

BESS work on control planes for DC overlay networks

A short overview

Jorge Rabadan

IETF99, July 2017
Prague

Agenda

EVPN in a nutshell

BESS work on EVPN for NV03 networks

EVPN in the industry today

Future work for NV03

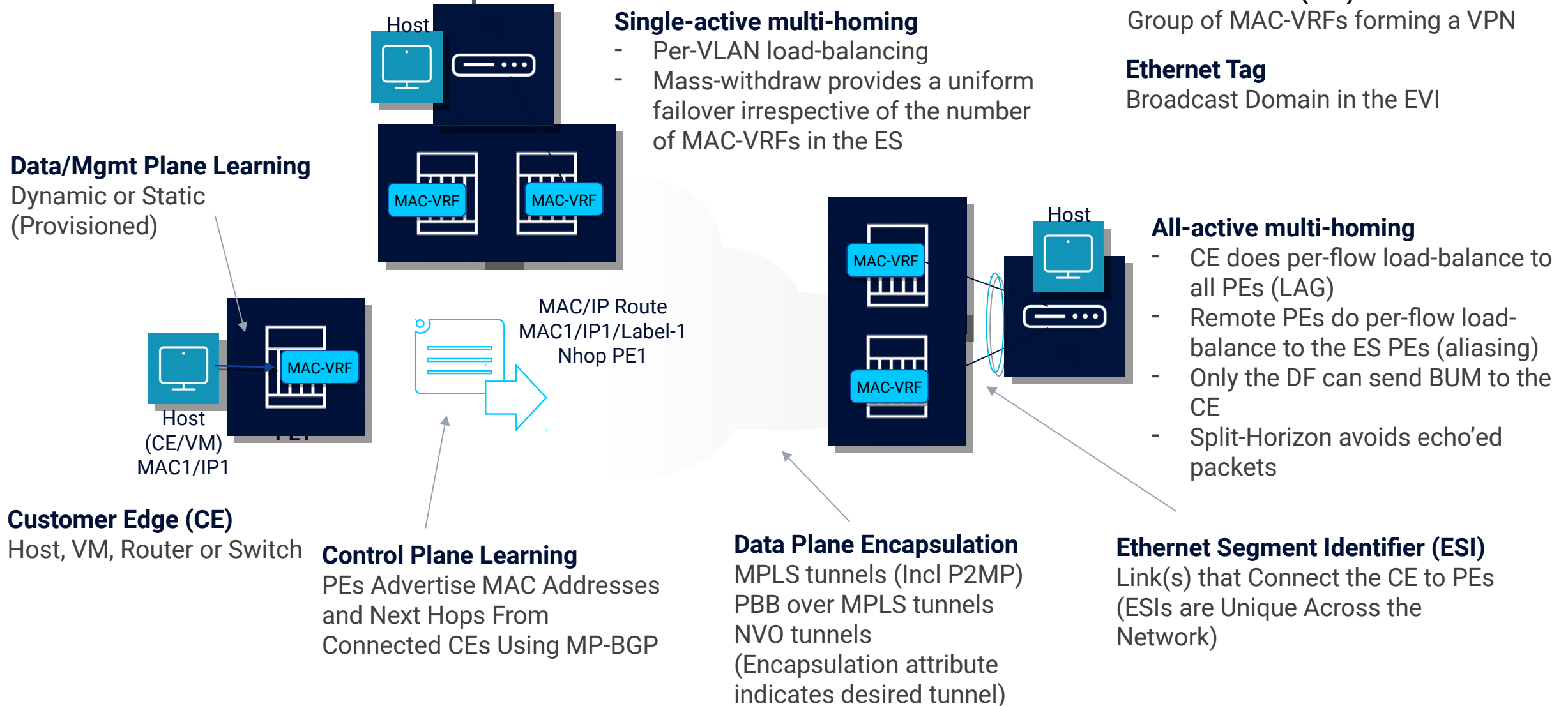
EVPN in a nutshell

A unified VPN control plane

- What is RFC7432 EVPN
 - IETF technology that allows multipoint layer-2 VPNs to be operated as RFC4364 IP-VPNs, where MACs and information to setup flooding trees are distributed by MP-BGP
- What were RFC7432's main objectives
 - Replace the old flood-and-learn behavior for BGP-based MAC learning (more control, mac-duplication, mac-protection, mac-mobility)
 - Efficient multi-destination delivery
 - All-active multi-homing
- How has EVPN evolved
 - EVPN is now a unified control plane protocol for E-LAN, E-Line, E-Tree, Layer-3 and Cloud services.
 - Transport agnostic
 - Advanced features

Ethernet Virtual Private Networks

RFC7432 main concepts



EVPN work in IETF

EVPN Application/Service Standard document

E-LAN

RFC7432 (EVPN)
RFC7623 (PBB-EVPN)

E-Line

draft-ietf-bess-evpn-vpws

E-Tree

draft-ietf-bess-evpn-etree

L3 VPN (Inter-subnet-forwarding)

draft-ietf-bess-evpn-inter-subnet-forwarding
draft-ietf-bess-evpn-prefix-advertisement

EVPN for DC

draft-ietf-bess-evpn-overlay
draft-ietf-bess-evpn-optimized-ir

EVPN for DCI

draft-ietf-bess-dci-evpn-overlay

Other applications or
enhancements:

- Multi-homing improvements
- Proxy-ARP/ND and security
- BUM optimizations
- IP Multicast optimizations L2/L3
- EVPN <-> IPVPN integration
- EVPN <-> VPLS integration

draft-ietf-bess-evpn-vpls-seamless-integ
draft-ietf-bess-evpn-df-election
draft-ietf-bess-evpn-ac-df
draft-ietf-bess-evpn-pref-df
draft-ietf-bess-evpn-proxy-arp-nd
draft-ietf-bess-evpn-bum-procedure-updates
draft-ietf-bess-evpn-igmp-mld-proxy



**Current DC Overlay
Network Deployments**

**Plus around 20 individual
EVPN-related drafts !!**

Why is EVPN used in DC Overlay Networks?

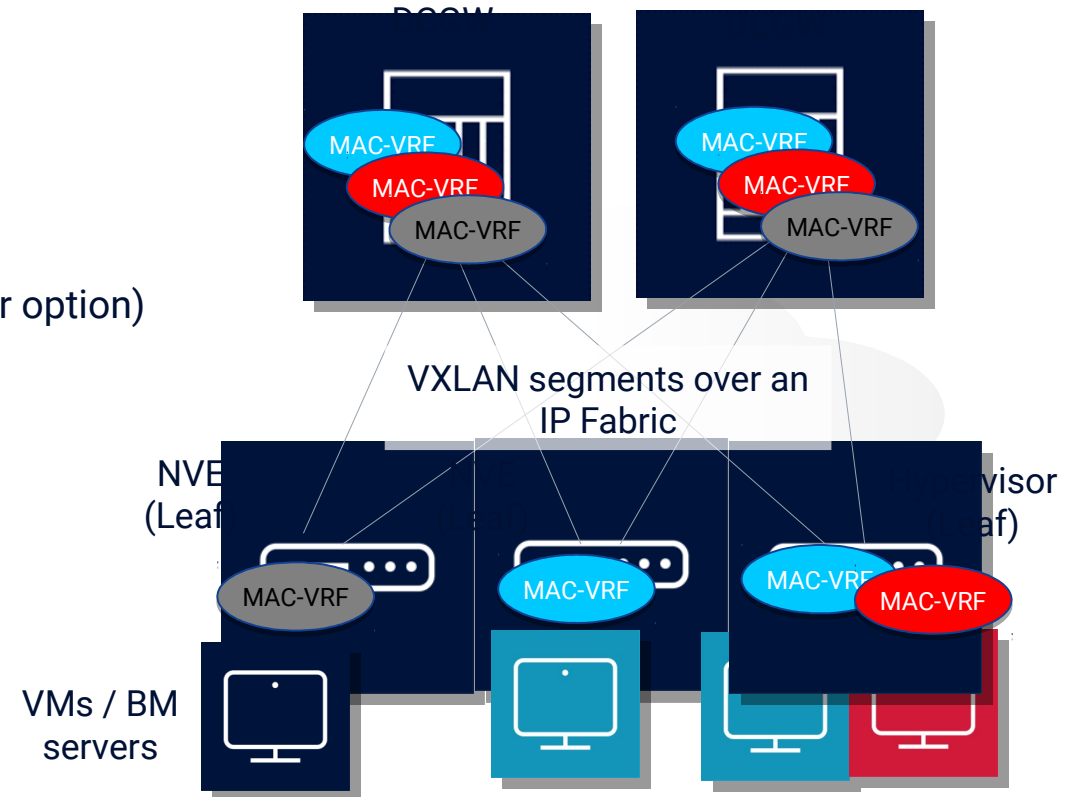
draft-ietf-bess-evpn-overlay

Modern DCs are based on:

- CLOS architecture and IP Fabrics
 - No loops, no flooding, fast convergence
 - ECMP
- Multi-tenancy with intra (L2) and inter-subnet (L3) connectivity
- IP Overlay tunnels are therefore needed (VXLAN is the most popular option)

Why do I need a control plane?

- Auto-discovery of the remote VTEPs
- Distribution of MAC/IP information in order to reduce/suppress flooding
- Other advanced options



EVPN for NVO tunnels

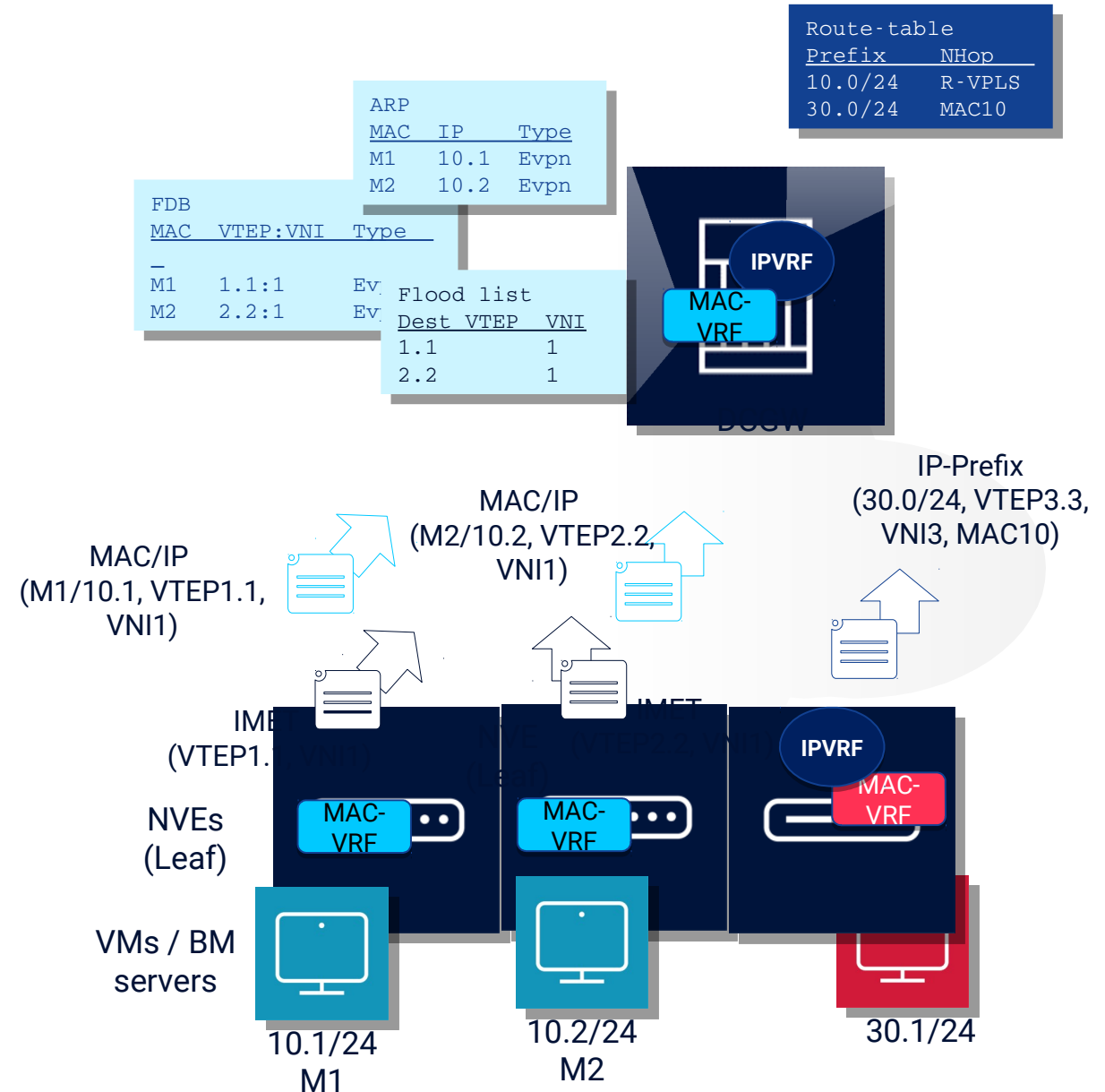
Endless possibilities

EVPN provides the basic Control Plane needs...

- Auto-discovery of remote VTEPs through Inclusive Multicast Routes (IMET)
- Distribution of MAC/IP information in order to reduce/suppress flooding through MAC/IP routes

But also advanced options

- All-active multi-homing
- MAC/IP mobility
- MAC protection, duplication detection, loop protection
- Proxy-ARP/ND, IGMP and PIM proxy, to reduce/suppress the flooding in the DC
- Assisted-Replication
- Distribution of IP-Prefix/host routes and inter-subnet-forwarding so that not all the subnets must be defined in all the NVEs
- Etc



EVPN for VXLAN tunnels is widely deployed today in Data Centers

Multiple vendor implementations

- Hardware based NVEs
- Software based NVEs

Interoperability publicly demonstrated

Interoperability Showcase 2017
White Paper

MPLS + SDN + NFV World Congress 2017 Multi-Vendor Interoperability Test

We configured the interface-full mode on the EVPN-VXLAN PE devices. We verified that the VXLAN virtual network identifier (VNI) directly mapped to the EVPN EVI. We confirmed that the RT2 Prefix advertisement route) carried the correct IP length, the Route Target Gateway is set to 0 and was learned on the peer leaf device's IRB interface via CLI. Additionally, a RT2 was used in interface-full mode. It carried MAC address length and MAC address. The IP length was set to 0. Following, we sent IPv4 test traffic from all IPv4 subnets to any other IPv4 subnets, and expected to receive traffic on all IPv4 subnets without any packet loss. Five vendors participated in the symmetric interface-less EVPN setup with the following DUTs (see Figure 2): Arista 7050SX, Cisco Nexus 9300 Series, Huawei NE40E-X2-MBA, Huawei NE40E-XBA, Ixia IxNetwork and Spirent TestCenter. Three vendors participated in the symmetric interface-full EVPN test with the following DUTs: Cisco Nexus 9300 Series, Ixia IxNetwork, and Nokia 7750SR. Spirent TestCenter and Ixia IxNetwork acted as the traffic generator. Cisco Nexus 7702 participated as the router server.

MAC Mobility. MAC mobility is a mechanism that is used to detect host moving from one Ethernet Segment to another. It is achieved by adding a sequence number into the MAC/IP advertisement route (RT2). When a host moves once, the sequence number is increased by one. The new PE sends an Ethernet MAC/IP advertisement route [EVPN RT2] to inform the other PEs to withdraw the route that has a smaller number. The host was simulated either by Spirent TestCenter or Ixia IxNetwork. We moved the host by setting up the same MAC address of the previous host on a different device and removed the MAC address from the previous device. We first verified that the sequence number was 0 (or not set) before the MAC [host] was moved. When the MAC [host] moved once, we observed that the sequence number was exactly increased by one as expected. We also measured the out of service time during the host movement. We used a constant rate of packets (1,000 packets/s) and performed the host movement. Then we measured the out of service time which we calculated based on the last number of packets. The out of service times from different vendors were between 22 ms to 130 ms. The following vendors participated in the test: Arista 7050SX, Huawei NE40E-XBA, Ixia IxNetwork and Nokia 7750SR. Nokia 7750SR joined as BGP route reflector in the EVPN overlay network. Cisco Nexus 7702 joined as route server. Spirent TestCenter and Ixia IxNetwork acted as the traffic generator.

Figure 3: MAC Mobility and Proxy ARP/ND

Initially, one vendor failed to learn the first update when the host was moved from its own device to a remote peer. The DUT replied with a sequence number that was higher than the received one. The DUT got a RT2 route sent by the remote site to notify the changes, indicating that the learning failed. This error prompted the process to be repeated one more time. Afterwards, the DUT successfully learned the updates from the second RT2 route sent by the remote peer.

Single Homed EVPN with Proxy ARP and Proxy ND. The goal was to use ARP Proxy/Proxy ND (Neighbor Discovery) to learn MAC addresses across the EVPN instance. This is achieved by using RT2 to exchange the MAC address and to store it locally. When one PE receives an ARP request/Neighbor solicitation, it will search locally first. Only if no proper result is found, the ARP request/Neighbor solicitation floods to the other remote PE.

The first step was to discover the prefix by having the PE connected at first with a host that was emulated by Ixia IxNetwork or Spirent TestCenter.

EVPN new/future work for NV03

- Support new NV03 data encapsulations
 - First attempt for Geneve - draft-boutros-bess-evpn-geneve-00
- Tunnel options/extensions negotiation
- Future extensions

Thank you