ITNOG2 :: Bologna 3/11/2016

VxLAN with eVPN control plane for L2 and L3 distributed architectures

Nicola Modena :: CCIE #19119 :: JNCIE-SP #986
Network Consultant :: nicola@modena.to :: @nmodena

Agenda

Datacenter evolution from STP to Ethernet and IP Fabric

Vxlan encapsulation

eVPN control-plane

eVPN function and capability

Vxlan and eVPN for DCI and distributed services

Future development

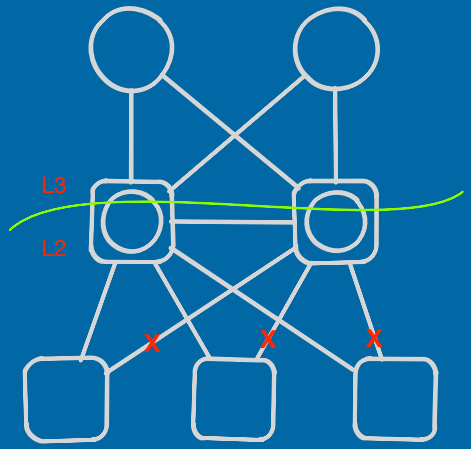
Disclaimer

All the information are results of my personal experience, test, reading and vision.

Although many things are already available, there is no guarantee that the above presented will be implemented by each vendor, platform or software versions available now or in the future.

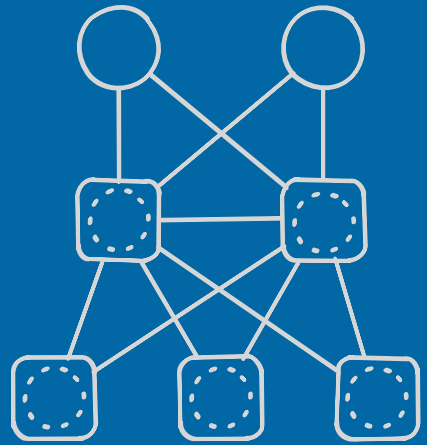
Data Center Evolution

Legacy (xSTP)



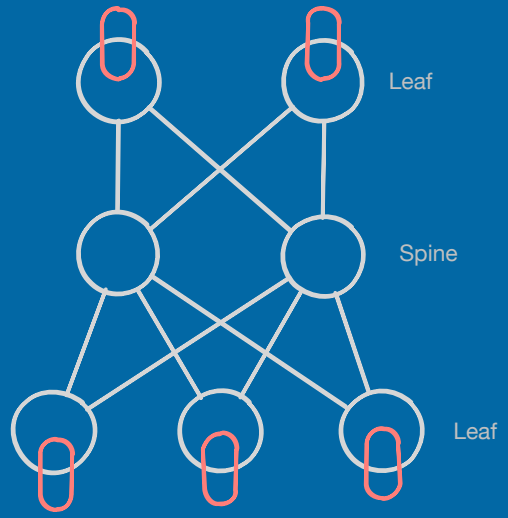
L2 Network with a loop free topology

Ethernet FABRIC



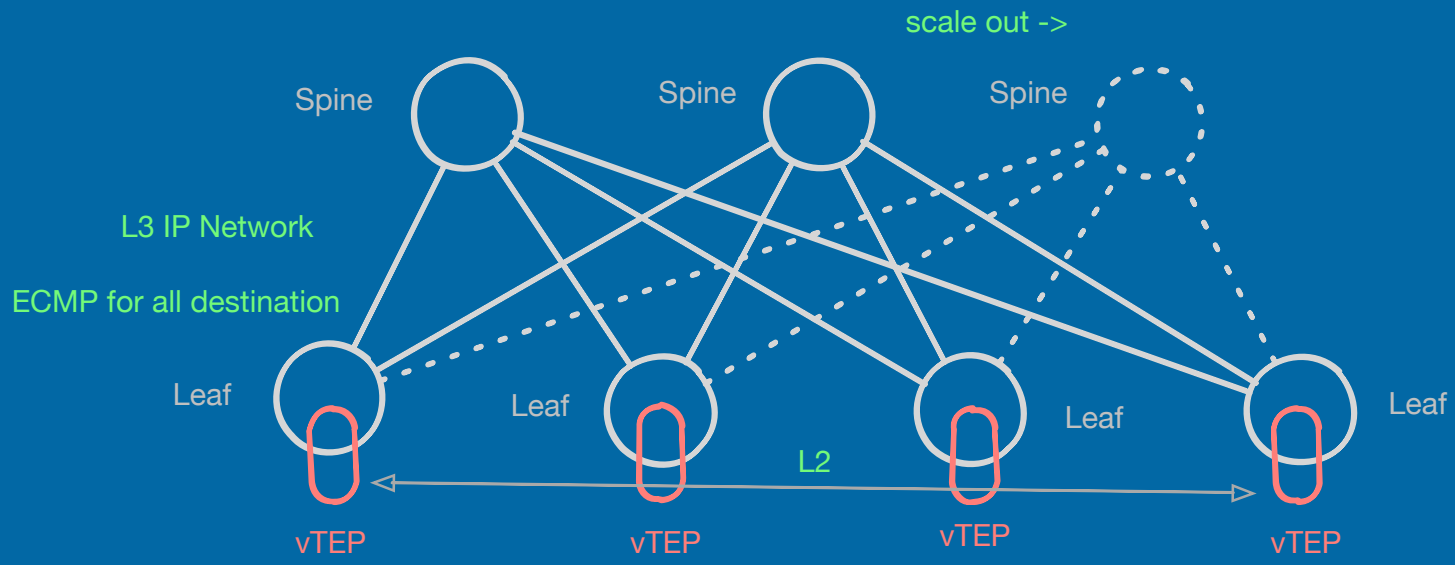
L2 Network with
L3 link-state topology with L2 packet encapsulation on rBridge

IP Fabric



L3 IP Network with
L2 packet encapsulation (overlay)

IP Fabric basics : underlay / overlay / vtep



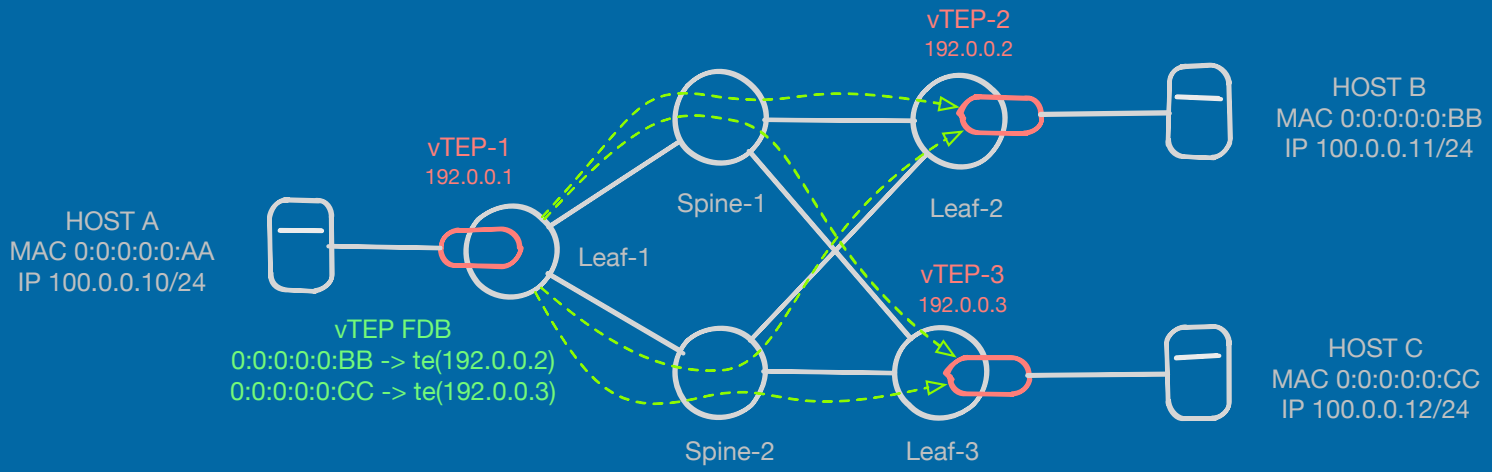
UNDERLAY: Routed IP network in CLOS (Leaf & Spine) topology
ECMP capable routing protocol (BGP as IGP is the preferred trend)
OVERLAY: Layer-2 ethernet frames are encapsulated and forwarded in IP packets
VTEP: VXlan Tunnel End Point perform encapsulation/decapsulation
OPTIONAL: Multicast support

VxLAN encapsulation

TENANT 1 : NVI 1000xxxx -> vlan 100 : VNI 10000100
TENANT 2 : NVI 1001xxxx -> vlan 100 : VNI 10010100

Virtual eXtensible Local Area Network (VXLAN) RFC7348
8 bytes header : 1 byte flags - 3 byte VNI - 4 byte reserved
Other overlay available : NvGRE / Geneve / STT / MPLS over GRE / MPLS over UDP

controller-less VxLAN operations

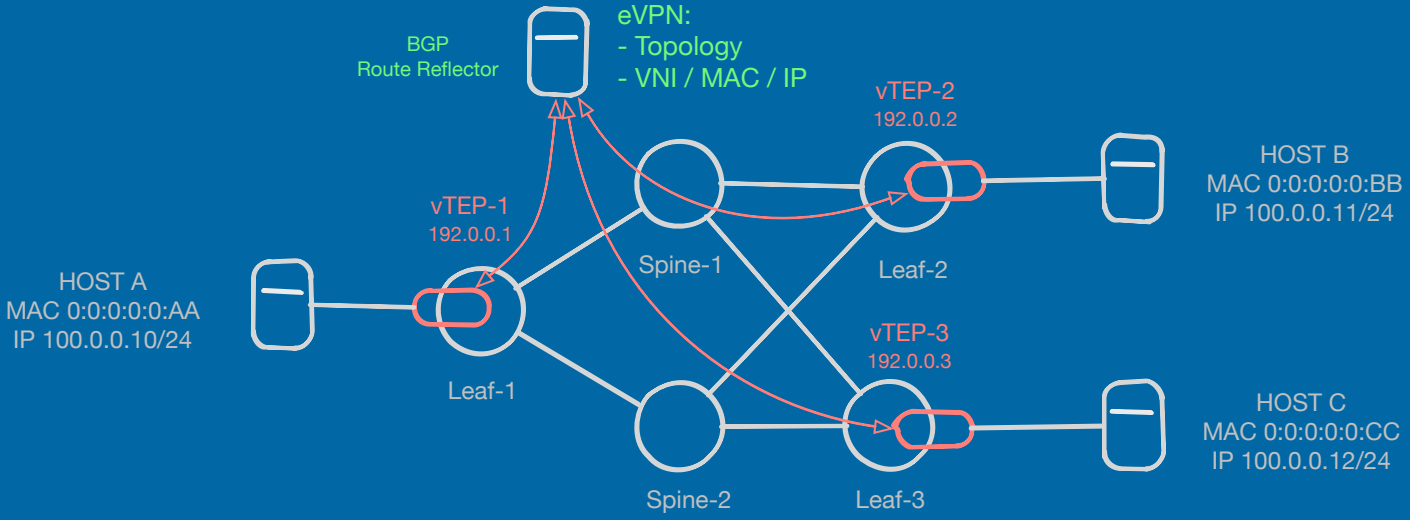


Learning based on Broadcast flooding

Broadcast distribution based on multicast or ingress-replication (with manual configuration)

Source MAC Address binding to the vTEP source IP address for unicast forwarding

eVPN control-plane for VxLAN

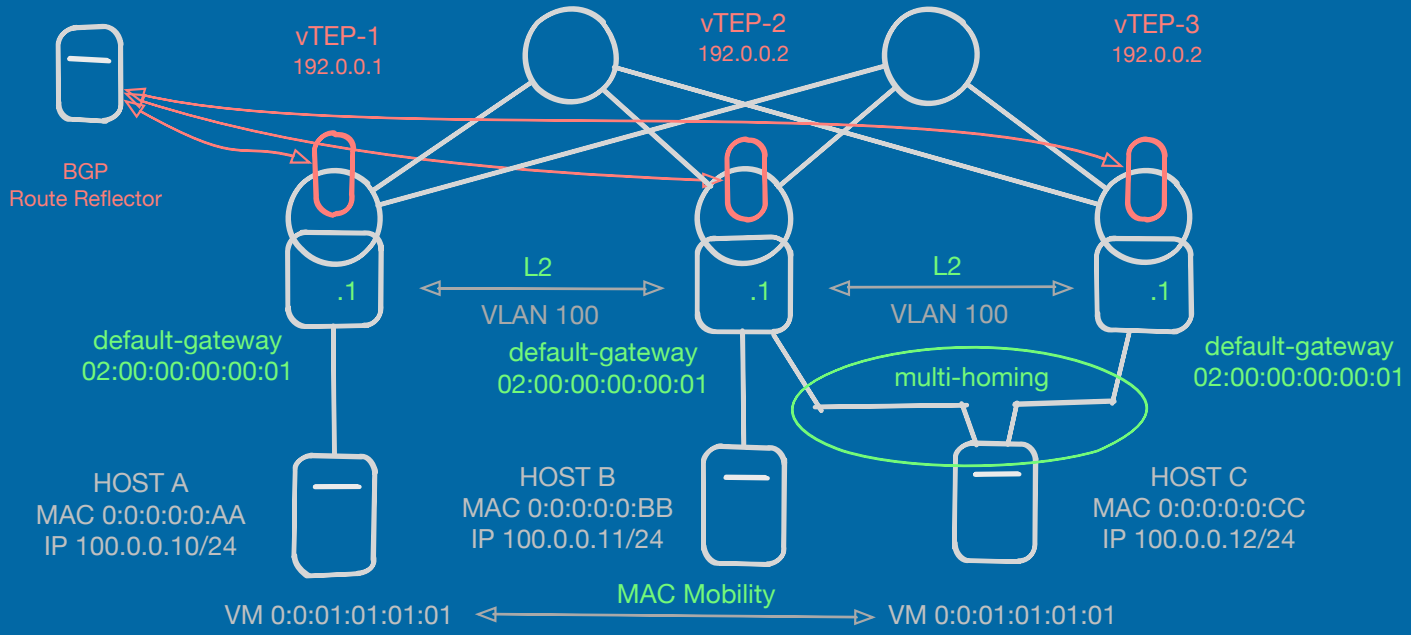


eVPN it’s a standard control-plane for Layer-2 ethernet MPLS VPN [RFC7432]

Advertise MAC addresses and topology informations in BGP

PROS regards VPLS: native layer-3 integration, fast convergence and active-active multihoming

eVPN features



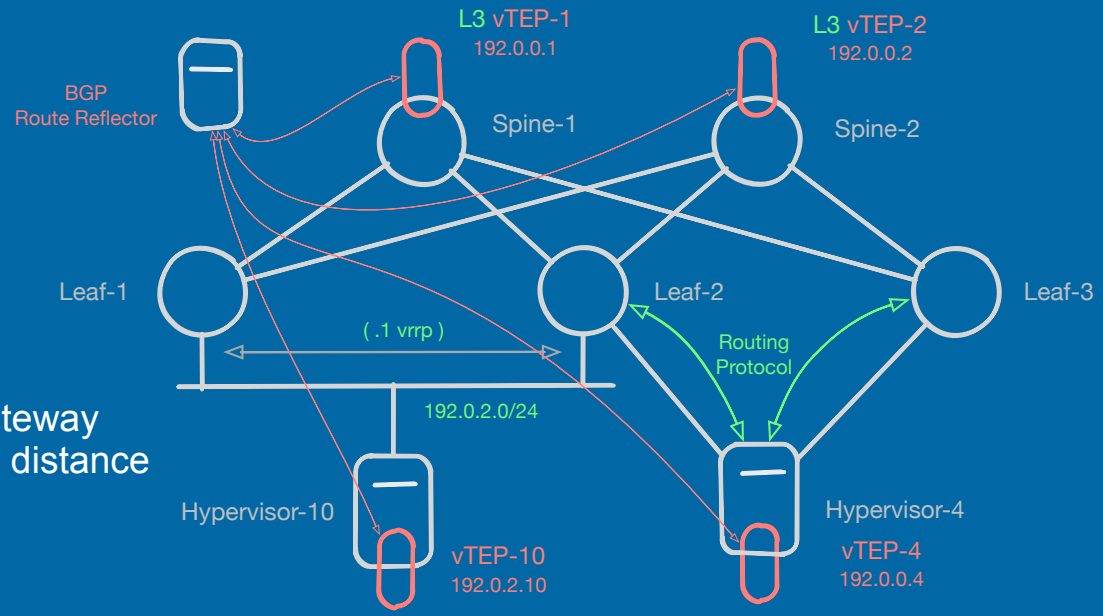
Integrated and distributed routing (IP Aware L2 & L3 services)

anycast default gateway (arp-suppression / proxy-arp)

mac-mobility for fast convergence

not mandatory in overlay implementation : active-active multihoming / massive withdraw

Alternate vTEP location



with only default-gateway
missing BGP next-hop distance

L3 vTEP in spine switch when leaf or hypervisor vTeps are only L2 (or for DCI)

for vTEP directly into hypervisor dynamic routing not mandatory: pay attention for LB and multicast

Take care that sometime “distance in overlay word” is not real

Fast Convergence (sub-second) it’s usually not required: HA and Failover handled at Layer-7

eVPN : BGP NLRI

```
Leaf-1# sh bgp l2vpn evpn 0000.0000.00AA
Route Distinguisher: 1:100    (L2VNI 50100)
BGP routing table entry for [2]:[0]:[0]:[48]:[0000.0000.00AA]:[0]:[0.0.0.0]/216, version 5
[…]
  AS-Path: NONE, path locally originated
    192.0.0.1 (metric 0) from 0.0.0.0 (192.0.0.1)
      Received label 50100
      Extcommunity:  RT:1:100

BGP routing table entry for [2]:[0]:[0]:[48]:[0000.0000.00AA]:[32]:[100.0.0.10]/272, version 18
[…]
  AS-Path: NONE, path locally originated
    192.0.0.1 (metric 0) from 0.0.0.0 (192.0.0.1)
      Received label 50100 50103
      Extcommunity:  RT:1:100 RT:1:103
```

pure Layer-2 context

L2 / L3 context
following ARP/DHCP

Label == VNI

Type-2 : MAC (and IP) Advertisement Route

Type-3 : Inclusive Mutlicast Route (also for ingress replication)

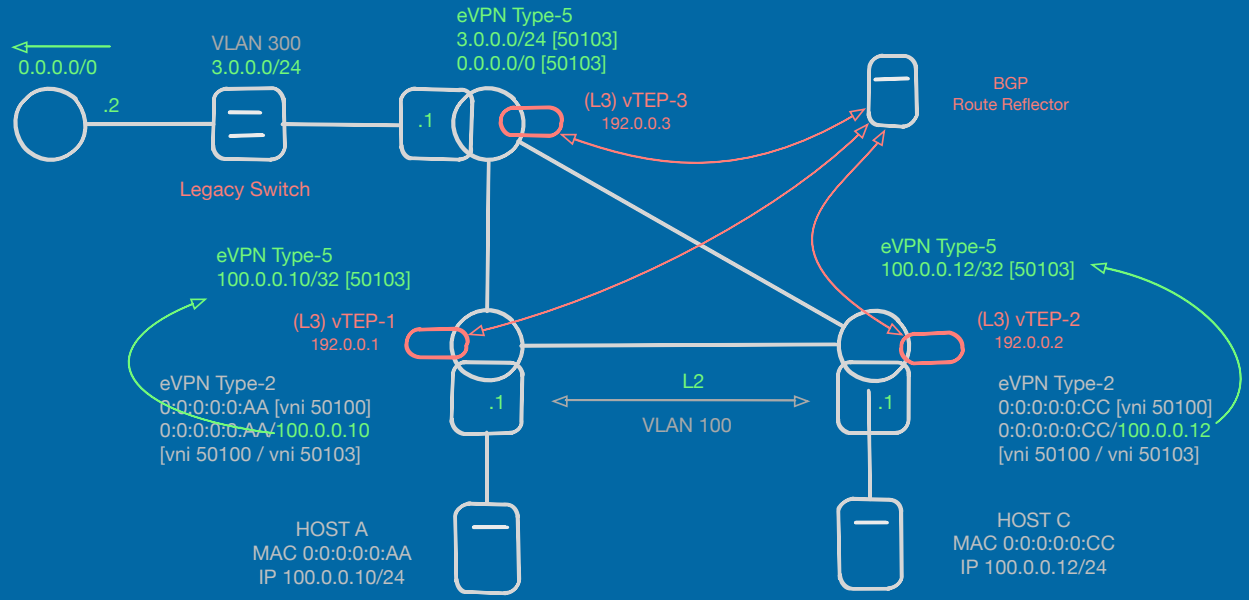
Type-5 : IP prefix Route

Not mandatory for overlay :

Type-1: Ethernet Auto-Discovery Route

Type-4 : Ethernet Segment Route

eVPN Integrated routing

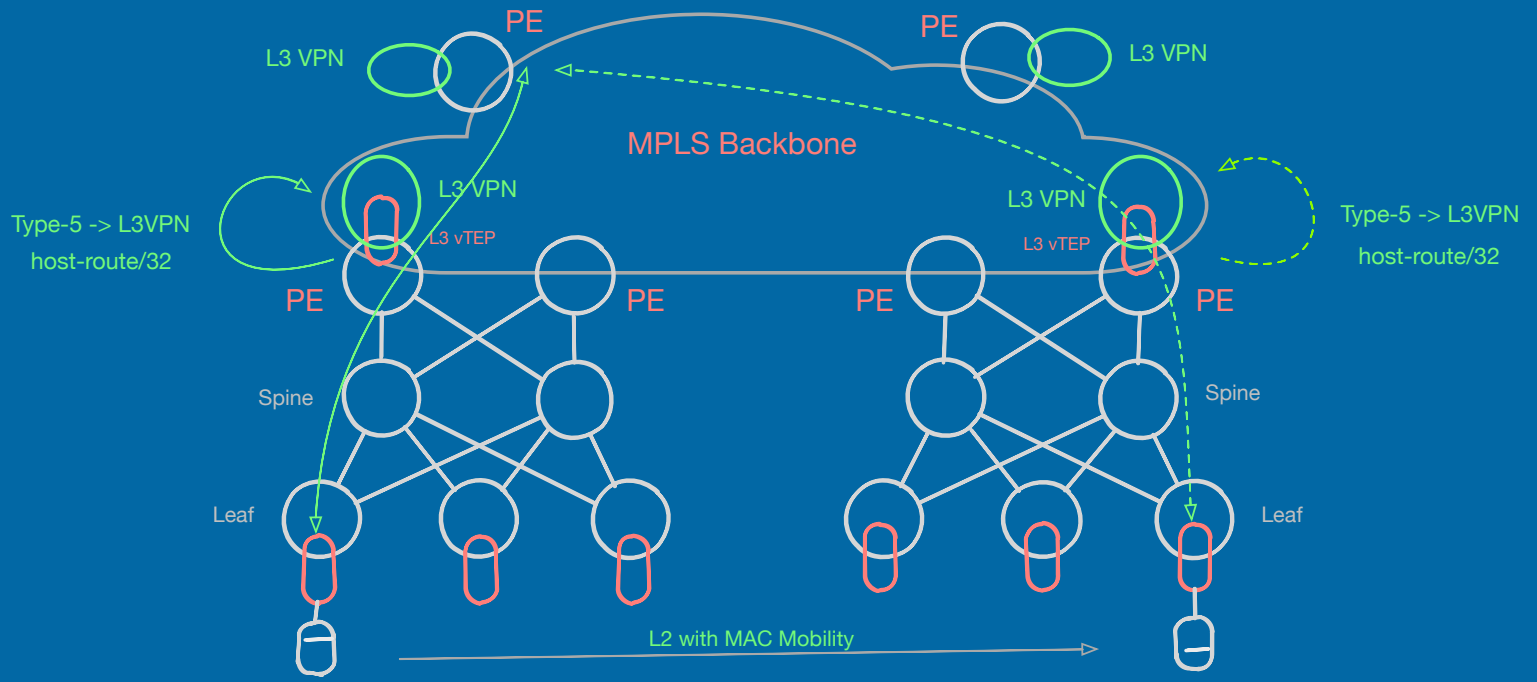


Type-5 routes enables routing integration without involving other AFI/SAFI (L3VPN) (es. external routing)

Can be used to automatically advertise local discovered hosts with a /32 netmask

Capable vTEP can perform proxy-arp for remot host in the same vlan without L2 (Type-2) route

Optimal service delivering

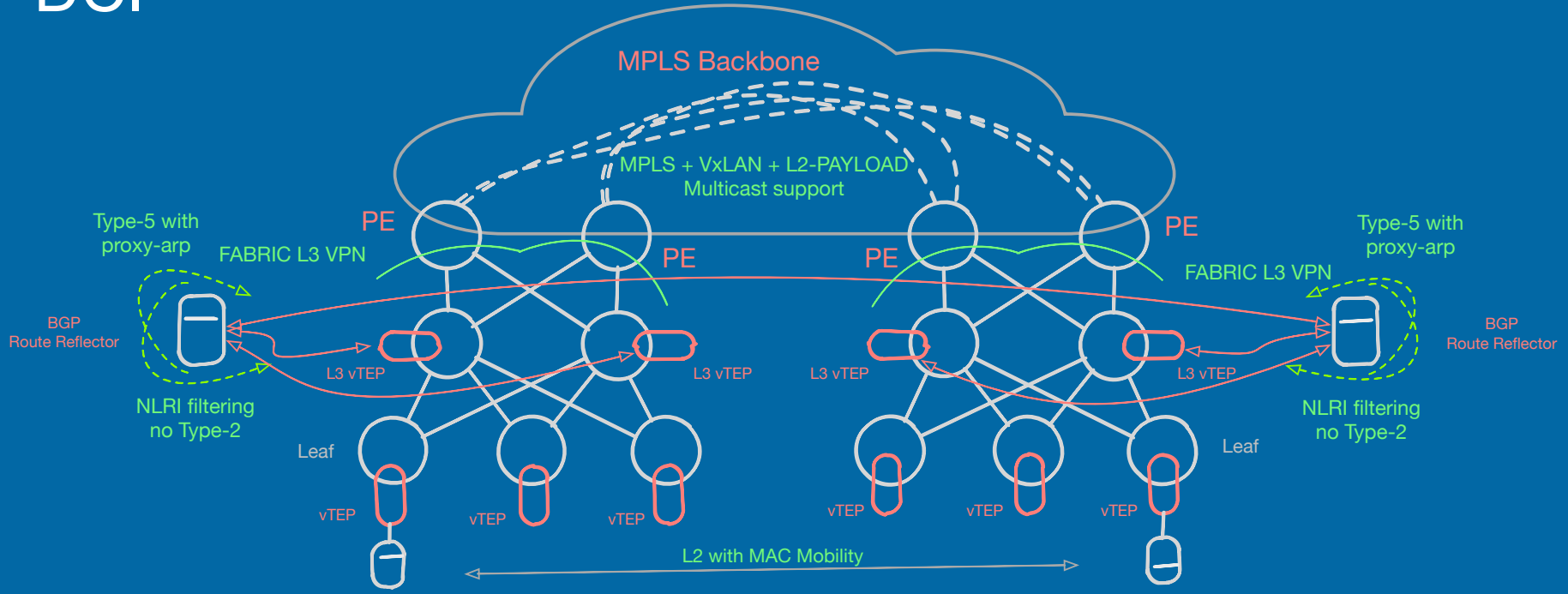


vTEP into PE permit redistribution of Type-5 prefix into customer L3VPN

Mac Mobility permits a fast convergence trough different DC

host-route redistribution and anycast default gw permits optimal traffic flow

DCI

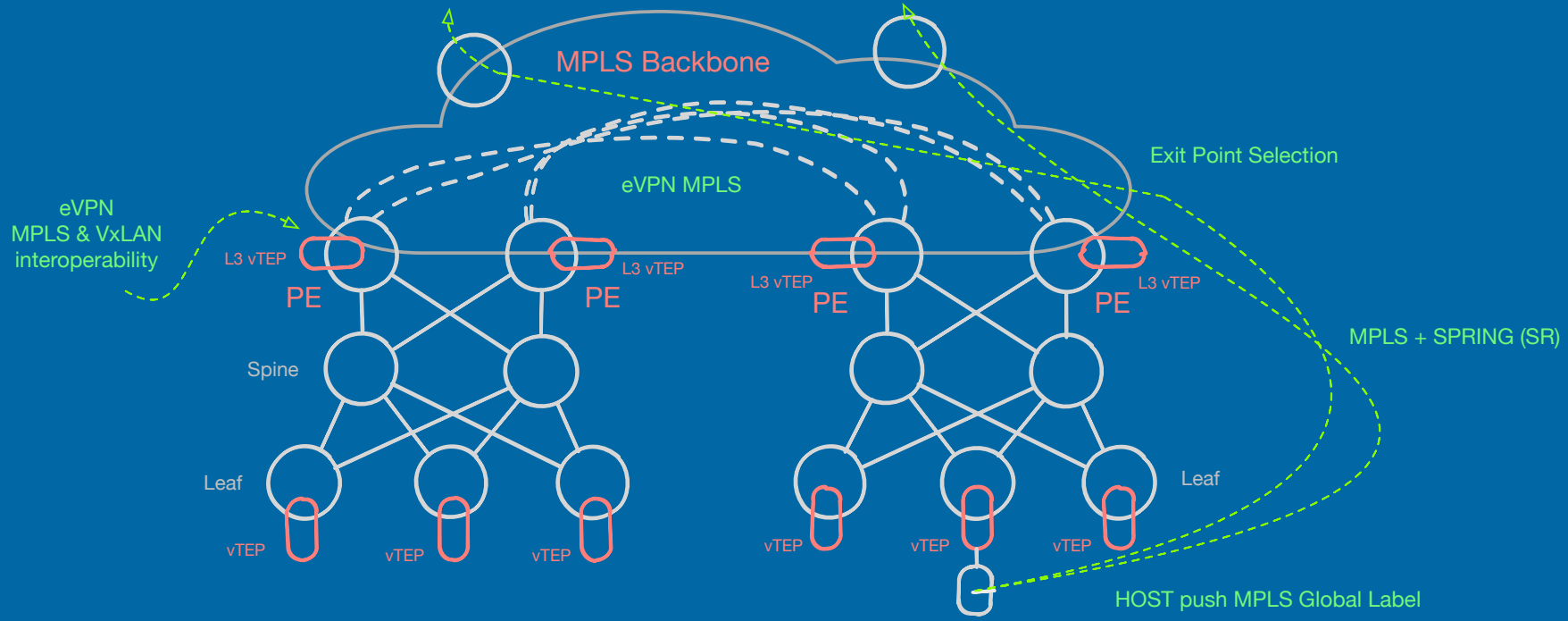


Reliable IP Connectivity (FABRIC L3VPN) for DCI best with multicast support (NG-mVPN)

eVPN NLRI control and filtering for optimal service localization

Type-5 with proxy-arp for broadcast suppression -> draft-ietf-bess-evpn-proxy-arp-nd-01

Evolution



Future developement:

- Integration between eVPN MPLS & VxLAN
- MPLS adoption in DC : Spring (SR) -> draft-lapukhov-segment-routing-large-dc-00

eVPN with overlay network

It’s IP and BGP: all I know is still applicable !

Simple troubleshooting

Integrate L2 and L3

Open Standard well documented

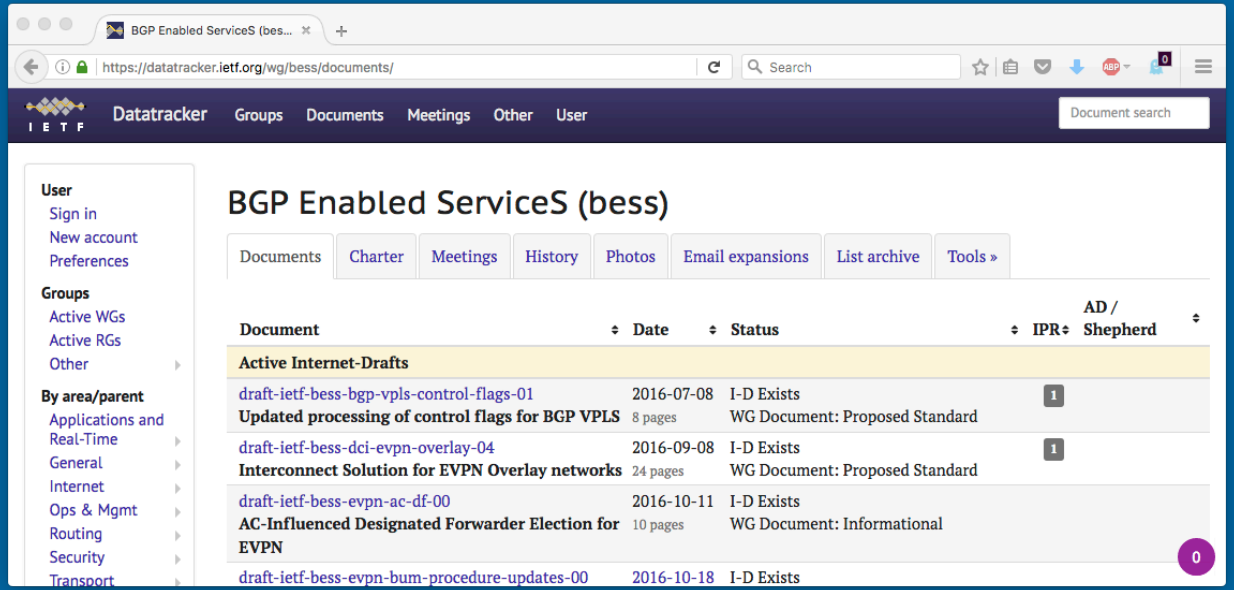
There are already open source implementations

Best option for DCI

Increasingly adoption by many vendors

eVPN as interconnection standard with legacy and proprietary Overlay Solutions

Sources



IETF BESS WG -> http://datatracker.ietf.org/wg/bess/documents

Ivan Pepelnjack blog -> http://blog.ipspace.net

Tiziano Tofoni blog -> http://blog.reissromoli.com

Thank You, Questions ?