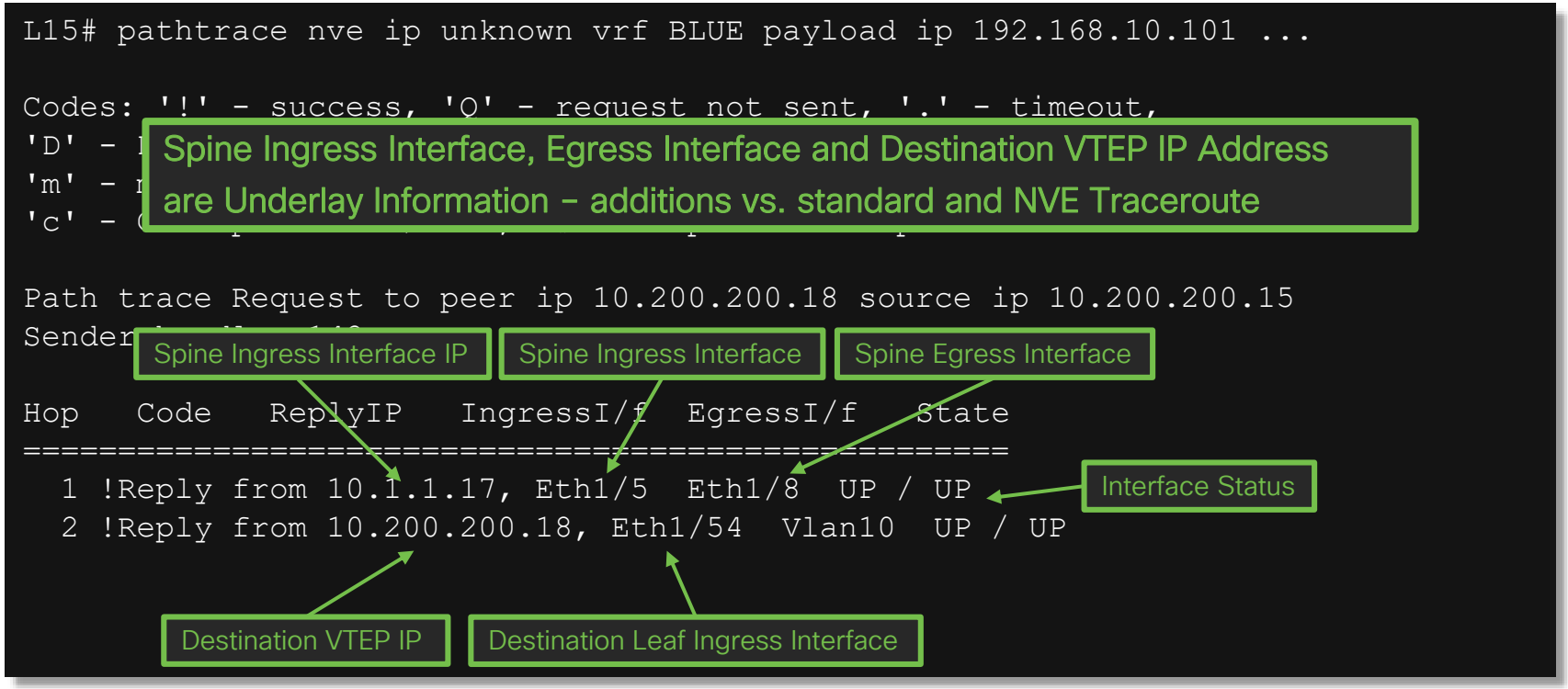
Why are we specifying Payload information?



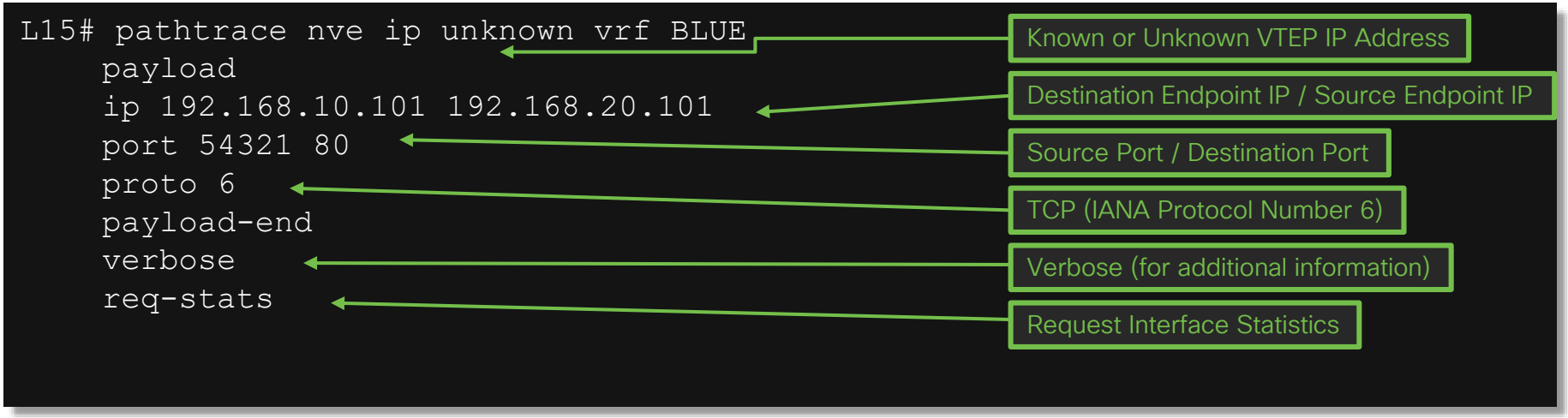
- VXLAN provides variable UDP Source Port in Outer Header
- Hash of the inner Layer-2/Layer-3/Layer-4 Headers of the original Ethernet Frame.
- Enables entropy for ECMP Load balancing in the Network

Which Path did your Application Traffic take?

Pathtrace – VXLAN OAM – Close-Up

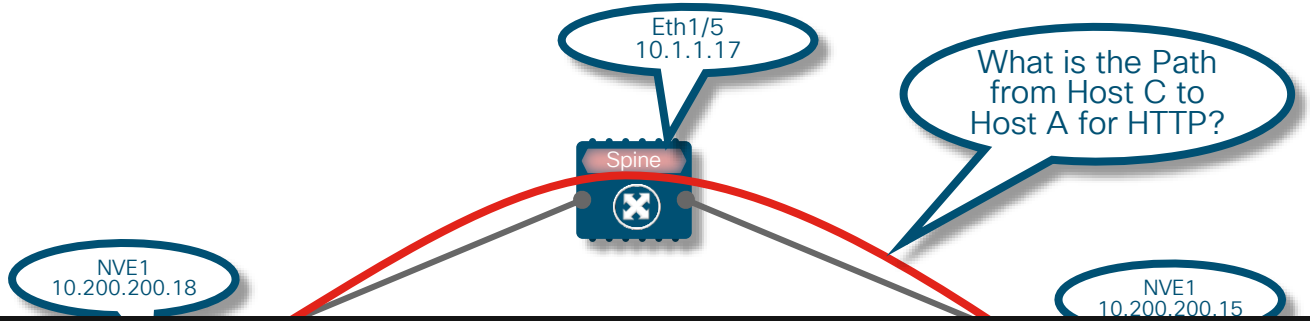


Pathtrace – VXLAN OAM – Extensions (Routing)



Pathtrace – VXLAN OAM – Extensions (Bridging)

Pathtrace with Unknown VTEP – Request Statistics




```
L15# pathtrace nve ip unknown vrf BLUE payload ip 192.168.10.101 192.168.20.101 payload-end verbose req-stats

Path trace Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 168

Hop   Code   ReplyIP   IngressI/f  EgressI/f   State
=====================================================
1 !Reply from 10.1.1.17, Eth1/5  Eth1/8  UP / UP
Input Stats: PktRate:0 ByteRate:0 Load:0 Bytes:66113123 unicast:140952 mcast:252611 bcast:2 discards:0 errors:0 unkno
Output Stats: PktRate:0 ByteRate:0 load:0 bytes:51359028 unicast:100504 mcast:252545 bcast:6 discards:0 errors:0 banc

2 !Reply from 10.200.200.18, Eth1/54  Vlan10  UP / UP
Input Stats: PktRate:0 ByteRate:0 Load:0 Bytes:52302926 unicast:99998 mcast:263225 bcast:4 discards:0 errors:0 unknow
```

Request Stats – VXLAN OAM – Close-Up

```
L15# pathtrace nve ip unknown vrf BLUE
     payload ip 192.168.10.101 192.168.20.101 payload-end
     verbose req-stats

Path trace Request to peer ip 10.200.200.18 source ip 10.200.200.15
Sender handle: 168

Hop   Code   ReplyIP   IngressI/f  EgressI/f   State
================================================
1 !Reply from 10.1.1.17, Eth1/5  Eth1/8  UP / UP
Input Stats: PktRate:0 ByteRate:0 Load:0 Bytes:66113123 unicast:140952 mcast:252611 bcast:
Output Stats: PktRate:0 ByteRate:0 load:0 bytes:51359028 unicast:100504 mcast:252545 bcast

2 !Reply from 10.200.200.18, Eth1/54  Vlan10  UP / UP
Input Stats: PktRate:0 ByteRate:0 Load:0 Bytes:52302926 unicast:99998 mcast:263225 bcast:4
```

OAM Session ID

Interface Statistics

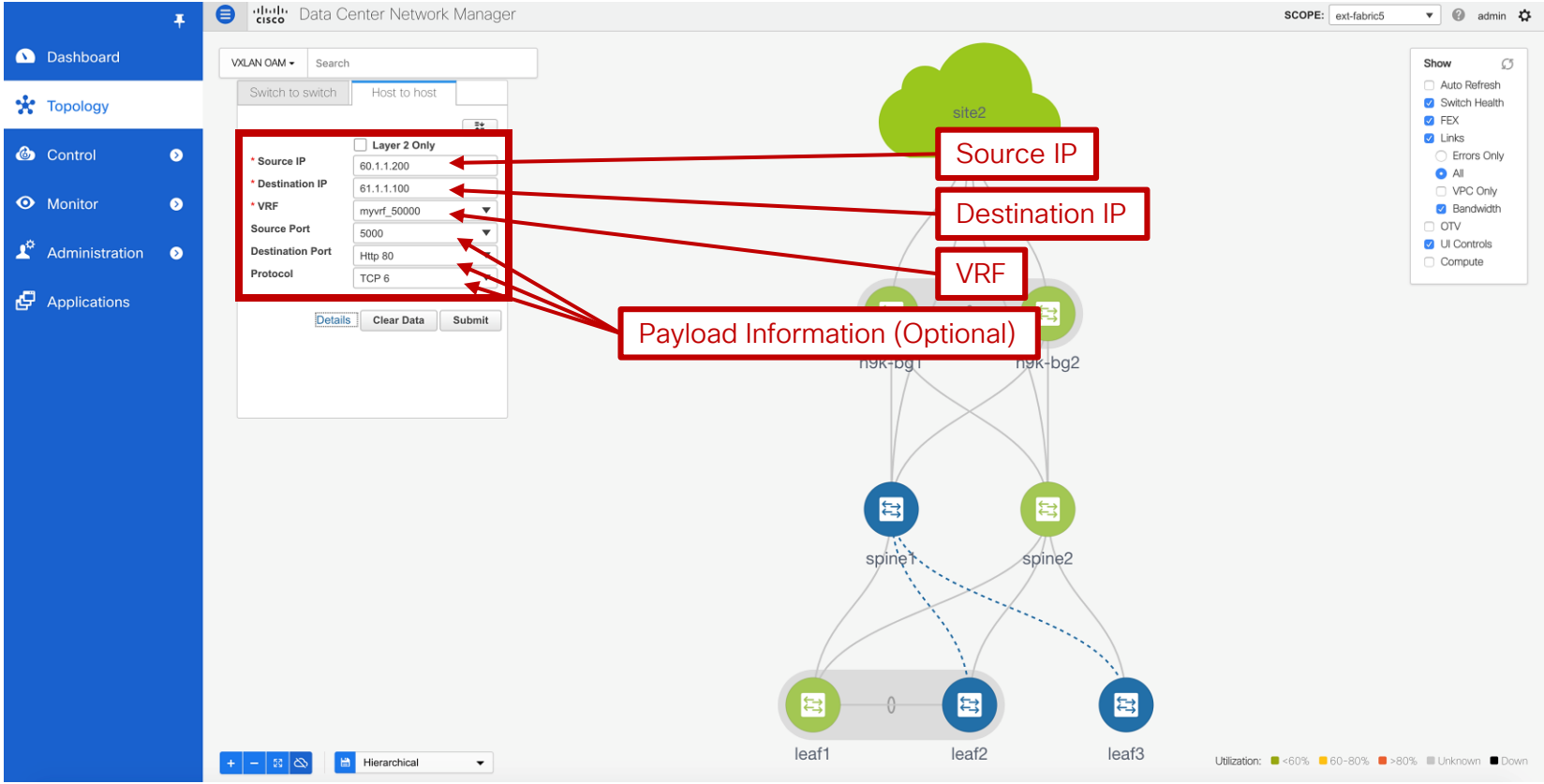
Database Output – VXLAN OAM – Close-Up

```
L15# show ngoam pathtrace database session 168 detail

Pathtrace entry for session id 168
==================================

Start time: Tue Jun 13 01:18:39.710 PDT
End time: Tue Jun 13 01:18:39.735 PDT

Last Clear of Summary Statistics: Never
Pathtrace Requests: sent (2)/received (0)/timeout (0)/unsent (0)
Pathtrace Replies: sent (0)/received (2)/unsent (0)/Duplicate (0)

! Reply from 10.1.1.17 on Eth1/5, state  UP. Sent on Eth1/8, state  UP.
 Interface stats for interface: Eth1/5
------------------------------
Rx Len                : 84
Rx Bytes              : 66113123
Rx Pkt rate           : 0
Rx Byte rate          : 0
Rx Load               : 0
```

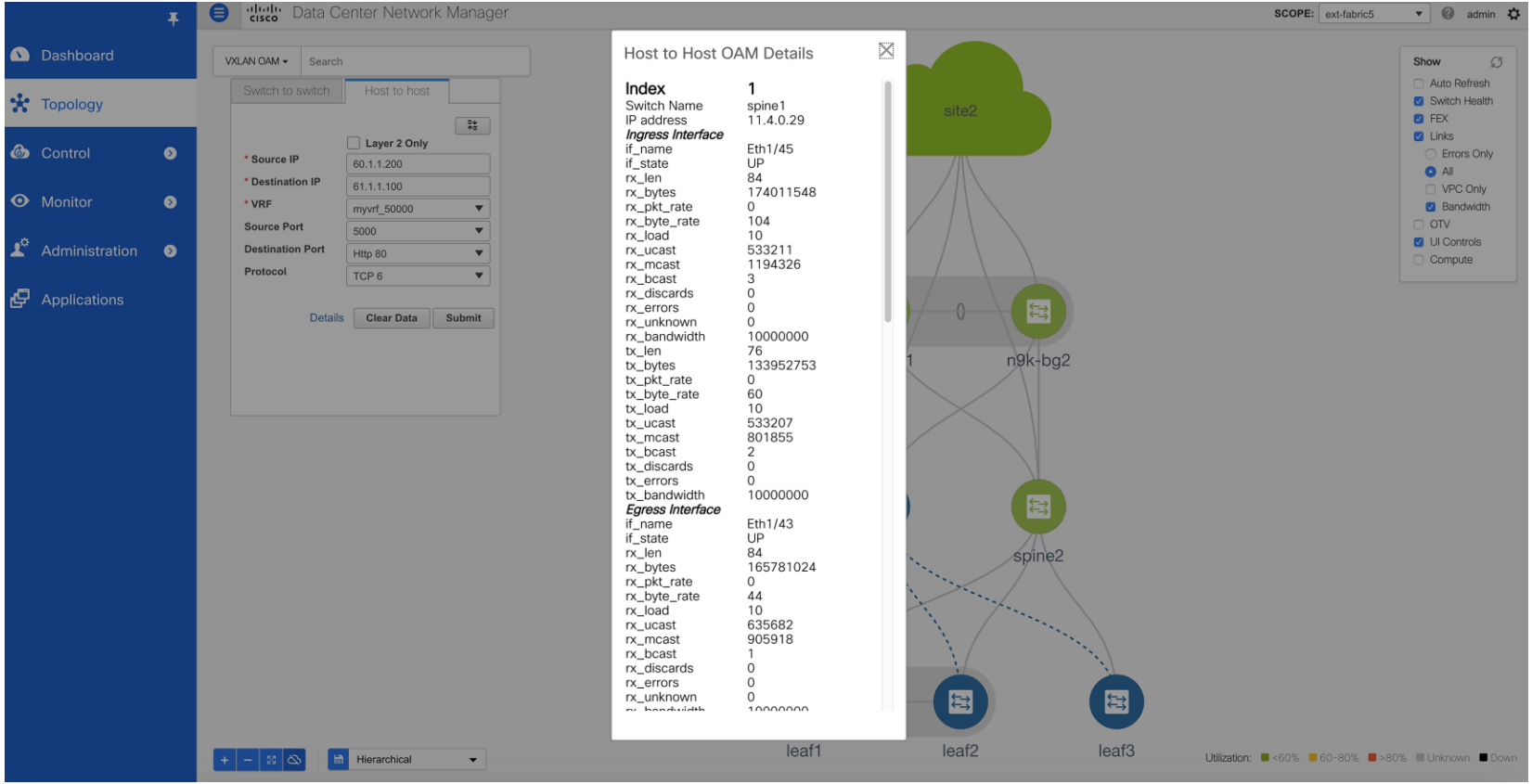
OAM Session ID

Interface Statistics

Endpoint Pathtrace with DCNM



Endpoint Pathtrace with DCNM

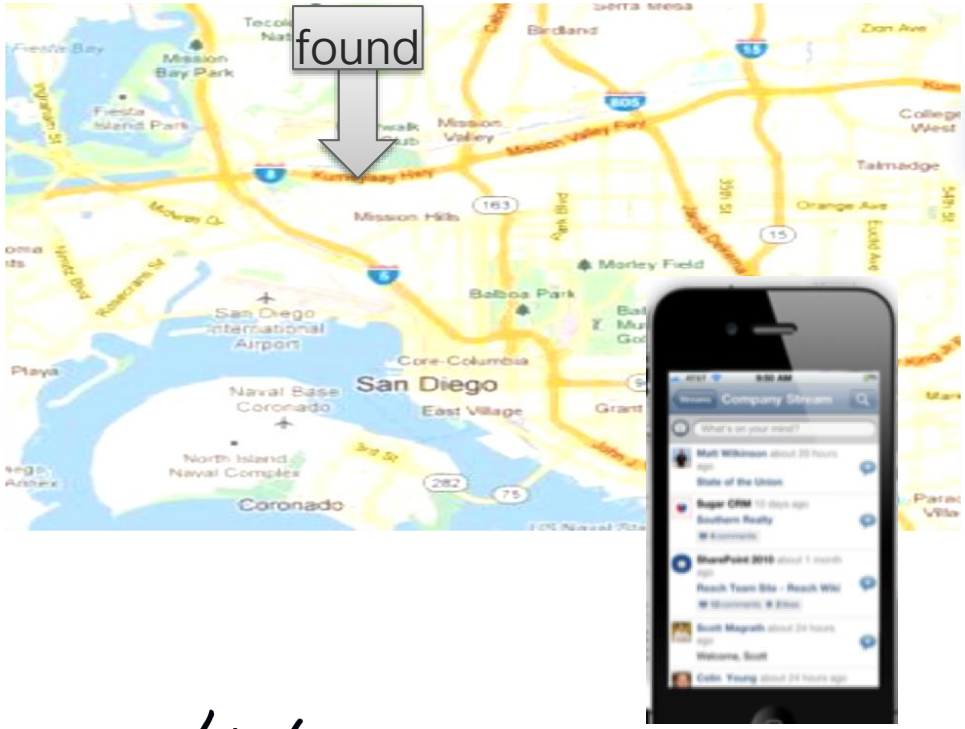


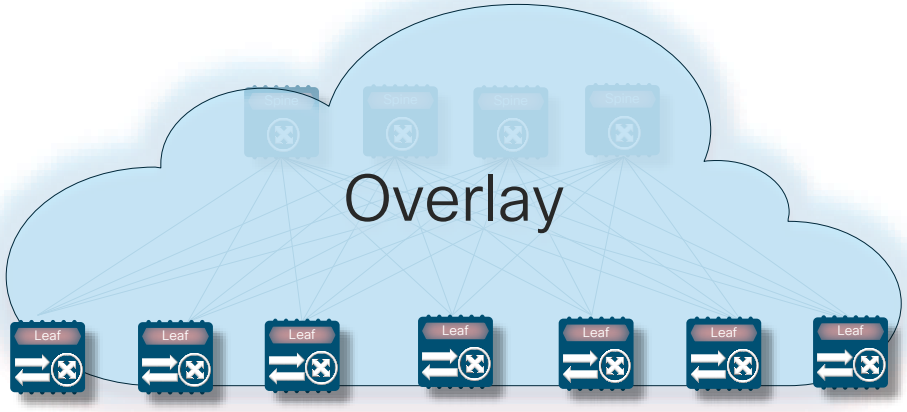
Demo – VXLAN OAM NVE PathTrace

Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

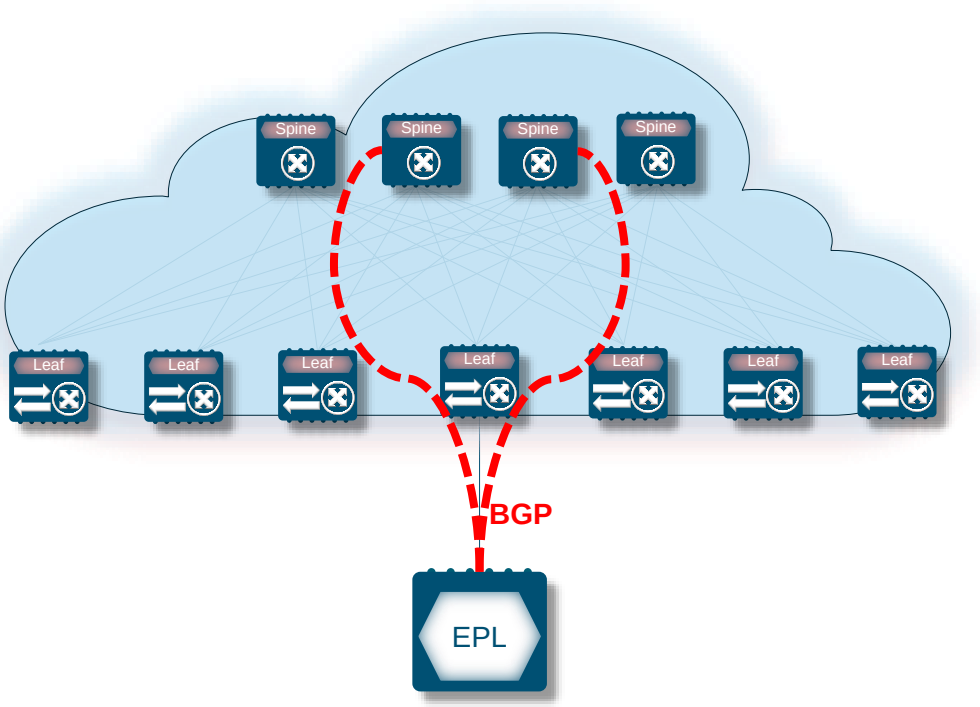
Where is my Endpoint?





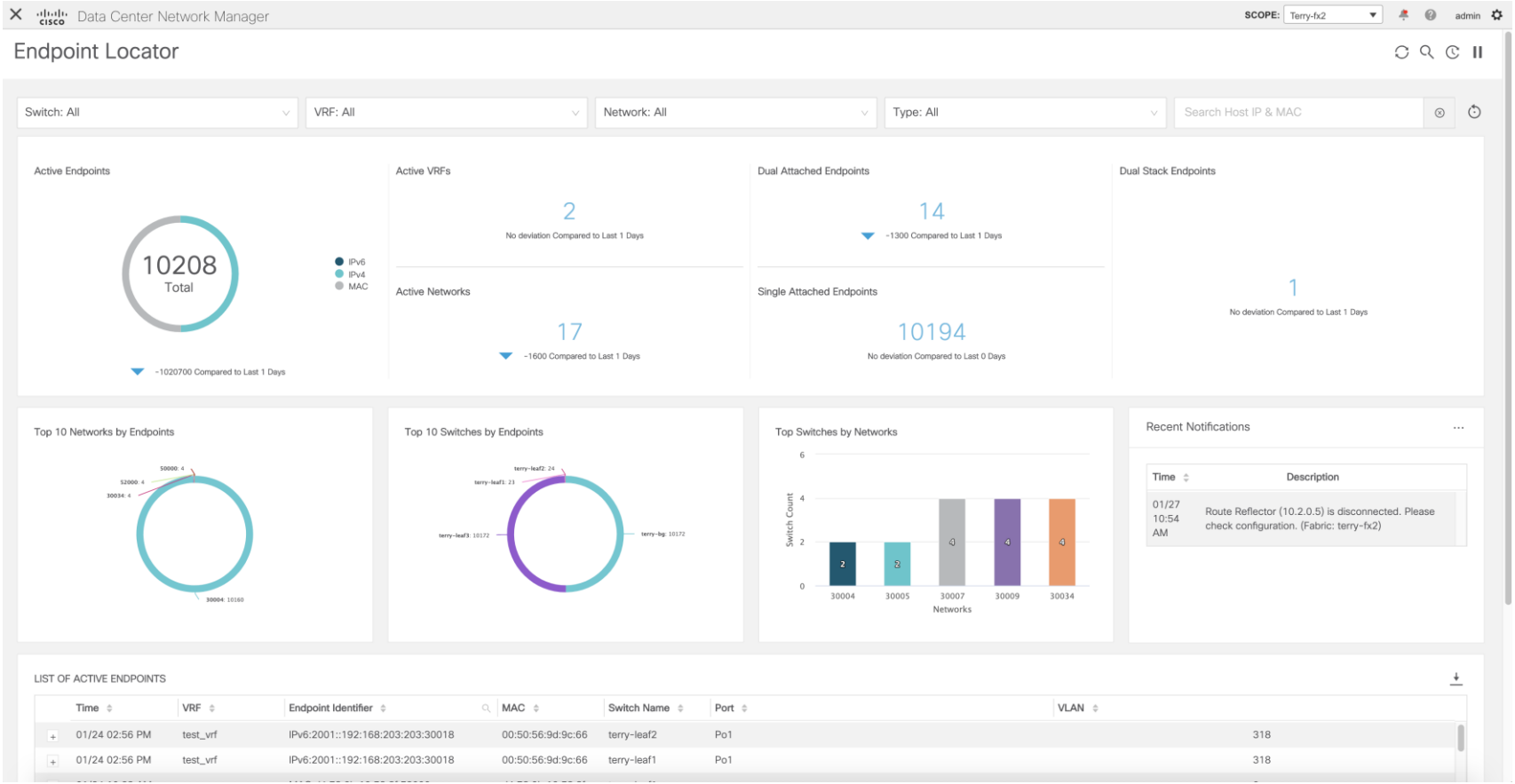
- Find Your Endpoints
- Correlate All Network Related Information for a given Endpoint
- Keep Endpoints Location History

Endpoint Locator (EPL) – Architecture



- Endpoint Locator (EPL)
  - Application in DCNM
  - Peers with the Overlay Control-Plane (i.e. BGP EVPN)
  - BGP Receiver only (Passive)
- Searchable and Scalable Database for Real-Time and Historic Data
- Stores every Endpoint Control-Plane Event
- Correlates with Inventory Data

Endpoint Locator

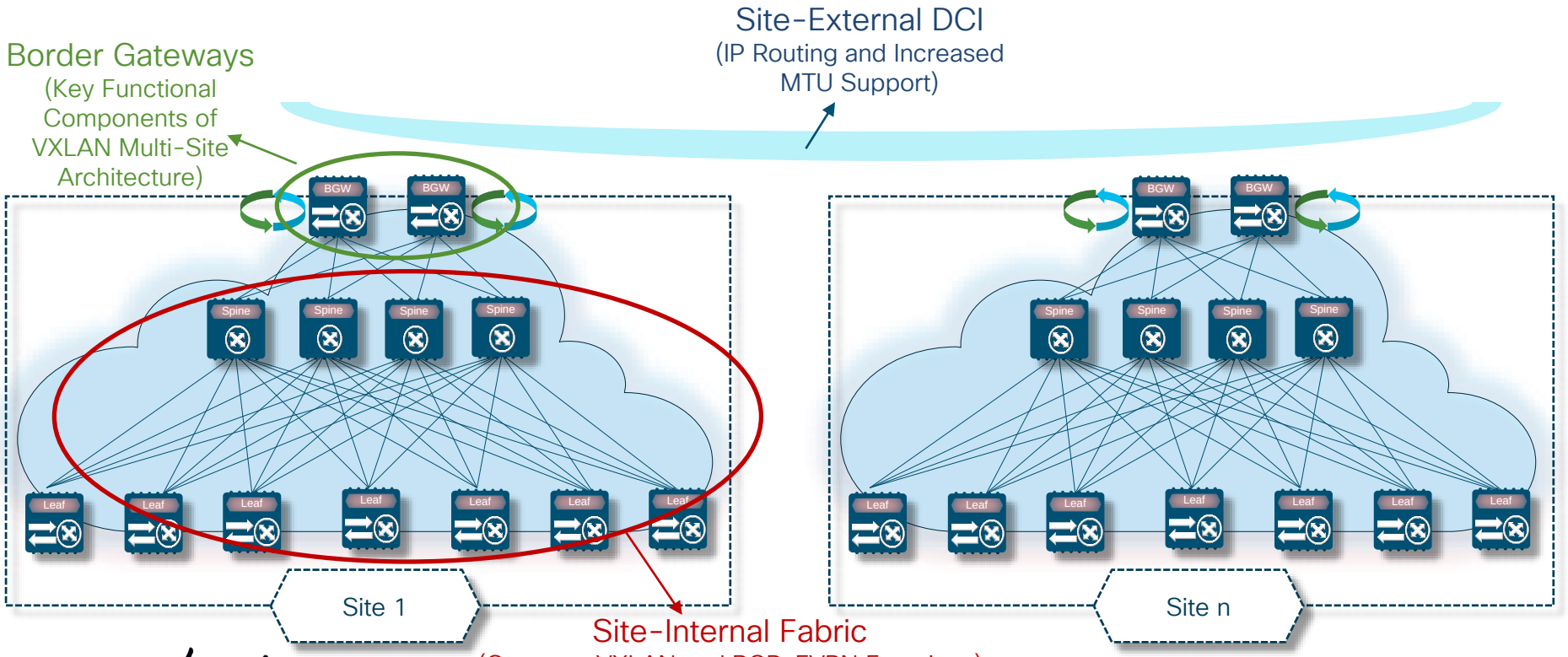


Demo –Endpoint Locator with DCNM

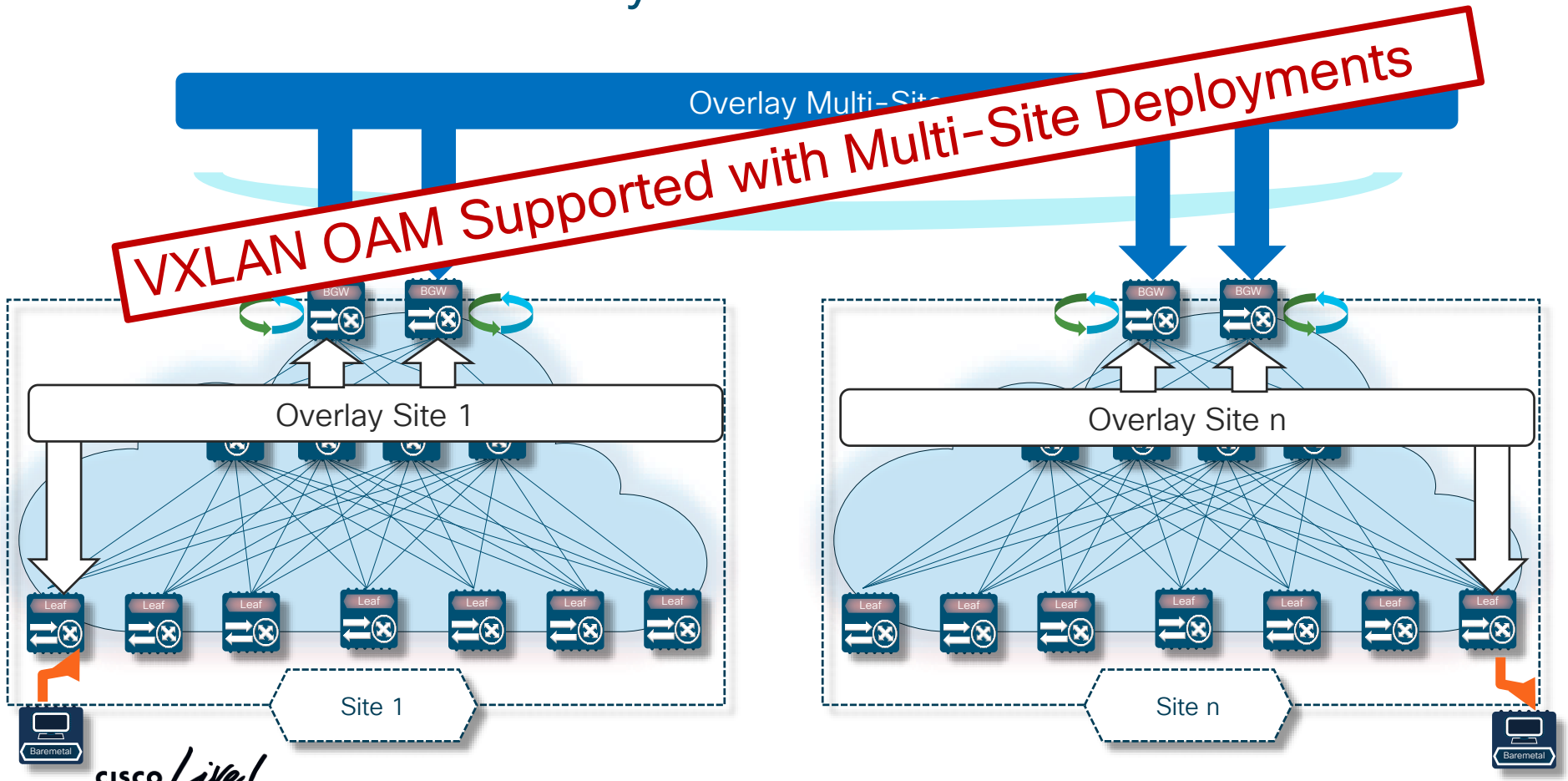
Agenda

- Overlays and Network abstraction
  - Underlay-Overlay Correlation
  - Motivation for Overlay OAM
- Operations, Administration and Management (OAM)
  - VXLAN OAM – NVE Ping, Traceroute, Pathtrace
  - Endpoint Visibility
  - EVPN Multi-Site
- Examples

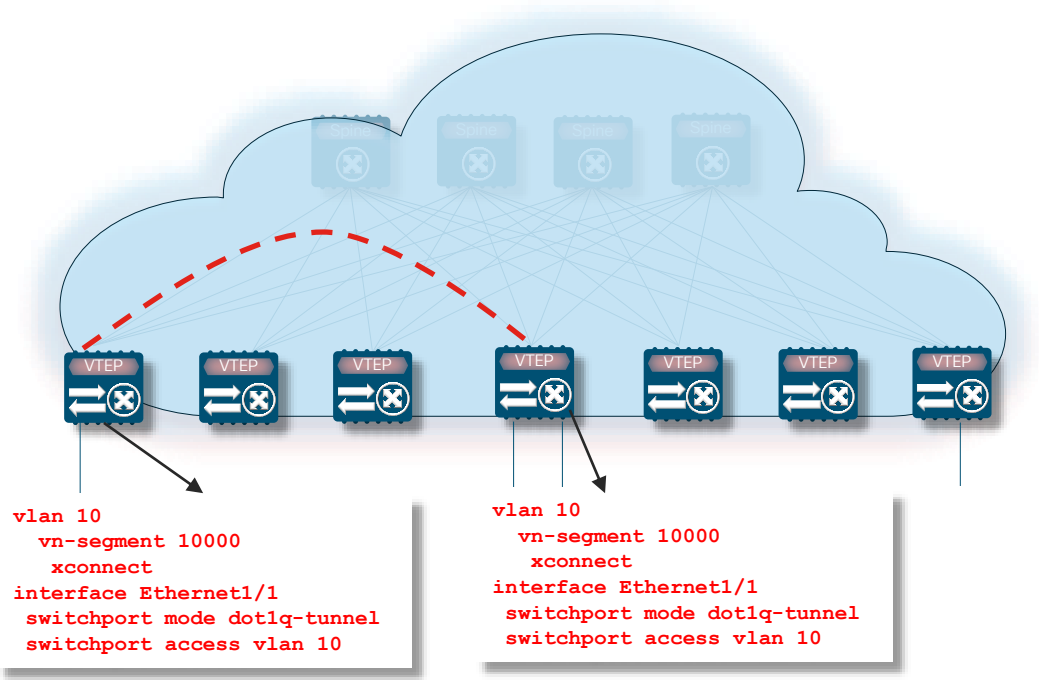
EVPN Multi-Site



Hierarchical Overlay Domains

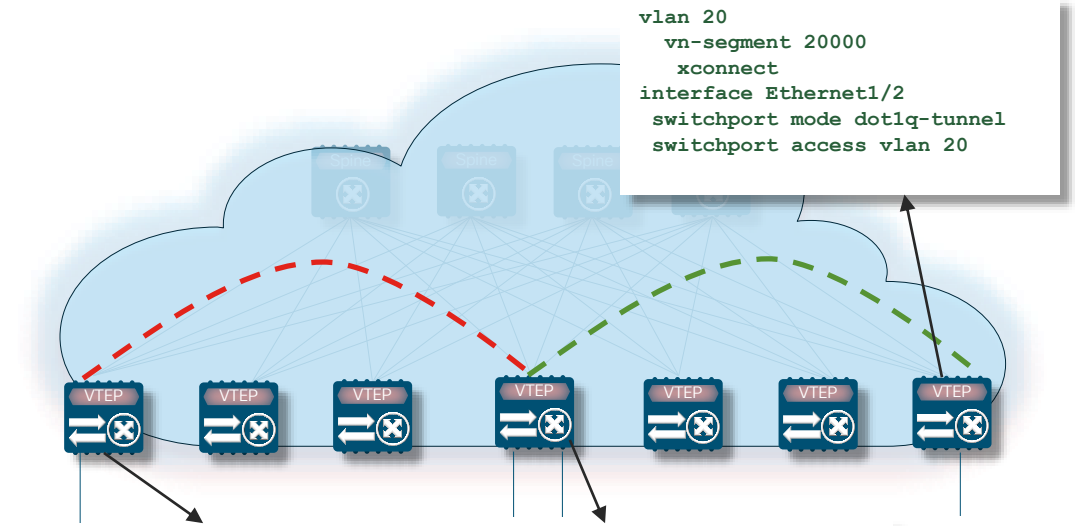


VXLAN XConnect



- MPLS Pseudowire like Tunneling with VXLAN
- Tunnel all control & data packets between VTEPs
- Attachment point is part of a unique provider VNI
- P2P

VXLAN XConnect




```
vlan 20
  vn-segment 20000
   xconnect
interface Ethernet1/2
 switchport mode dot1q-tunnel
 switchport access vlan 20
```

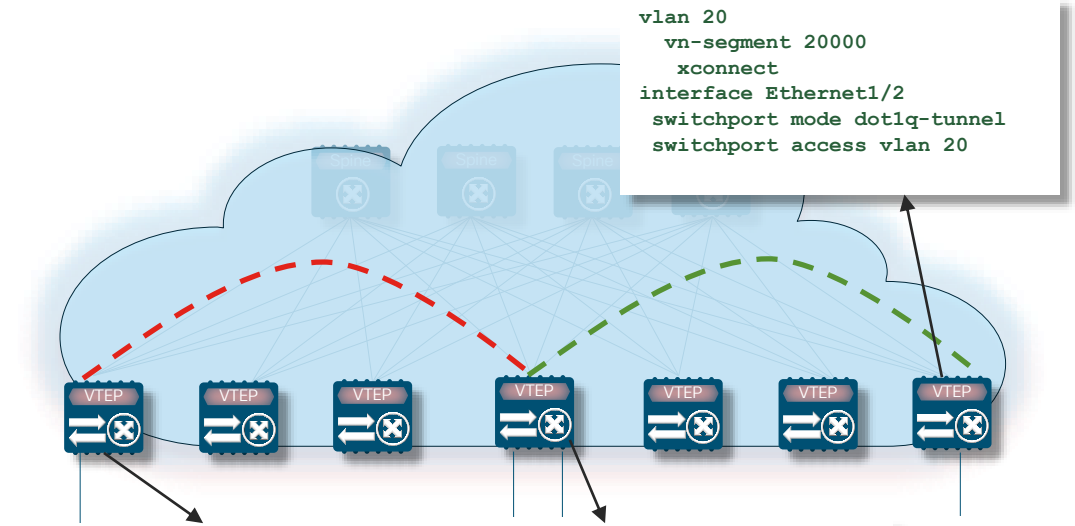
```
vlan 10
  vn-segment 10000
   xconnect
interface Ethernet1/1
 switchport mode dot1q-tunnel
 switchport access vlan 10
```

```
vlan 10
  vn-segment 10000
   xconnect
vlan 20
  vn-segment 20000
   xconnect

interface Ethernet1/1
 switchport mode dot1q-tunnel
 switchport access vlan 10
interface Ethernet1/2
 switchport mode dot1q-tunnel
 switchport access vlan 20
```

- MPLS Pseudowire like Tunneling with VXLAN
- Tunnel all control & data packets between VTEPs
- Attachment point is part of a unique provider VNI
- P2P

cisco Live!

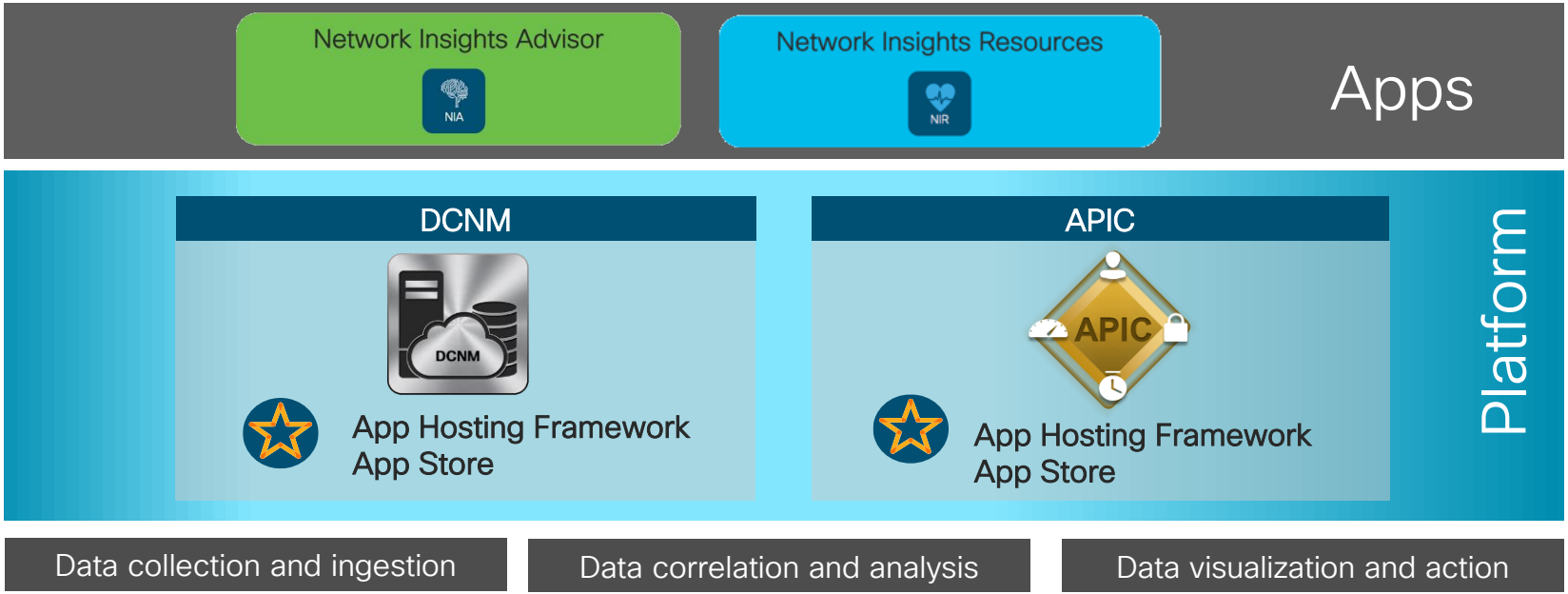
OAM for VXLAN XConnect



- Pro-active OAM
  - Monitor reachability to the remote VTEPs/VNI
  - Configurable heart-beat interval
- Failure Propagation
  - Convey local vni/vlan/interface failures to the remote VTEPs
  - Error-disabled mode

Network Insights

Network Insights Applications



Visibility
Learn from your network and recognize anomalies

Insights
See problems before your end users do

Proactive Troubleshooting
Find root cause faster with granular details

Data Center Visibility Use Cases

Network Health

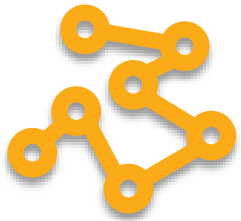
- CPU and memory utilization
- Forwarding table utilization
- Protocol state and events
- Environmental data

Path and Latency Measurement

- End-to-end visibility
- Path tracing over time
- Flow latency monitoring

Network Performance

- Interface utilization
- Buffer monitoring
- Microburst detection
- Drop event correlation

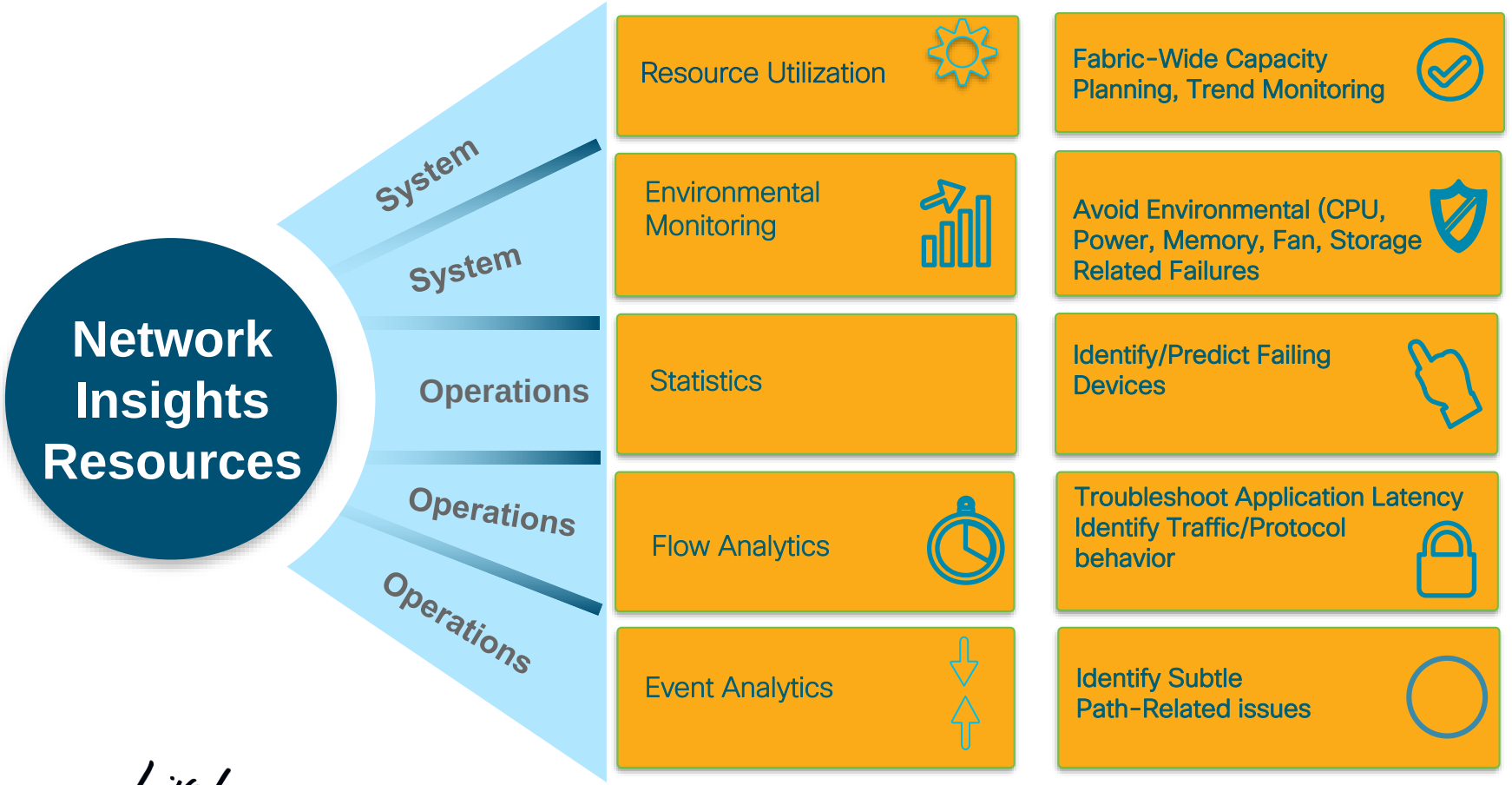


Network Insights Resources

- *Analysis and correlation* of software and hardware telemetry data with focus on Day 2 network operations use-cases
- Focus on *identifying anomalies* and *providing quick drill-down* to specific issues

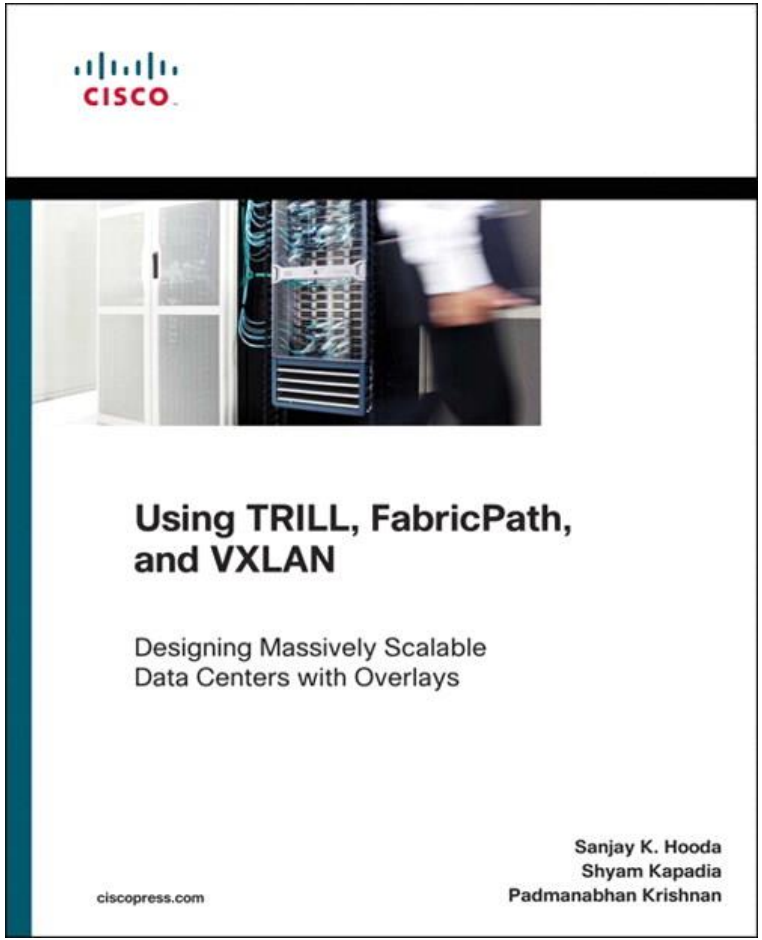
NIR

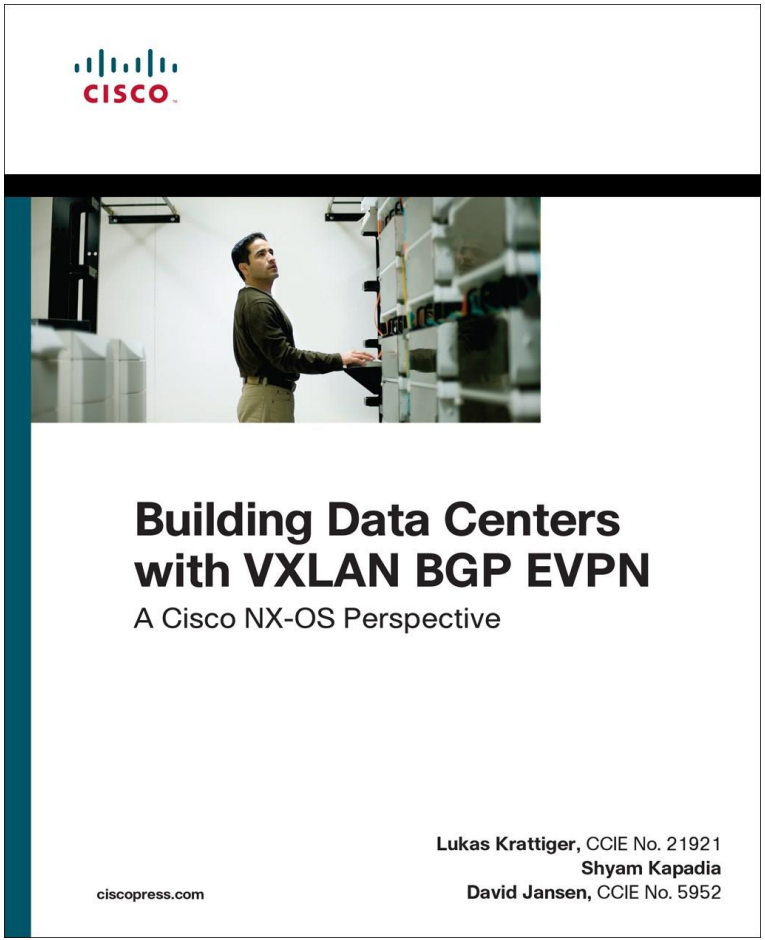
Network Insights Resources – Customer Benefits



Key Takeaways

- Modern Overlays require Modern Tools
- ICMP isn't sufficient
  - Who can Ping What, When?
- VXLAN OAM/NGOAM for Single & Multi-Site deployments
- Pro-active OAM
  - Enables use cases like Xconnect tunnel monitoring
- Use all available tools for quicker resolution
  - VXLAN OAM, Endpoint Locator, Network Insights with DCNM





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

cisco Live!

You make **possible**