



IP-EVPN-Multitenancy

Bringing web-scale networking

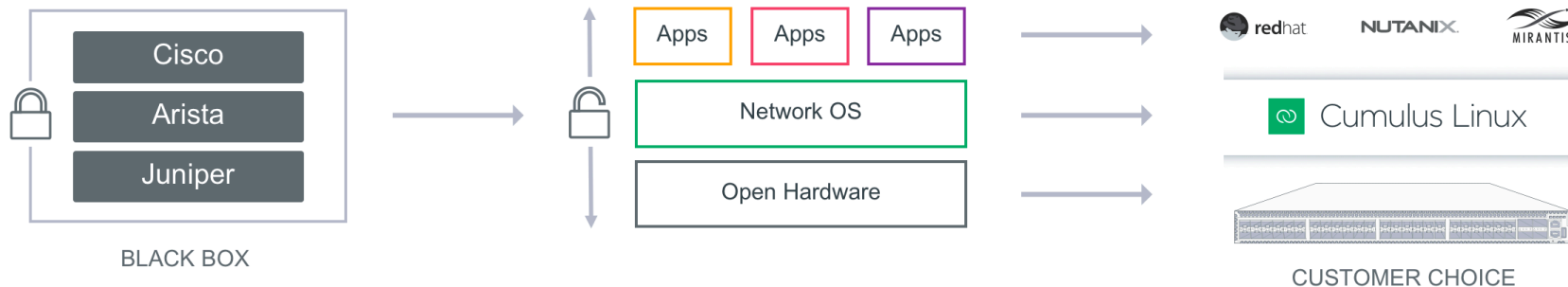
May 30th, 2017

Atilla de Groot | SE, HCIE #3494 | Cumulus Networks

Cumulus Networks brings Web-Scale Networking to Enterprise Cloud



Unlocking the vertical network stack to build the modern data center



1

Economical scalability

With commodity hardware and a standardized Linux stack, achieving a lower TCO by up to 60%

2

Built for the automation age

Making networking repeatable and consistent

3

Standardized toolsets

Easily enable Linux tools: automation, monitoring, analytics...

4

Choice and flexibility

50+ hardware platforms, from 11 vendors, and 2 silicon

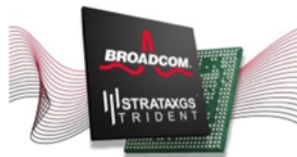
Open hardware and disaggregation

Opening up possibilities



Juniper	Cisco	Arista	Extreme	Brite-Box White Box
 QFX 3500	 Nexus 3100 Nexus 9000	 7500E 7250X	 X770	 HCL: Agema, Edge-Core, Penguin Computing, Quanta, Dell, HP, Mellanox   

powered by

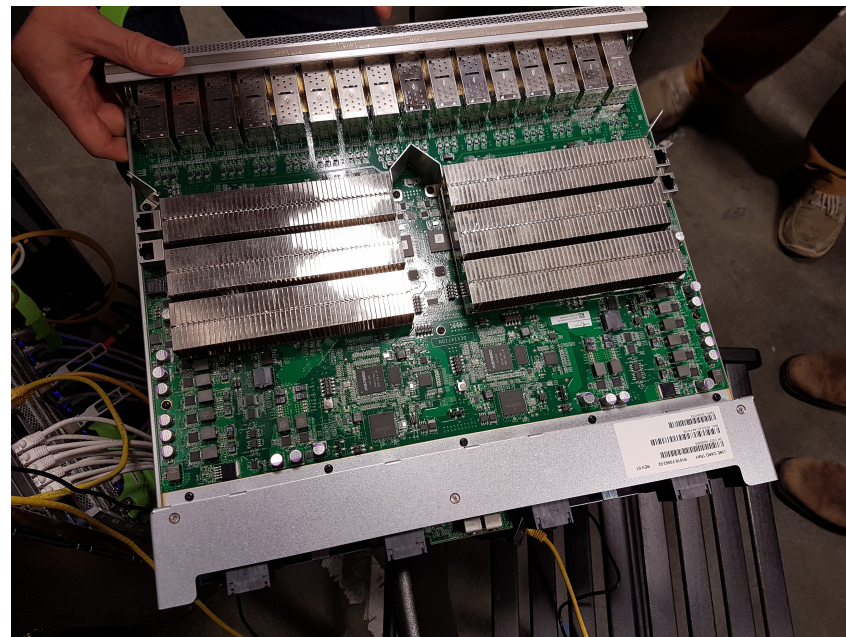
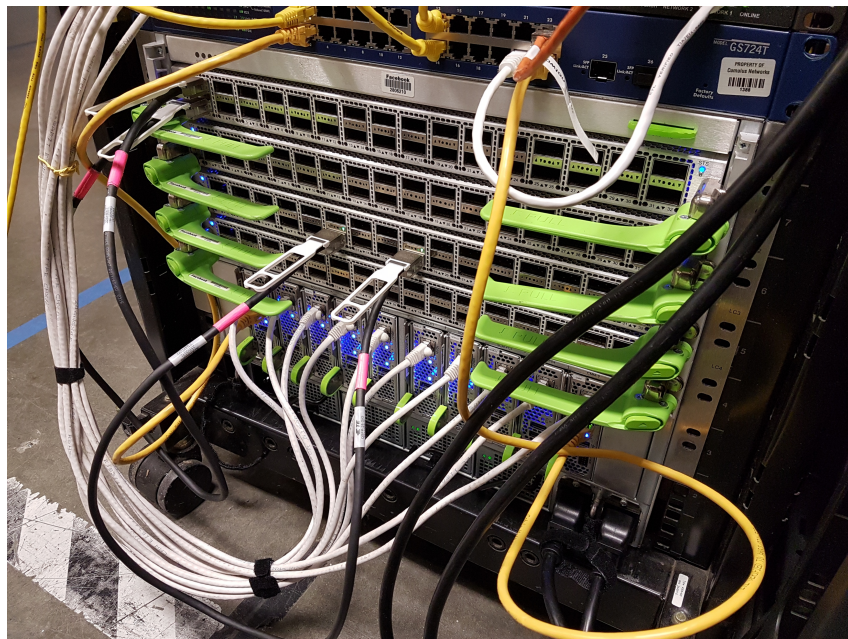


Trident II



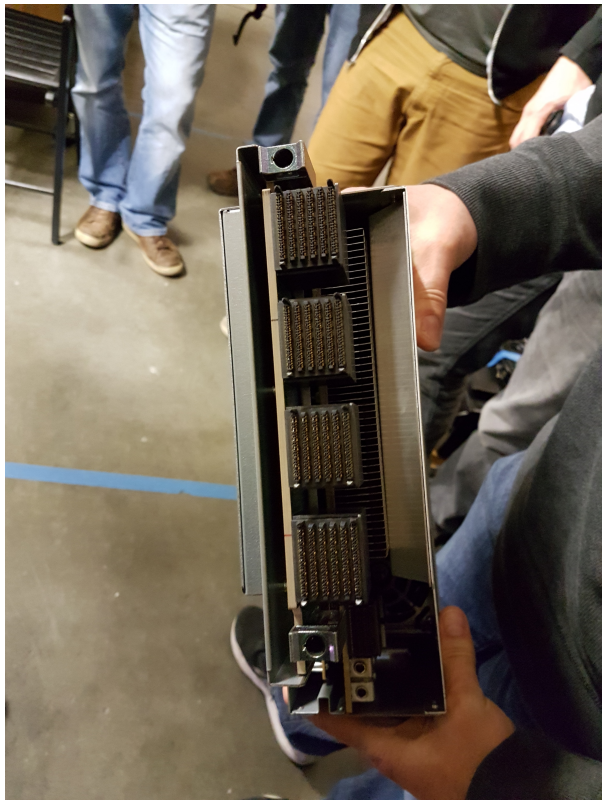
Open hardware and disaggregation

Facebook Backpack



Open hardware and disaggregation

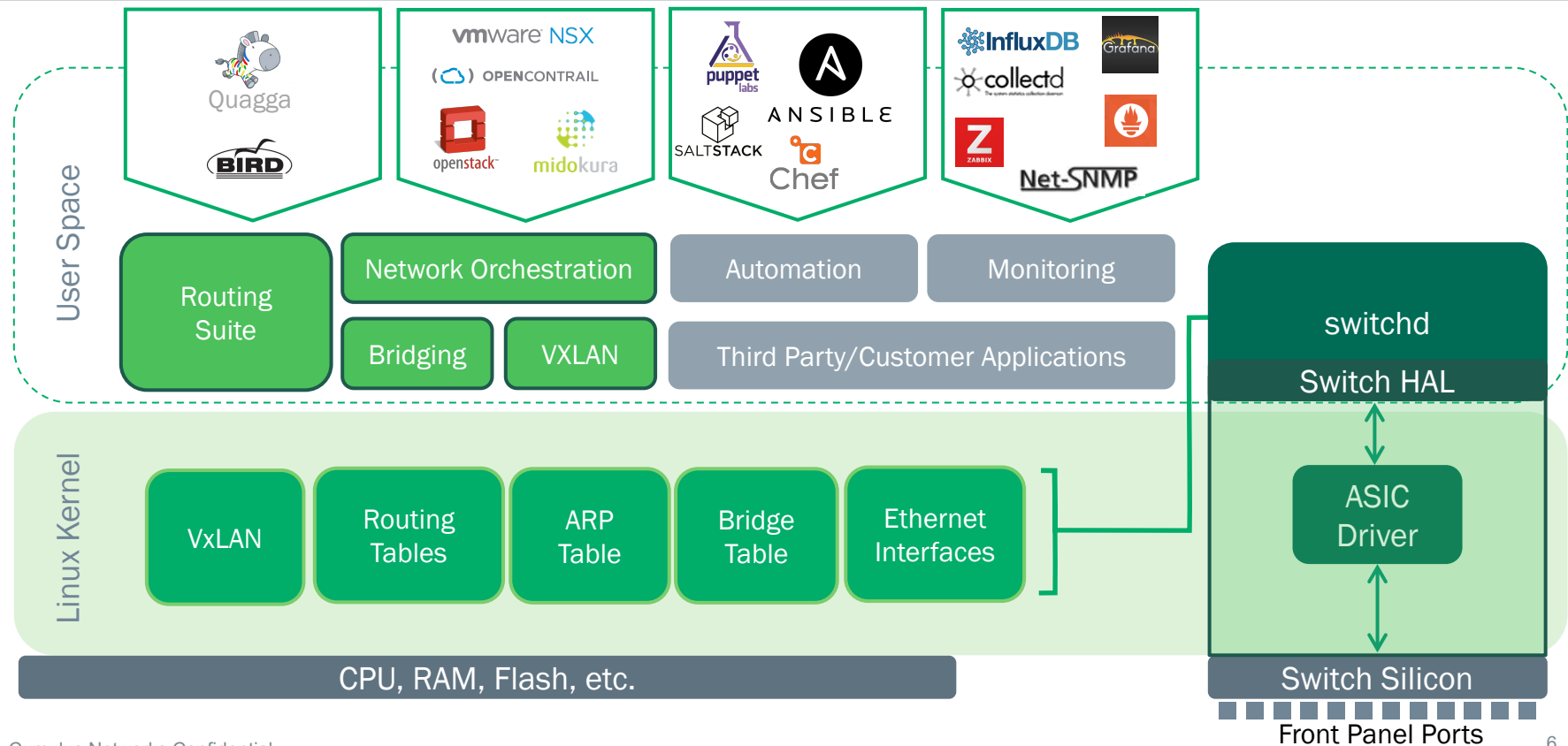
Facebook Backpack





Cumulus Linux architecture

Uniform operating model – write any tools, use any apps





Open networking concept

- ONIE
- APD

Linux networking development

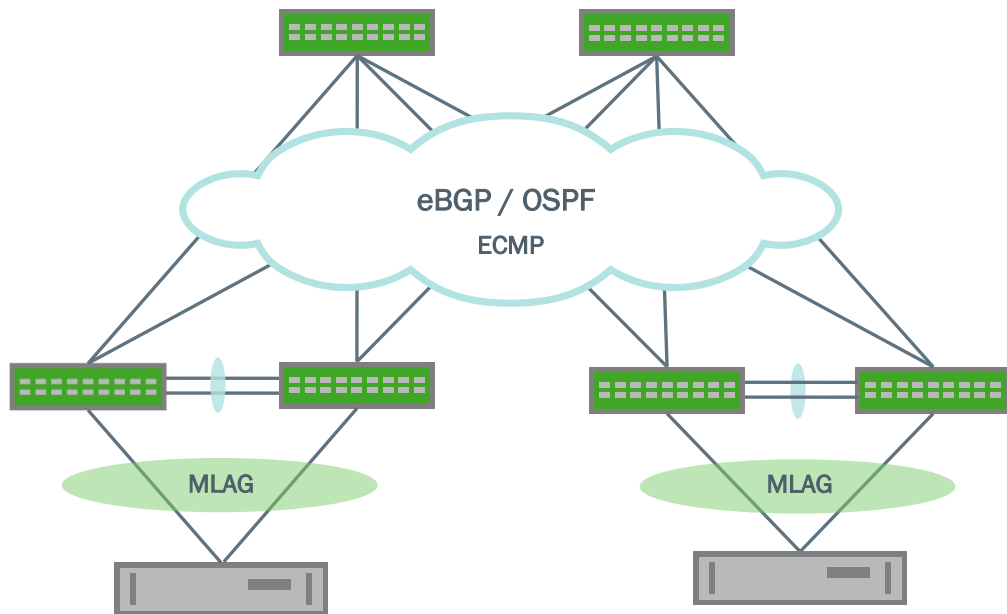
- VRF
- Quagga / Free Range Routing
- Ifupdown2
- Iproute2



L3 to ToR

L3 to the ToR

- CLOS Fabric
- Routing protocol to TOR
- LACP from servers
- One device from the server
- ECMP over multiple spines
- No STP in the backbone





Optimizing configuration

BGP Unnumbered

- Peers on IPv6 link-local
- Neighbor detection on RA messages
- RFC 5549
IPv4 over IPv6 neighbor
- Removing unnecessary variables
- Troubleshooting additions

Interfaces

```
auto lo
iface lo inet loopback
        address 10.100.1.1

auto swp1
iface swp1
        address 172.30.1.1/30
        address 2001:DB8:1::1/64

auto swp2
iface swp2
        address 172.30.1.5/30
        address 2001:DB8:2::1/64
```

```
auto lo
iface lo inet loopback
        address 10.100.1.1

auto swp1
iface swp1

auto swp2
iface swp2
```

BGP configuration

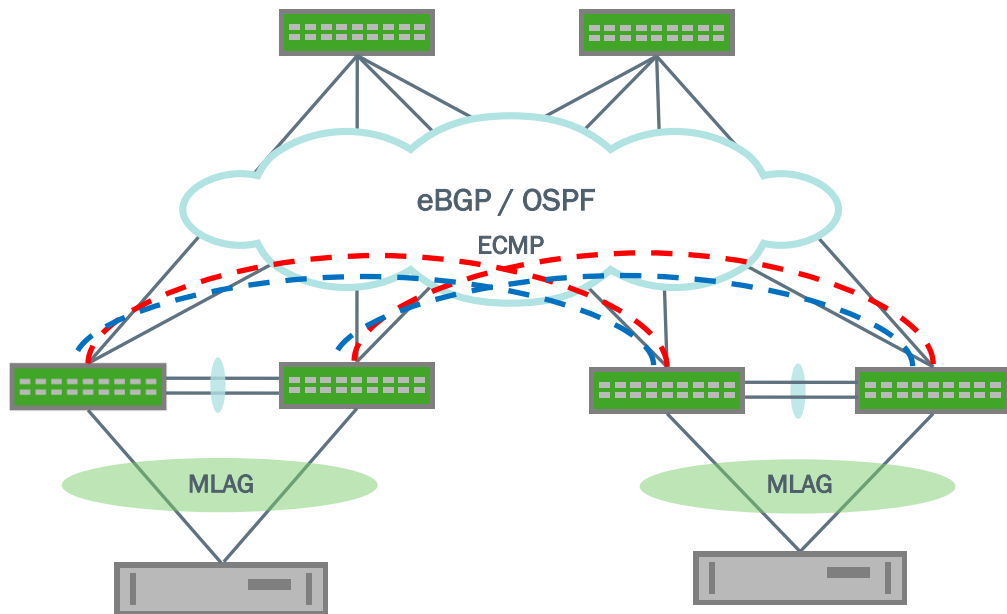
```
router bgp 64501
bgp log-neighbor-changes
bgp router-id 10.0.0.1
!
neighbor 10.1.1.1 remote-as 65000
neighbor 10.1.2.1 remote-as 65000
neighbor 10.1.3.1 remote-as 65000
neighbor 10.1.4.1 remote-as 65000
```

```
router bgp 64501
bgp log-neighbor-changes
bgp router-id lo
!
neighbor swp1 remote-as external
neighbor swp2 remote-as external
neighbor swp3 remote-as external
neighbor swp4 remote-as external
```

Overlay Networking

Overlay networking

- Routing protocol to TOR
- VxLAN encapsulation
- Active/Active redundancy
- EVPN Control plane

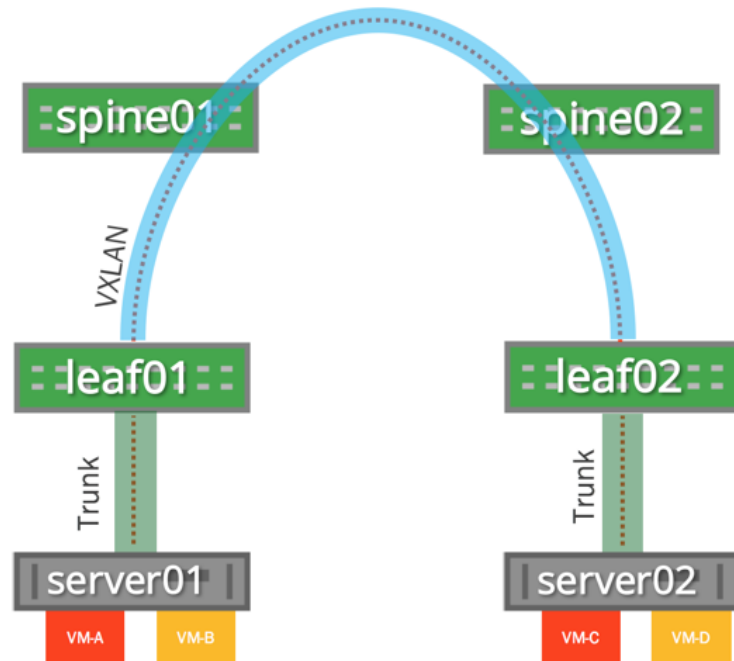




Overlay Networking

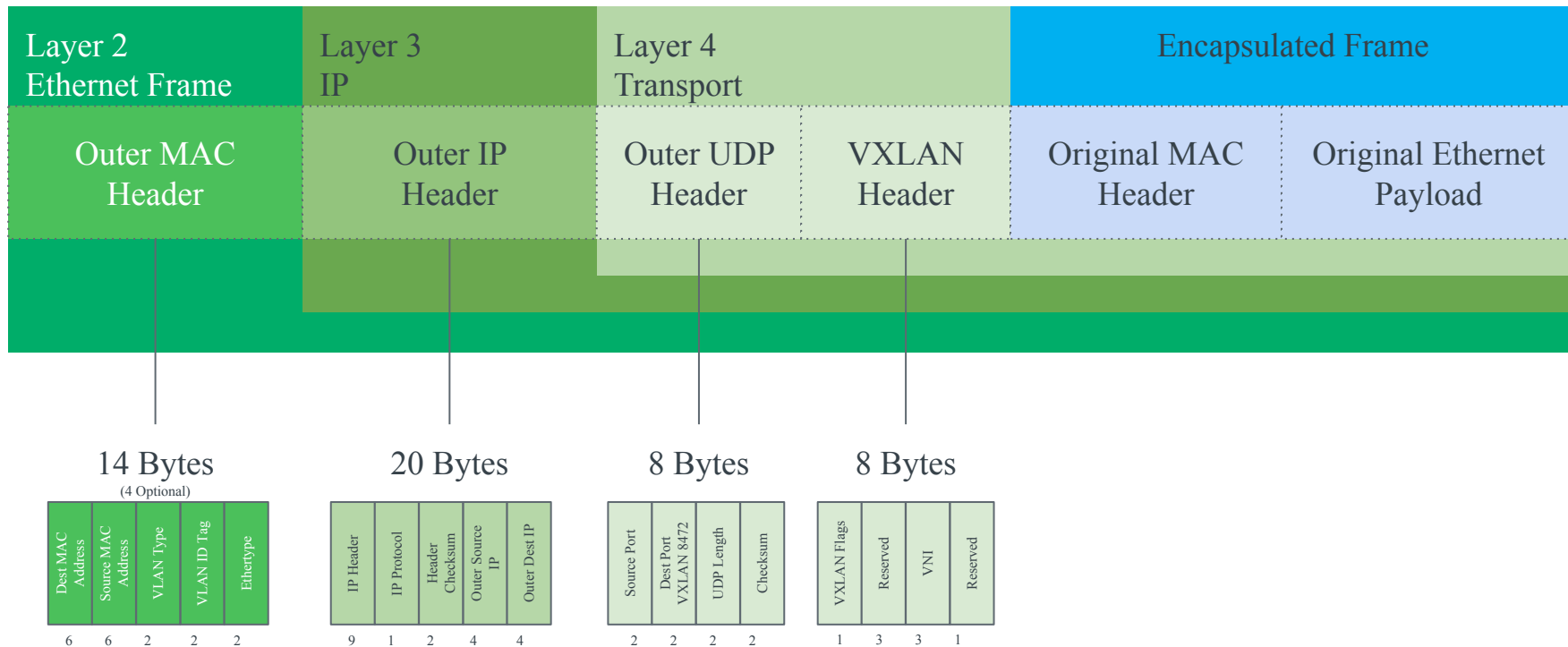
L2 adjacency between racks

- Simple L2 tunnel
- Built over IP
- VLANs or bridges are tied to VxLANs
- L2 functionality with L3 stability
- Hardware Acceleration

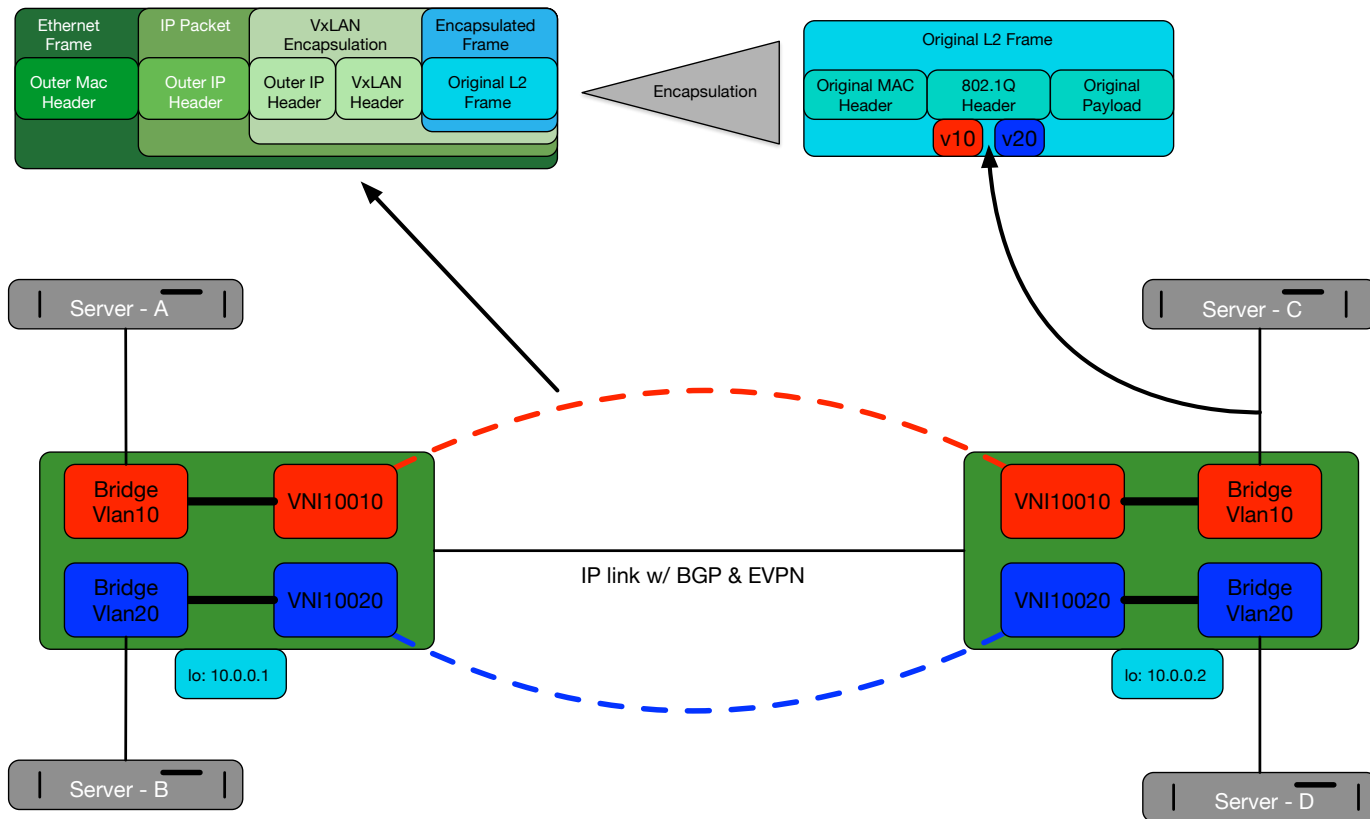




VxLAN Encapsulation overhead



VxLAN encapsulation



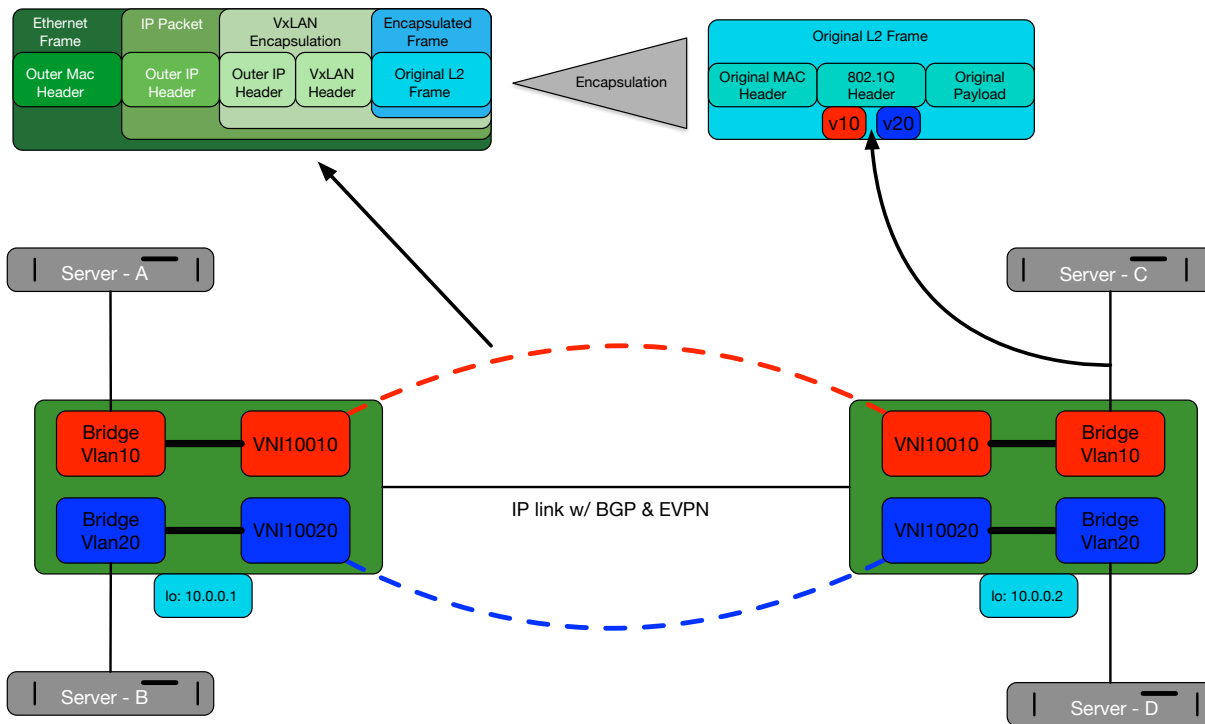


VxLAN configuration

```
auto bridge
iface bridge
    bridge-vlan-aware yes
    bridge-ports uplink regex vni.*
    bridge-stp on
    bridge-vids 10 20
    bridge-pvid 1
```

```
auto vni-10010
iface vni-10010
    alias CUSTOMER X VLAN 10
    vxlan-id 10010
    bridge-access 10
    vxlan-local-tunnelip 10.0.0.1
    vxlan-remoteip 10.0.0.2
```

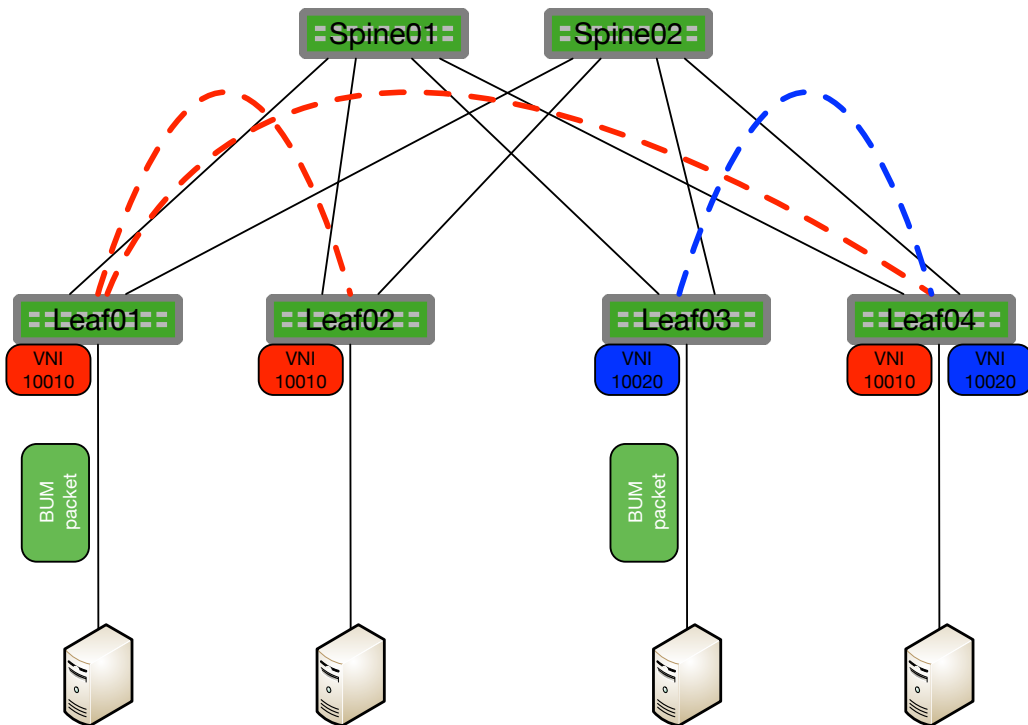
```
auto vni-10020
iface vni-10020
    alias CUSTOMER X VLAN 20
    vxlan-id 10020
    bridge-access 20
    vxlan-local-tunnelip 10.0.0.1
    vxlan-remoteip 10.0.0.2
```



VxLAN dataplane learning

Head end replication

- RFC 7348
 - Multicast replication
- HER:
 - Anycast replication
- Only sent to interested VTEPs
- Asic handling



Next-generation “L2VPN” control plane

Based on Standards:

- RFC 7432
- draft-ietf-bess-evpn-overlay

For VXLAN, EVPN uses MP-BGP (RFC4760) to exchange MAC addresses between VTEPs

- BGP is very scalable
 - BGP MAC routing
- Switch advertises local MAC addresses to all VTEPS within a VXLAN

Exchanging both MAC address and IP address allows for:

Inter-VXLAN routing (Future T2+, Spectrum Only)
Local VTEP to reply to ARP request from local host

REDUCES BUM TRAFFIC ACROSS VXLAN TUNNELS



EVPN Type3

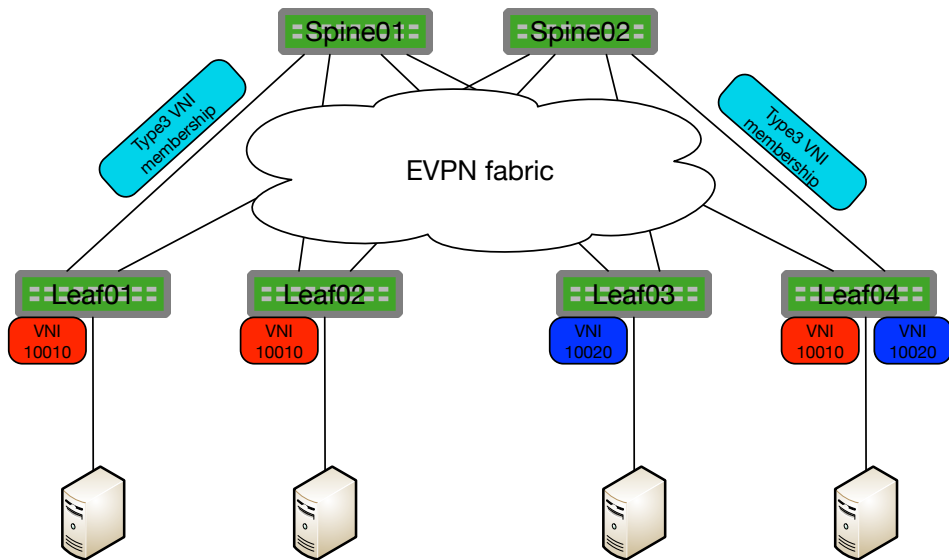
Type3 VNI Membership

Route Distinguisher

Ethernet Tag ID

IP Address Length

Origin IP



```
cumulus@leaf01:mgmt-vrf:~$ net show bgp evpn vni
Advertise All VNI flag: Enabled
Number of VNIs: 1
Flags: * - Kernel
  VNI    Orig IP    RD          Import RT    Export RT
* 10100  10.100.100.1 10.0.0.11:1 65011:10100 65011:10100
```



EVPN Type2

EVPN Type2

- Mac-IP route advertisement
- BUM suppression
- Proxy arp

```
auto vtep100
iface vtep100
    bridge-access 10
    bridge-arp-nd-suppress on
    bridge-learning off
    vxlan-id 10010
    vxlan-local-tunnelip 10.0.0.1
```

Route Distinguisher

Ethernet Segment Identifier

Ethernet Tag ID

MAC Address length

MAC Address

IP Address Length

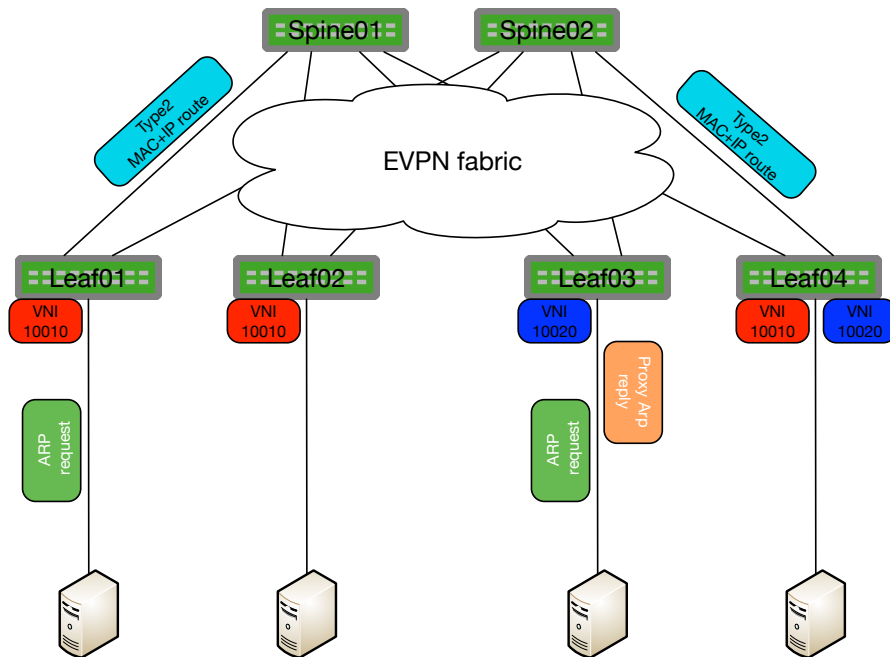
IP Address

VNI-ID

EVPN Broadcast suppression

Broadcast suppression

- Local ARP proxy
- Translation to MAC+IP route
- IPv6 ND support
- Smaller L2 “domains”
- Legacy applications



```
cumulus@leaf01:mgmt-vrf:~$ net show evpn mac vni all
```

VNI 10100 #MACs (local and remote) 4

MAC	Type	Intf/Remote	VTEP	VLAN
46:38:39:00:00:5e	remote	10.100.100.5		
46:38:39:00:00:6a	local	server01		100
46:38:39:00:00:2a	local	server02		100
46:38:39:00:00:44	remote	10.100.100.5		

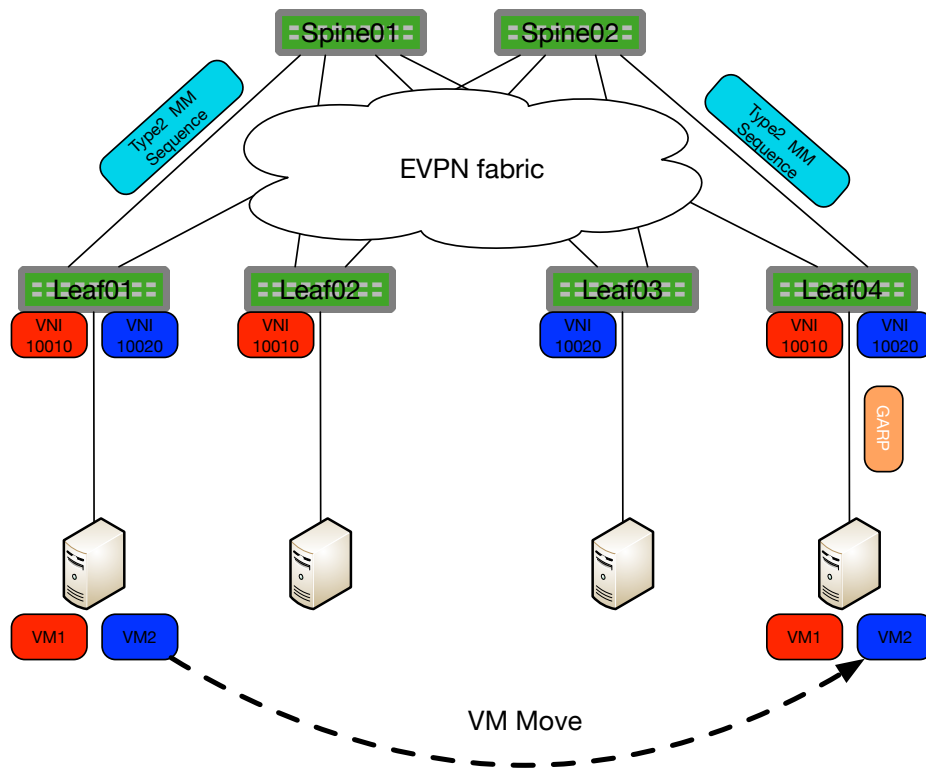
EVPN Mac mobility

Mac mobility

- Local ARP proxy
- VM movement
- Gratuitous ARP

```
cumulus@leaf01:mgmt-vrf:~$ net show bgp evpn route vni
10100 mac 46:38:39:00:00:5e
```

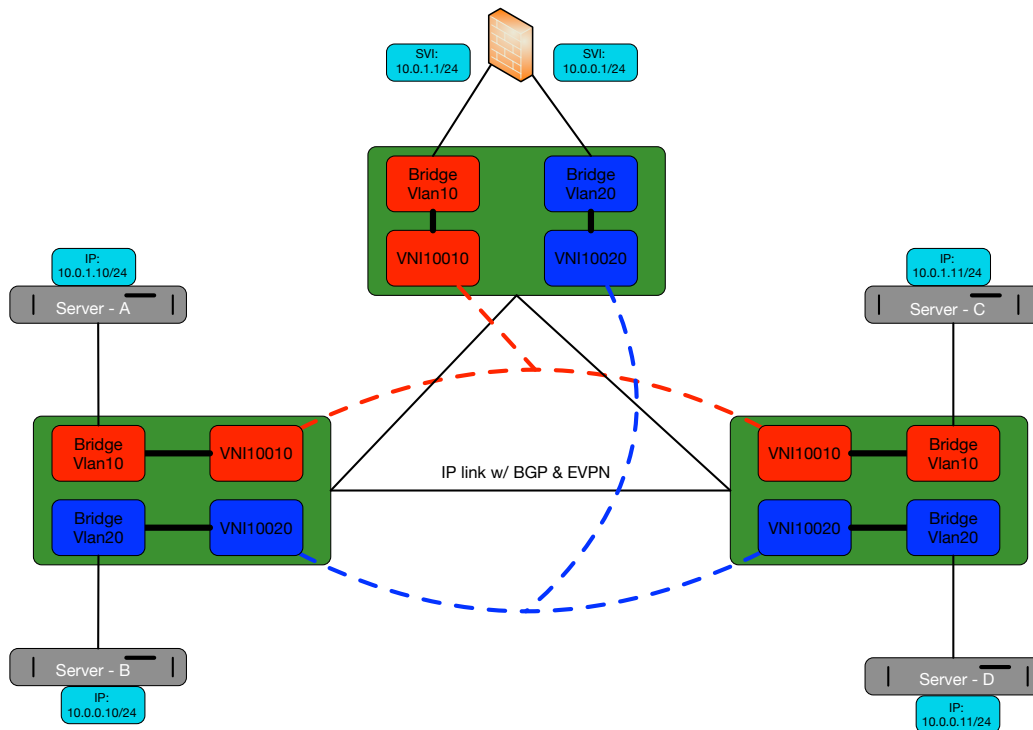
```
BGP routing table entry for [2]:[0]:[0]:[48]:[46:38:39:00:00:5e]
Paths: (2 available, best #2)
Not advertised to any peer
Route [2]:[0]:[0]:[48]:[46:38:39:00:00:5e] VNI 10100
Imported from 10.0.0.15:1:[2]:[0]:[0]:[48]:[46:38:39:00:00:5e]
65020 65030 65021 65015
10.100.100.5 from spine01(swp49) (10.0.0.21)
Origin IGP, localpref 100, valid, external
Extended Community: RT:65015:10100 ET:8 MM:3
AddPath ID: RX 0, TX 571
Last update: Tue May 2 10:31:21 2017
```



Routing in VxLAN

Router on a stick

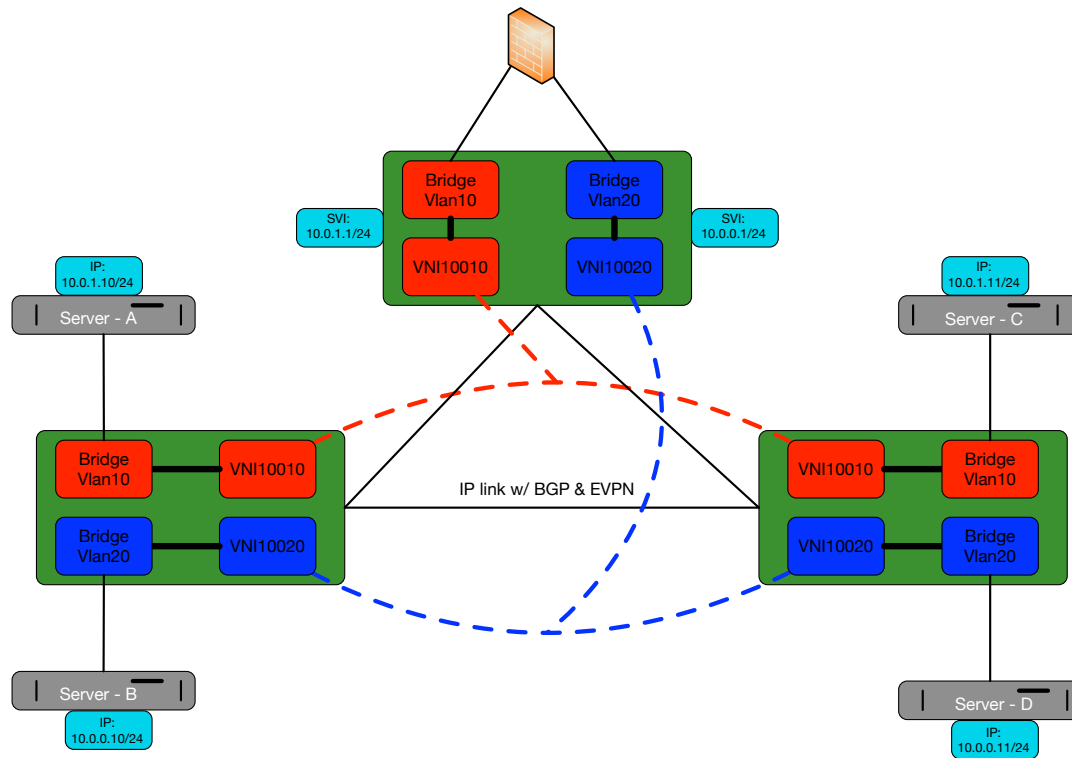
- Traffic tromboning
- Bandwidth limitation
- Costs



VxLAN routing

VxLAN routing

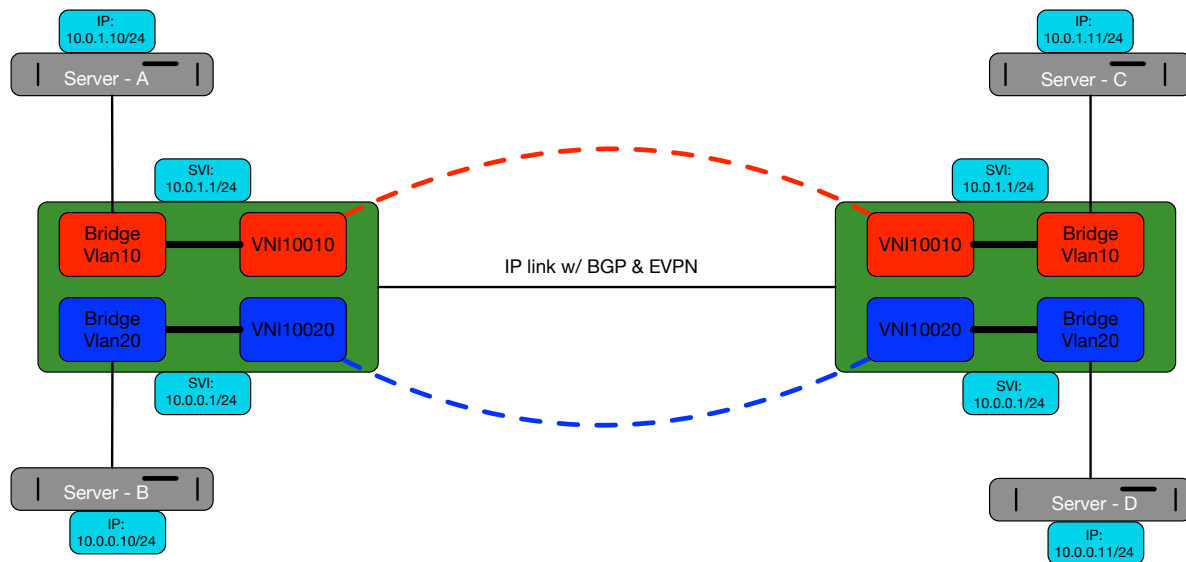
- Hardware dependency
 - Trident2+
 - TH+
 - Spectrum
- Centralized routing
- Distributed / Anycast routing
- Asymmetric routing
- Symmetric routing



VxLAN routing

VxLAN routing

- Hardware dependency
 - Trident2+
 - TH+
 - Spectrum
- Centralized routing
- Distributed / Anycast routing
- Asymmetric routing
- Symmetric routing





VxLAN routing

VxLAN routing

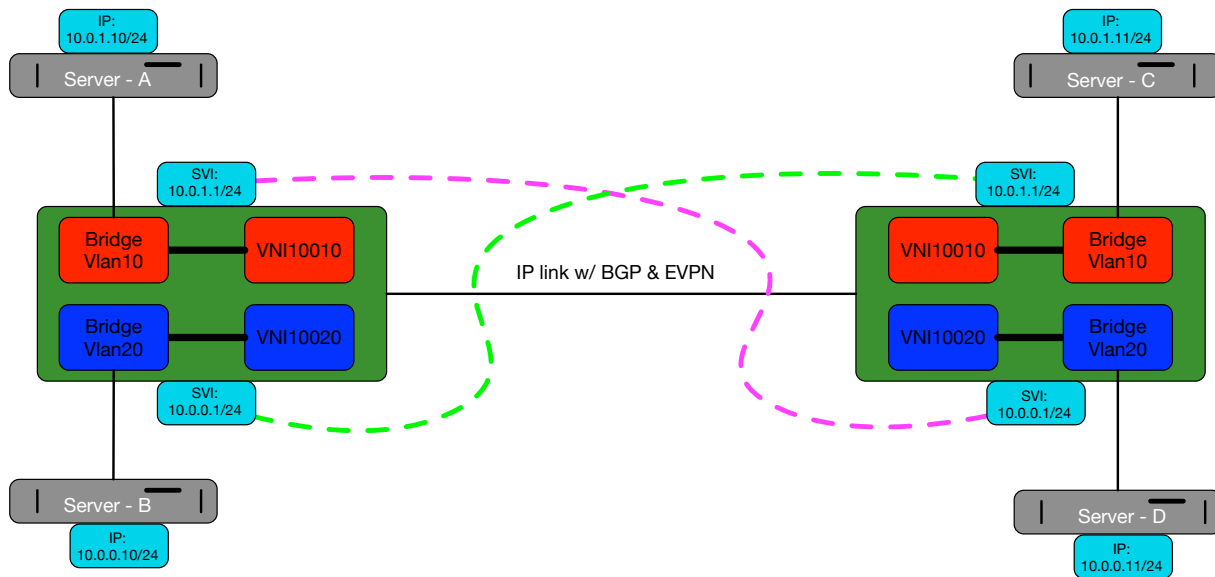
- Hardware dependency

Trident2+

TH+

Spectrum

- Centralized routing
- Distributed / Anycast routing
- Asymmetric routing
- Symmetric routing





There is more!





Future EVPN features

EVPN Routing

- L3 Multitenancy
- Anycast routing
- Symmetric routing

EVPN Bonding

- Type 1 & 4
- Standardized MLAG
- More than 2 TORs



L3 Multi tenancy

EVPN Type 5

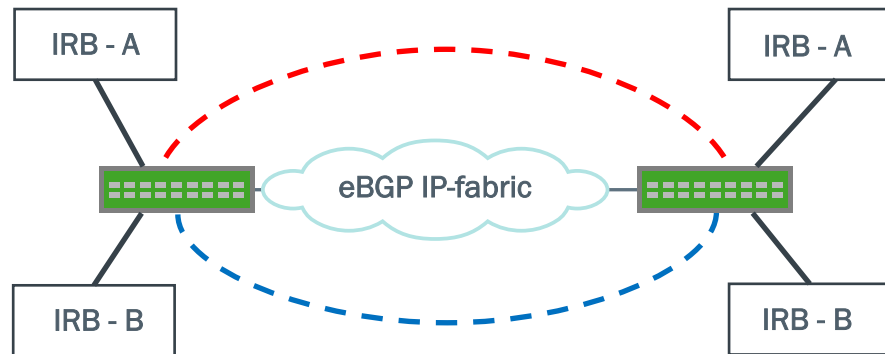
- Mac route advertisement
- Integrated routing bridging
- L3VNI

Route Distinguisher
Ethernet Segment Identifier
Ethernet Tag ID
IP Prefix Length
IP Address
GW IP Address
VNI ID

L3 Multi tenancy

EVPN Segment routing

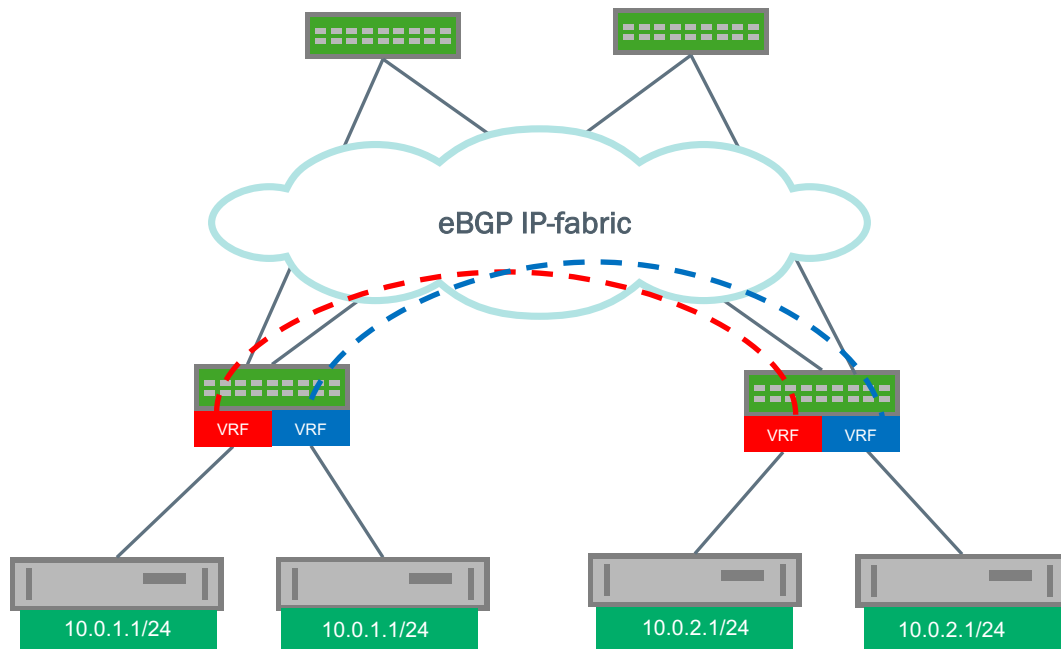
- VRFs
- Overlapping subnets
- Security boundaries
- L3VNI
- EVPN Type-5



L3 Multi tenancy

EVPN Segment routing

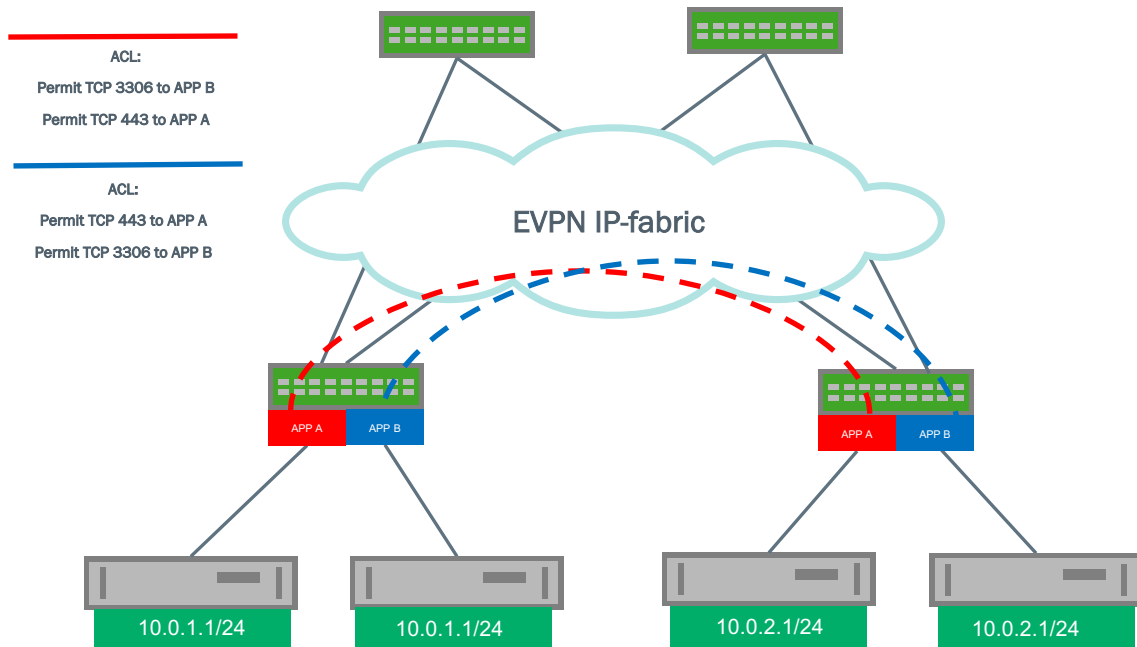
- VRFs
- Overlapping subnets
- Security boundaries
- L3VNI
- EVPN Type-5



Use case: Security

Network security

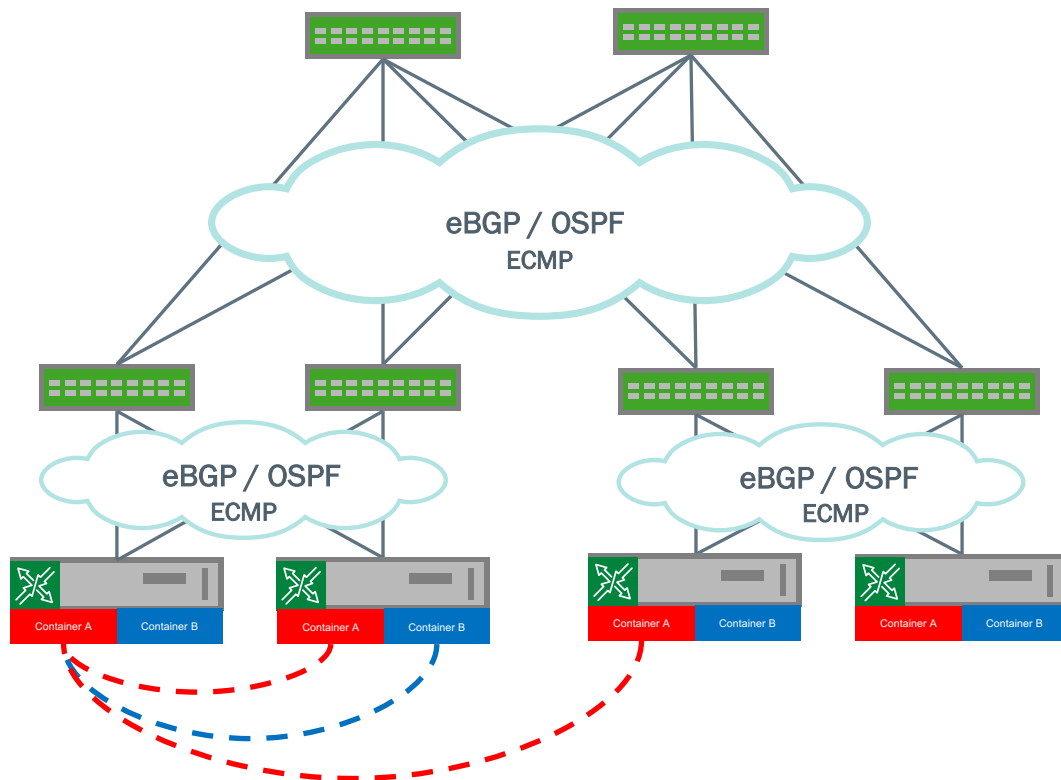
- Default deny
- Application groups
- Multi layer security



Use case: Multitenancy w/ Containers or VMs

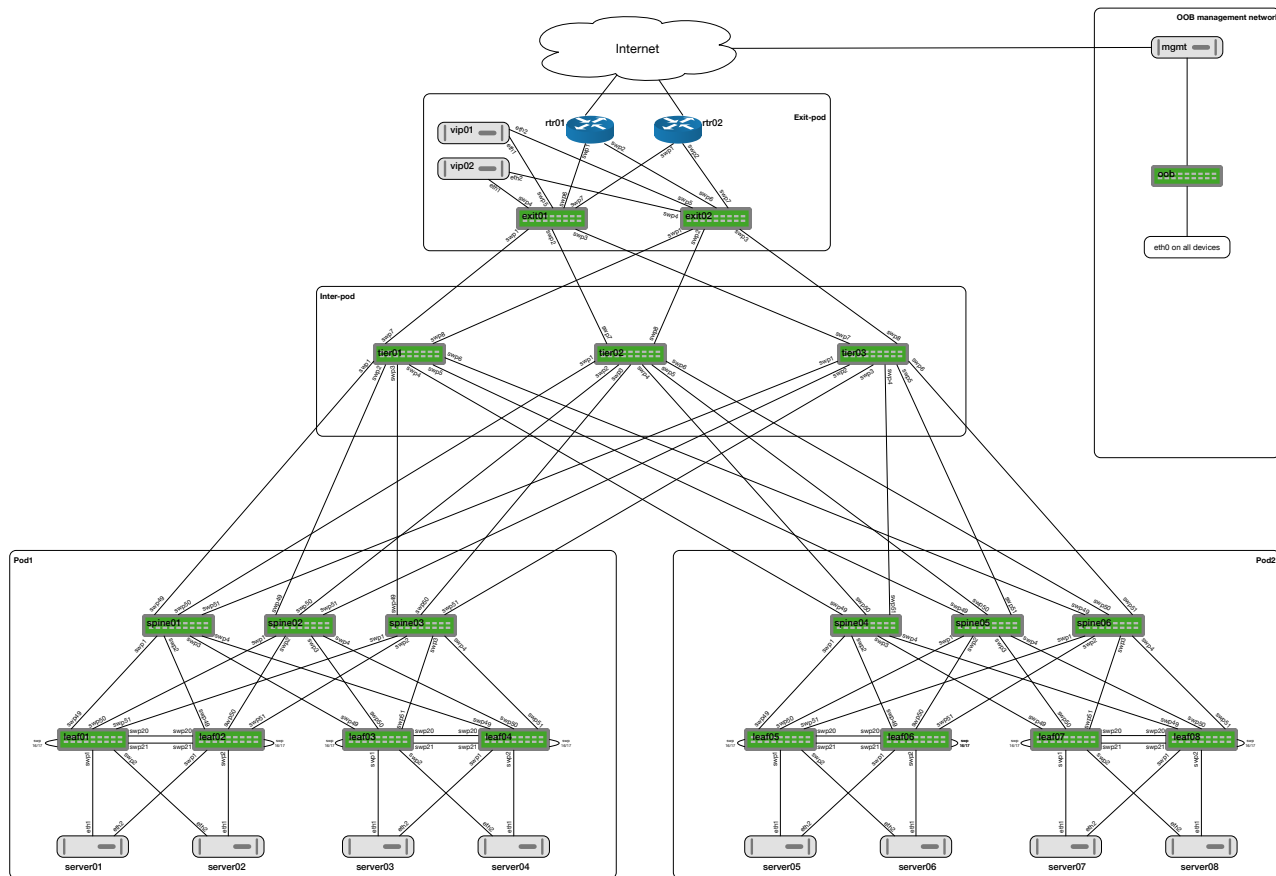
EVPN to the host

- VRFs on the host
- Segregating hypervisor infra
- Hybrid overlay





EVPN demo





Join the Cumulus Networks Community!

Get involved in the conversation:

- Join the community & share your Cumulus Linux tips and experiences: <https://getsatisfaction.cumulusnetworks.com/cumulus>
- Chat with others on our Slack channel: <https://cumulusnetworks.slack.com/>
- Keep on learning: <https://community.cumulusnetworks.com>

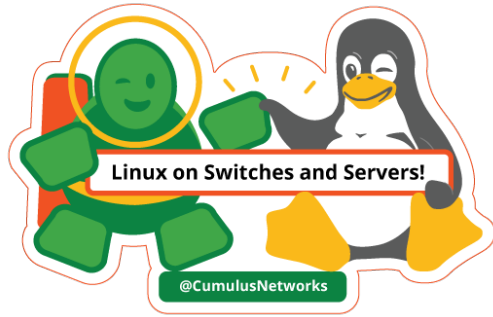


Download Cumulus VX to work on proofs of concept, test your automation and continue practicing:

- <https://cumulusnetworks.com/cumulus-vx/>

Learn at your own pace with demos using our reference topology:

- <https://github.com/CumulusNetworks/cldemo-vagrant>



Thank you!