EVPN
Unifying control plane

Sudarshan Murali, Product Manager, SP Routing

Jiri Chaloupka, Technical Marketing Engineer, SP Routing

Agenda

- Why EVPN
- EVPN Introduction & Value Proposition
- EVPN Technical Overview
- EVPN Components
- EVPN Life of a Packet
- EVPN Demo
- Summary

Cisco Open Network Architecture

Cisco Innovations – Simplify, Automate, Virtualize

Efficiency | **Agility** | **Innovation**

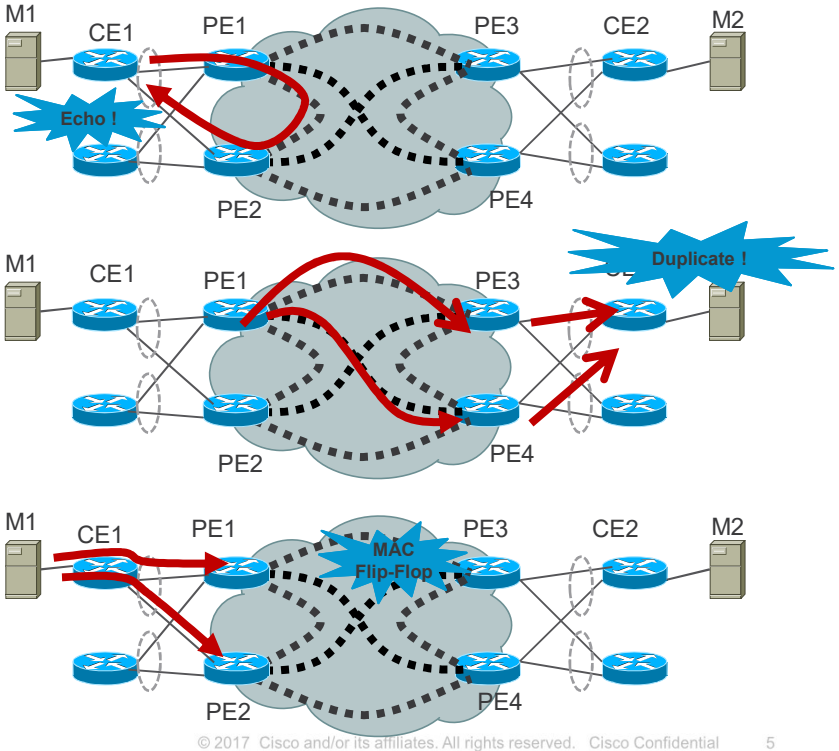
Why is EVPN needed?

SIMPLIFY

- Legacy L2 technologies (VPLS, PBB) still rely on flooding and learning to build Layer 2 forwarding database
- Network Operators have emerging needs in their network:
  - Data center interconnect operation (DCI)
  - Cloud and Services virtualization
  - Reduce protocol stack and Simplify Network
  - Integrated of Layer 2 and Layer 3 Services over the same VPN

Solving VPLS challenges for per-flow Redundancy

- Existing VPLS solutions do not offer an All-Active per-flow redundancy
- Looping of Traffic Flooded from PE
- Duplicate Frames from Floods from the Core
- MAC Flip-Flopping over Pseudowire
  - E.g. Port-Channel Load-Balancing does not produce a consistent hash-value for a frame with the same source MAC (e.g. non MAC based
    Hash-Schemes)



Data Center Interconnect requirements

not fully addressed by current L2VPN technologies

- Per-Flow Redundancy and Load Balancing
- Simplified Provisioning and Operation
- Optimal Forwarding
- Fast Convergence
- MAC Address Scalability

EVPN with a choice of data plane encapsulation (MPLS, VxLAN, PBB) is the designed to address these requirements.

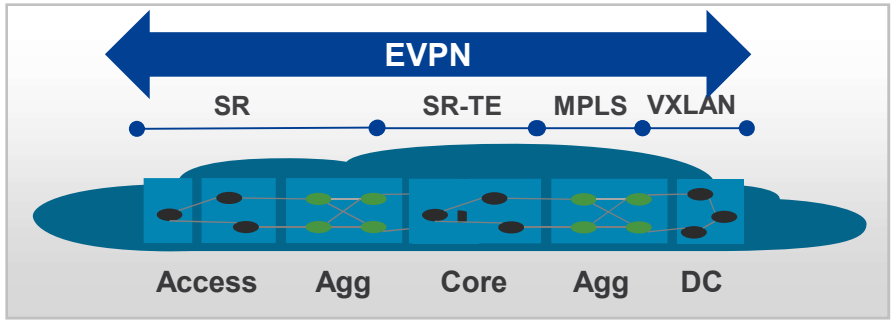
EVPN
Next generation network services

Single service for any application

ELINE	ELAN	ETREE	L3 VPN	DC Fabric	DCI
VPWS	VPLS	P2MP VPLS	RFC 2547	VXLAN	VPLS / L3 VPN

EVPN

E2E control and automation across domains



Optimized CapEx:
- Open Standards & Multi-vendor
- Active-Active multi-homing
- Enhanced load balancing

Reduced OpEx:
- Integrated L2 & L3 service, any application: faster time to market, certification
- E2E control and automation

Increased Customer Value
- Inter-domain SLA, faster convergence
- Better stability: no flood
- Granular policy control

EVPN: Unifying control plane

Control Plane

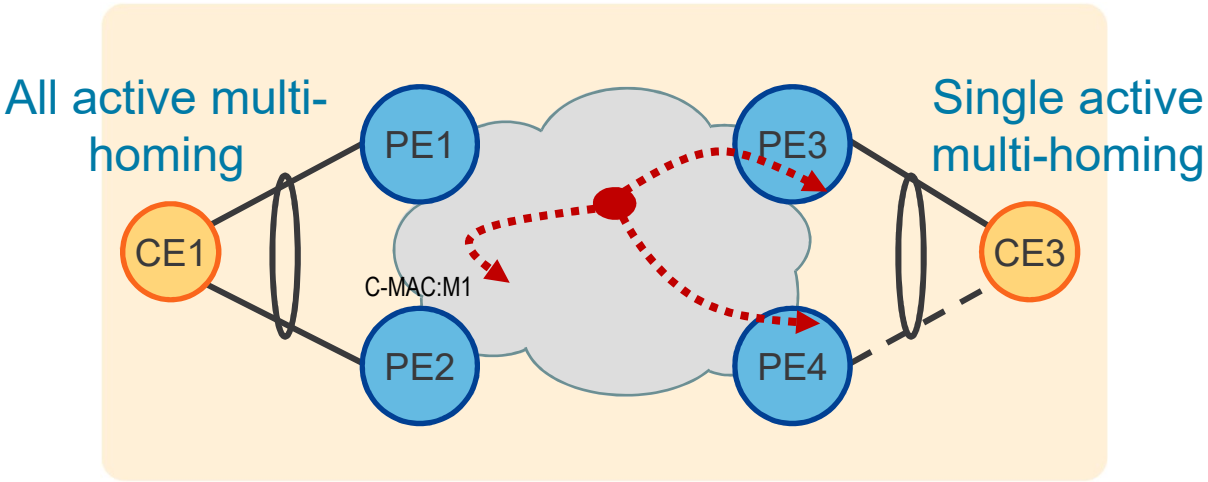
EVPN MP-BGP

Data Plane

MPLS RFC-7432	NVO evpn-overlay	PBB RFC-7623	VPWS evpn-vpws
• EVPN over MPLS (E-LAN service) • All-active Multi-homing • SR/SR-TE as underlay • EVPN-VPWS (E-Line service)	• EVPN over NVO Tunnels (VxLAN, NVGRE) • Data Center Fabric Encapsulation • Integrated L2/L3 services DCI • Overlays over IP networks	• PBB Encapsulation for CMAC scalability • MPLS as underlay • All-active Multi-homing • BMACs advertisement	• P2P services • MPLS as underlay • All-active Multi-homing • Flexible xconnect services

Ethernet VPN (EVPN) - Overview

MAC Routing: Control plane (BGP)
advertise the learnt MACs from CE



Data Plane: IP or MPLS

Network Efficiency

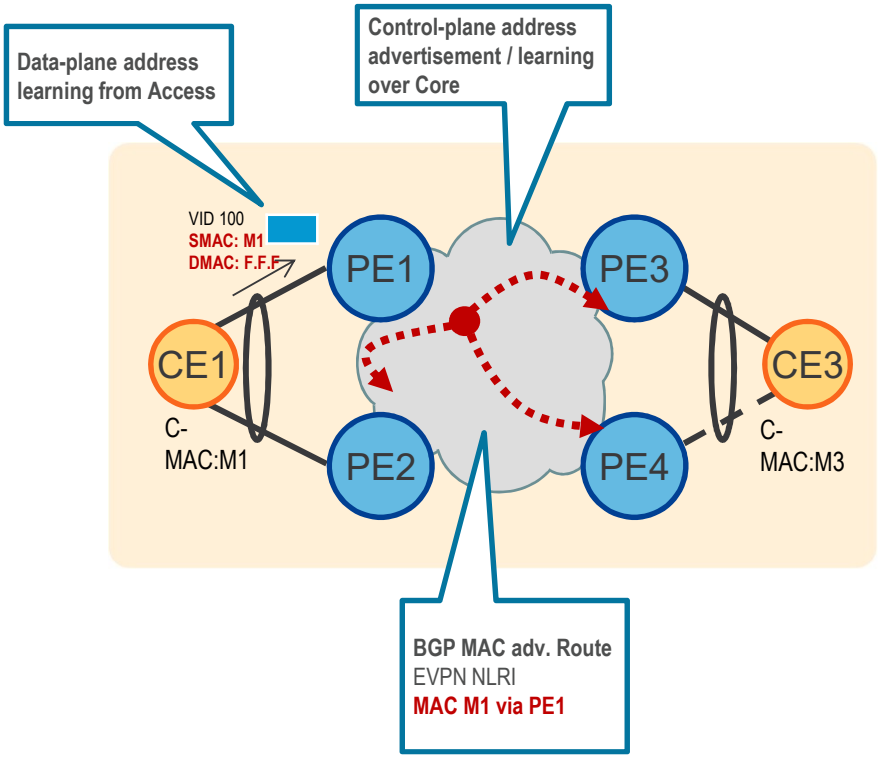
Common L2/L3 VPN Operational Mode

Consolidated VPN service with x-EVPN

Ethernet VPN

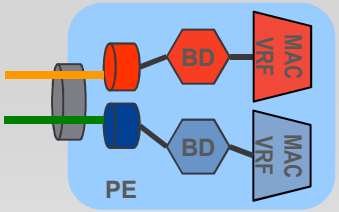
Highlights

- Next generation solution for Ethernet multipoint (E-LAN) services
- PEs run Multi-Protocol BGP to advertise & learn Customer MAC addresses (C-MACs) over Core
  - Same operational principles of L3VPN
- Learning on PE Access Circuits via data-plane transparent learning
- No pseudowire full-mesh required
  - Unicast: use MP2P tunnels
  - Multicast: use ingress replication over MP2P tunnels or use LSM
- Standardized at IETF – RFC 7432



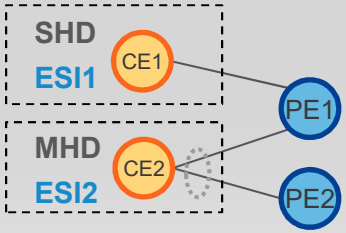
EVPN - Concept

EVPN Instance (EVI)



- **EVI spans all PEs participating in an EVPN**
- **MAC-VRF: A VRF table for MACs on a PE**
- **Encompass one or more bridge-domains, depending on service interface type**
  - Port-based
  - VLAN-based (shown above)
  - VLAN-bundling
  - VLAN aware bundling (NEW)

Ethernet Segment



- **Represents a 'site' connected to one or more PEs**
- **Uniquely identified by a 10-byte global Ethernet Segment Identifier (ESI)**
- **Could be a single device or an entire network**
  - Single-Homed Device (SHD)
  - Multi-Homed Device (MHD)
  - Single-Homed Network (SHN)
  - Multi-Homed Network (MHN)

BGP Routes

Route Types
[1] Ethernet Auto-Discovery (AD) Route
[2] MAC Advertisement **Route**
[3] Inclusive Multicast Route
[4] Ethernet Segment Route
(5) IP Prefix Advertisement Route

- **EVPN and PBB-EVPN define a single new BGP NLRI used to carry all EVPN routes**
- **NLRI has a new SAFI (70)**
- **Routes serve control plane purposes, including:**
  - MAC / IP address reachability
  - MAC mass withdrawal
  - Split-Horizon label adv.
  - Aliasing
  - Multicast endpoint discovery
  - Redundancy group discovery
  - Designated forwarder election

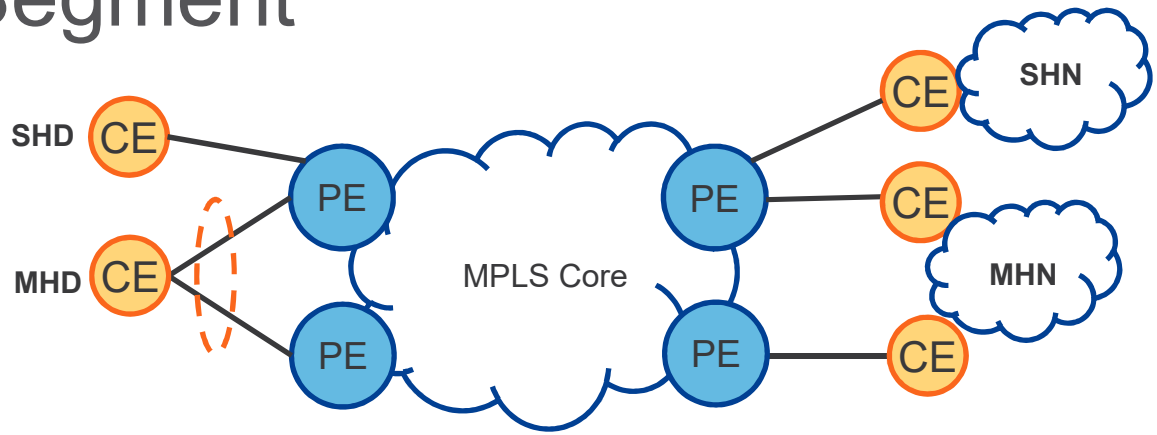
BGP Route Attributes

Extended Communities
ESI MPLS Label
ES-Import
MAC Mobility
Default Gateway
Router's MAC

- **New BGP extended communities defined**
- **Expand information carried in BGP routes, including:**
  - MAC address moves
  - C-MAC flush notification
  - Redundancy mode
  - MAC / IP bindings of a GW
  - Split-horizon label encoding

Ethernet Segment

Definition



- Ethernet Segment is a ‘site’ connected to one or more PEs.
- Ethernet Segment could be a single **device** (i.e. CE) or an entire **network**.
  - Single-Homed Device (SHD)
  - Multi-Homed Device (MHD)
  - Single-Homed Network (SHN)
  - Multi-Homed Network (MHN)
- Uniquely identified by global Ethernet Segment Identifier (**ESI**).

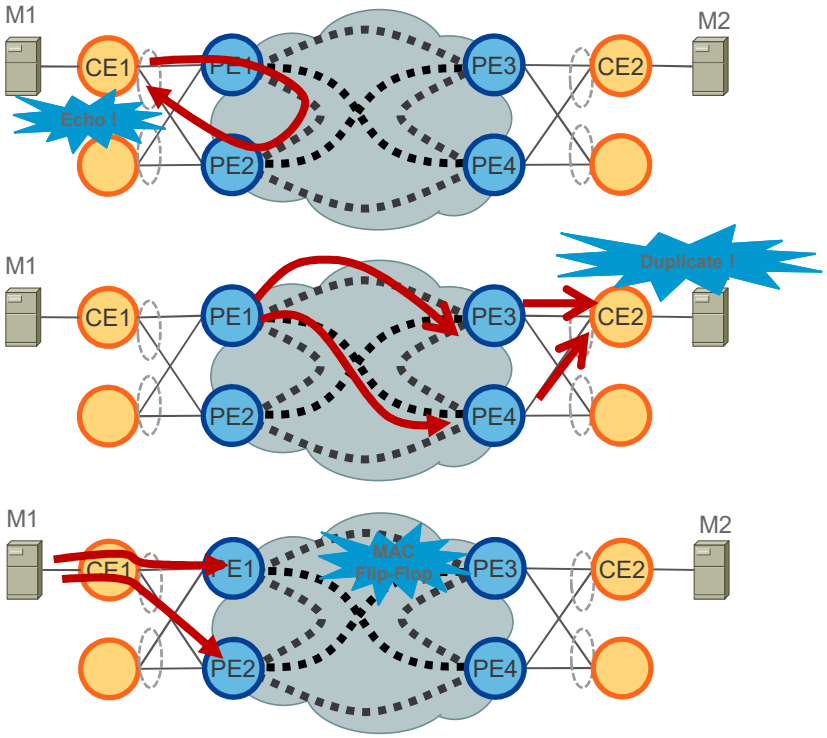
EVPN BGP route type

Route type	Usage	EVPN	PBB-EVPN	EVPN VPWS
0x1 Ethernet Auto-Discovery (A-D) Route	• MAC Mass-Withdraw • Aliasing (load balancing) • Split-Horizon "Tagged with ESI Label Extended Community"	✓	**NOT used**	✓
0x2 MAC Advertisement Route	• Advertises MAC addresses /IP for VM reachability • Provides MAC/IP address bindings for ARP broadcast suppression "Tagged with MAC Mobility Extended Community"	✓	✓	**NOT used**
0x3 Inclusive Multicast Route	• Indicates interest of BUM traffic for attached L2 segments • Multicast tunnels used to BUM frame "Tagged with PMSI tunnel attribute" (P tunnel type & ID) – RFC6514	✓	✓	**NOT used**
0x4 Ethernet Segment Route	• Auto discovery of Multi-homed Ethernet Segments, i.e. redundancy group discovery • Designated Forwarder (DF) Election "Tagged with ES-Import Extended Community"	✓	✓	✓
0x5 IP Prefix Route	• Advertises IP prefix for a subnet for L3 NLRI only inter-subnet routing via EVPN address family	✓	✓	**NOT used**

Next-Generation Solutions for L2VPN

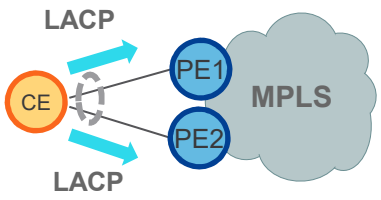
Solving VPLS challenges for per-flow Redundancy

- Existing VPLS solutions do not offer an All-Active per-flow redundancy
- Looping of Traffic Flooded from PE
- Duplicate Frames from Floods from the Core
- MAC Flip-Flopping over Pseudowire
  - E.g. Port-Channel Load-Balancing does not produce a consistent hash-value for a frame with the same source MAC (e.g. non MAC based
    Hash-Schemes)



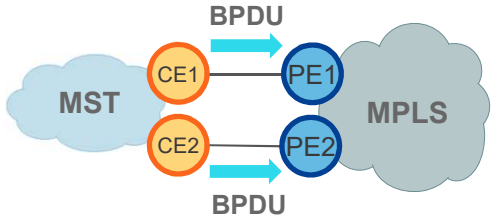
Ethernet Segment

ESI Auto-Sensing



MHD with Multi-chassis LAG

- ESI is auto-discovered via LACP.
- ESI is encoded using the CE's LACP parameters:



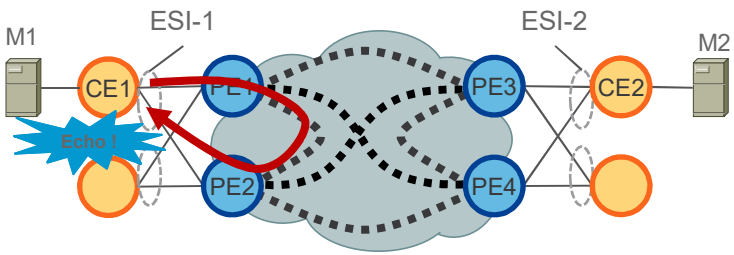
MHN with MST

- ESI is auto-discovered via MST BPDU snooping.
- ESI is encoded using the IST's root parameters:

Split Horizon

For Ethernet Segments – E-VPN

Challenge:
How to prevent flooded traffic from echoing back to a multi-homed Ethernet Segment?

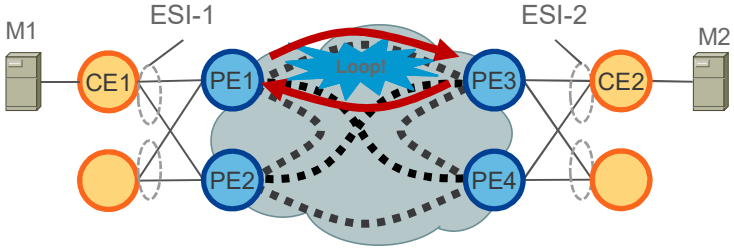


- PE advertises in BGP a split-horizon label (ESI MPLS Label) associated with each multi-homed Ethernet Segment.
- Split-horizon label is only used for multi-destination frames (Unknown Unicast, Multicast & Broadcast).
- When an ingress PE floods multi-destination traffic, it encodes Split-Horizon label identifying source Ethernet Segment in packet
- Egress PEs use this label to perform selective split-horizon filtering over attachment circuit

Split Horizon

For Core Tunnels

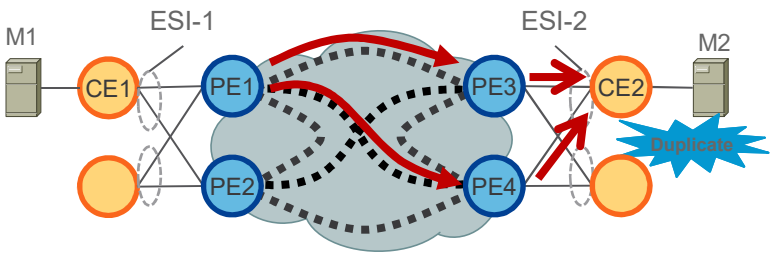
Challenge:
How to prevent flooded traffic from looping back over the core?



- Traffic received from an MPLS tunnel over core is never forwarded back to MPLS core
- Similar to VPLS split-horizon filtering rule

Designated Forwarder (DF)

DF Election



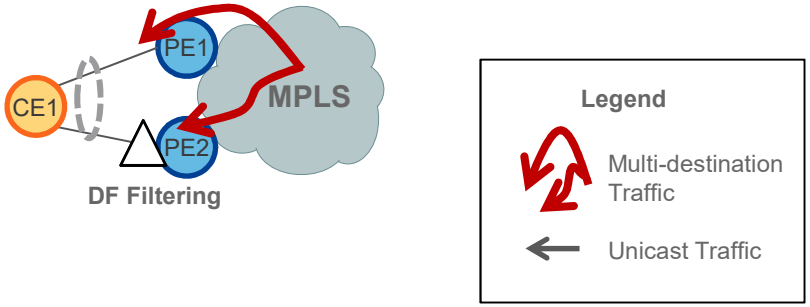
Challenge:
How to prevent duplicate copies of flooded traffic from being delivered to a multi-homed Ethernet Segment?

- PEs connected to a multi-homed Ethernet Segment discover each other via BGP
- PEs then elect among themselves a Designated Forwarder responsible for forwarding flooded multi-destination frames to multi-homed Segment
- DF Election granularity can be:
  - Per Ethernet Segment (Single PE is the DF)
  - Per EVI (E-VPN) on Ethernet Segment (Multiple DFs for load-balancing)

Designated Forwarder (DF)

DF Filtering

MHD All-Active with Per-Flow Load Balancing

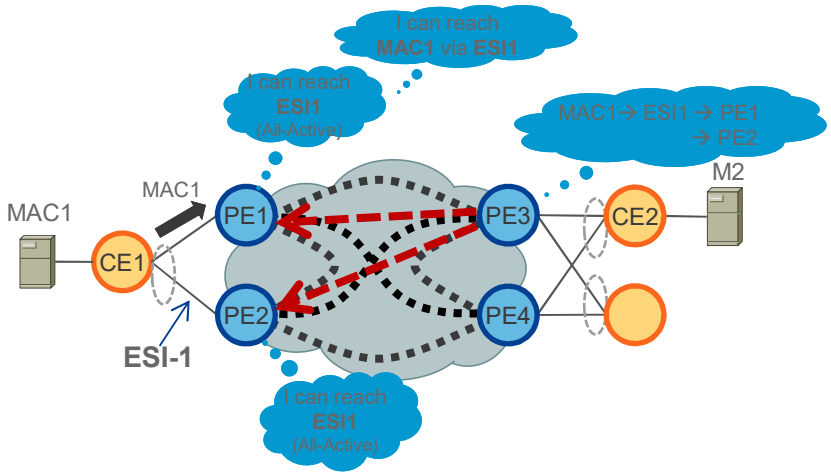


Filtering Direction:	Core to Segment
Filtered Traffic:	Flooded multi-destination

Aliasing

E-VPN

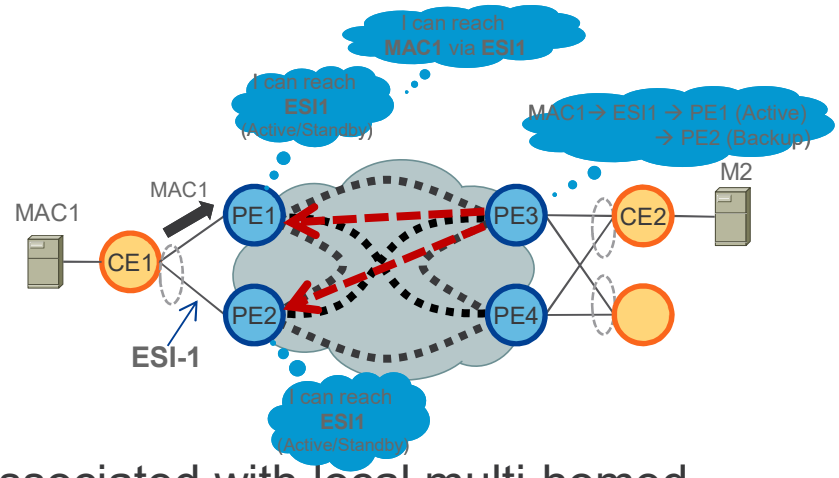
Challenge:
How to load-balance traffic towards a multi-homed device across multiple PEs when MAC addresses are learnt by only a single PE?



- PEs advertise in BGP the ESIs of local multi-homed Ethernet Segments
  - All-Active Redundancy Mode indicated
- When PE learns MAC address on its AC, it advertises MAC in BGP along with ESI of Ethernet Segment from which MAC was learnt
- Remote PEs can load-balance traffic to a given MAC address across all PEs advertising same ESI

Backup Path

Challenge:
How to identify PEs that have a backup path to a multi-homed Ethernet Segment?



PEs advertise in BGP connectivity to ESIs associated with local multi-homed Ethernet Segments

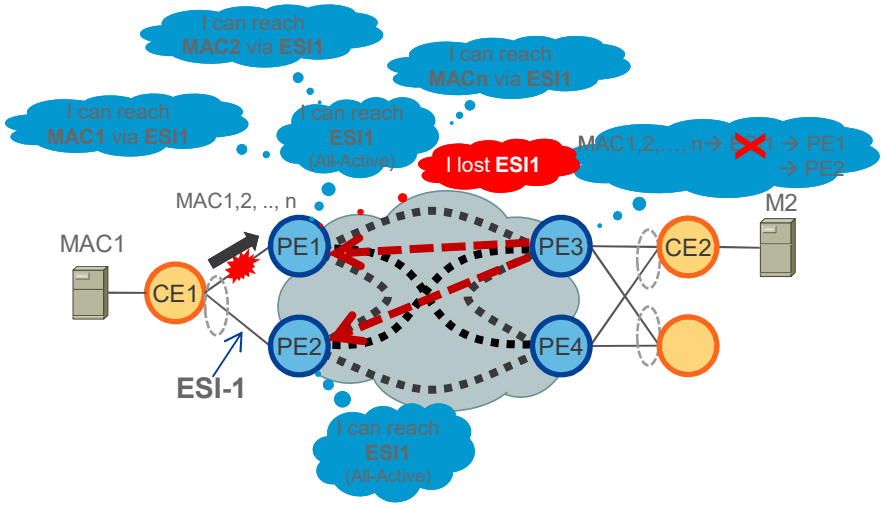
- Active/Standby Redundancy Mode is indicated

When PE learns a MAC address on its AC, it advertises MAC in BGP along with ESI of Ethernet Segment from which MAC was learnt

Remote PEs will install:

- **active path** to PE that advertised **both MAC Address & ESI**
- **backup path** to PE that advertised **ESI only**

MAC Mass-Withdraw



Challenge:
How to inform remote PEs of a failure affecting many MAC addresses quickly while the control-plane re-converges?

PEs advertise two sets of information:

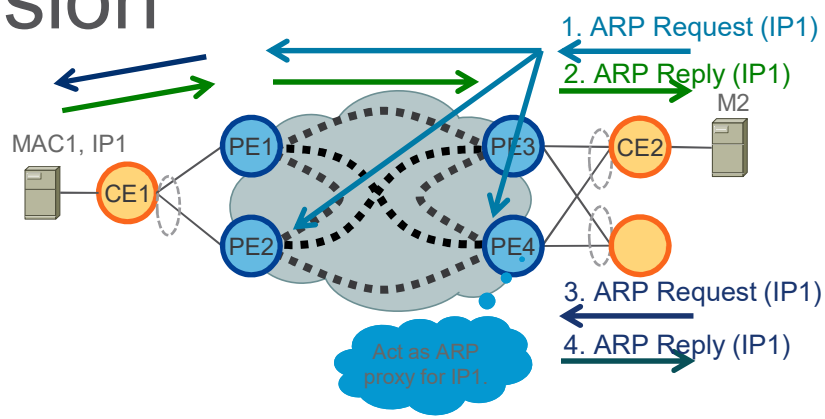
- MAC addresses along with ESI from address was learnt
- Connectivity to ESI(s)

If a PE detects a failure impacting an Ethernet Segment, it withdraws route for associated ESI

- Remote PEs remove failed PE from path-list for **all MAC addresses associated with an ESI**.
- This effectively is a MAC ‘mass-withdraw’ function

ARP Broadcast Suppression

Challenge:
How to reduce ARP broadcasts over the MPLS/IP network, especially in large scale virtualized server deployments?



Construct ARP caches on E-VPN PEs and synchronize them either via BGP or data-plane snooping

PEs act as ARP proxies for locally attached hosts, thereby preventing repeated ARP broadcast over the MPLS/IP network

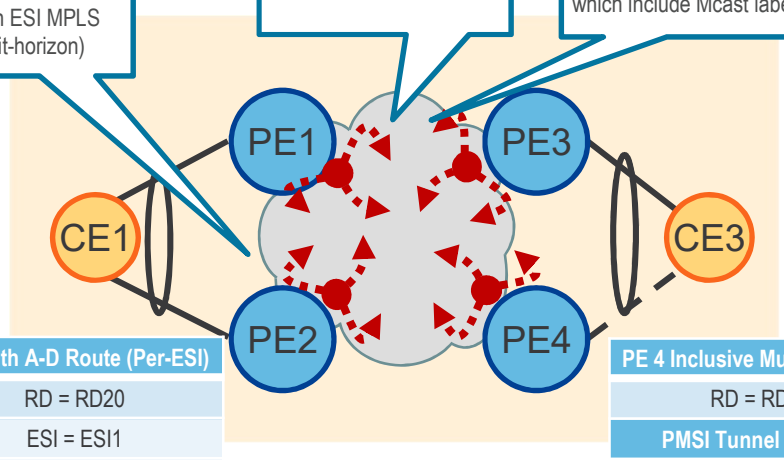
Life of a Packet

Ingress Replication – Multi-destination Traffic Forwarding

During start-up sequence, PE2 sent Per-ESI Ethernet AD route with ESI MPLS label (split-horizon)

During start-up sequence, DF election between PE1-PE2 and PE3-PE4

During start-up sequence, PE1, PE2, PE3, PE4 sent Inclusive Multicast route which include Mcast label



PE 2 Eth A-D Route (Per-ESI)
RD = RD20
ESI = ESI1
ESI MPLS Label ext. comm.
Redund. Flag = All-Active
Label = L5
RT ext. community
RT-a, RT-b, RT-c, RT-d

ESI MPLS Label – used by local PEs for split-horizon - downstream assigned (for ingress replication)

PE 4 Inclusive Multicast Route
RD = RD-4a
PMSI Tunnel Attribute
Tunnel Type = Ing. Repl.
Label = L4
ommunity
T-a

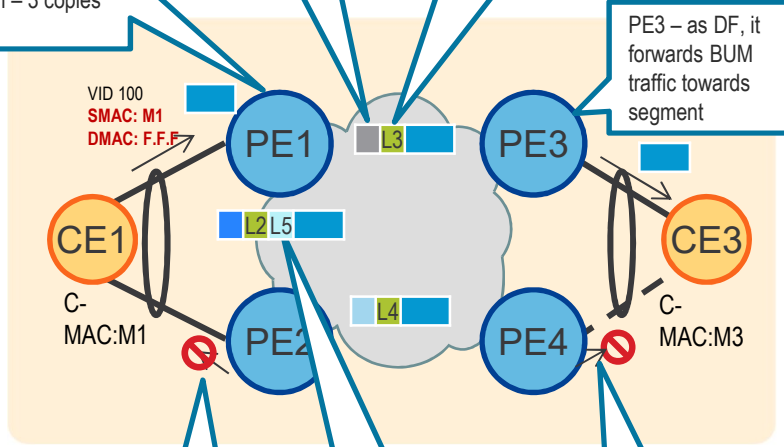
Mcast MPLS Label – used to transmit BUM traffic - downstream assigned (for ingress replication)

PE1 receives broadcast traffic from CE1. PE1 forwards it using ingress replication – 3 copies created

PSN MPLS label to reach PE3

Mcast MPLS Label assigned by PE3 for incoming BUM traffic on a given EVI

PE3 – as DF, it forwards BUM traffic towards segment



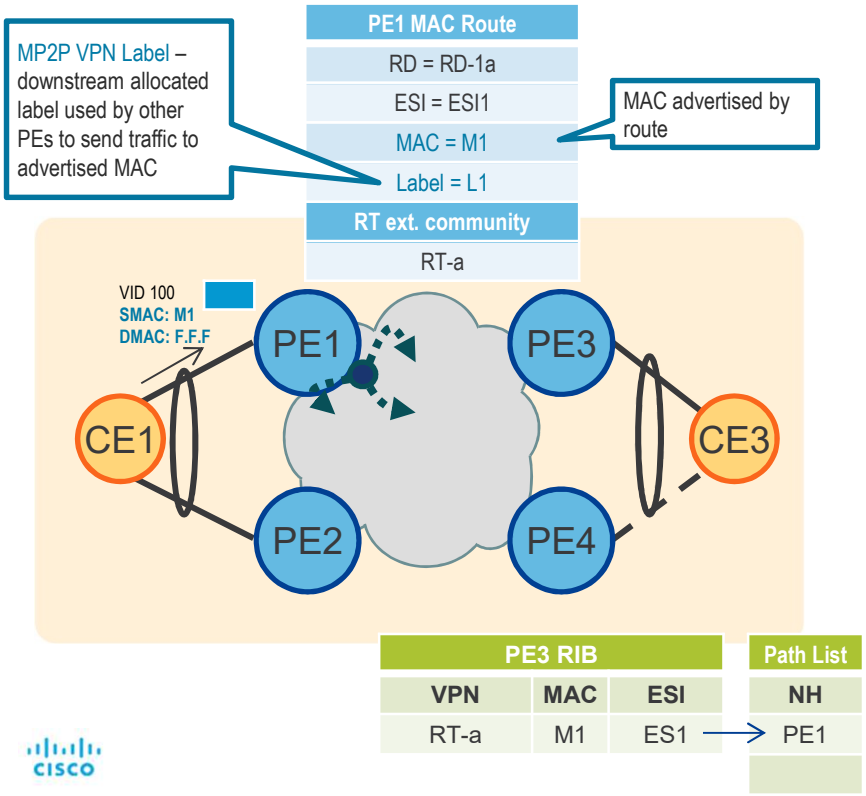
PE2 – drops BUM traffic originated on ES1

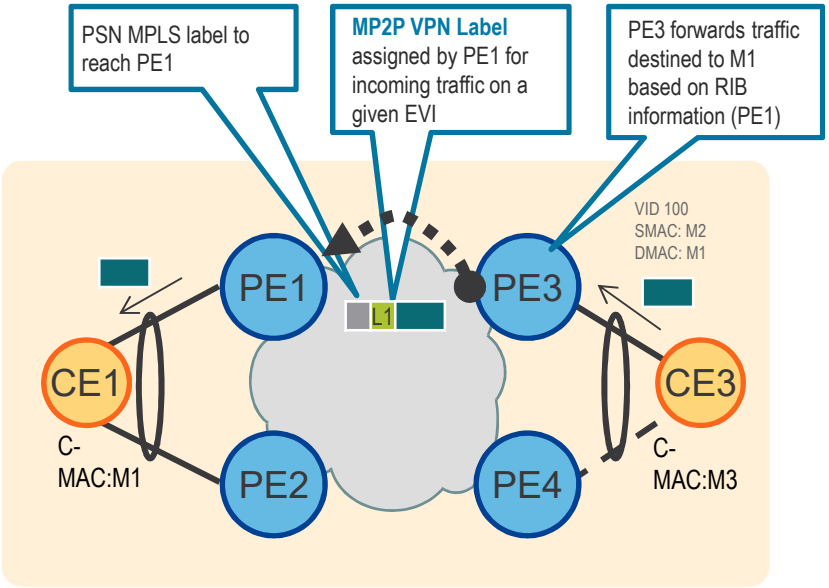
ESI (split-horizon) MPLS label allocated by PE2 for segment ES1

PE4 – non-DF for given EVI drops BUM traffic

Life of a Packet

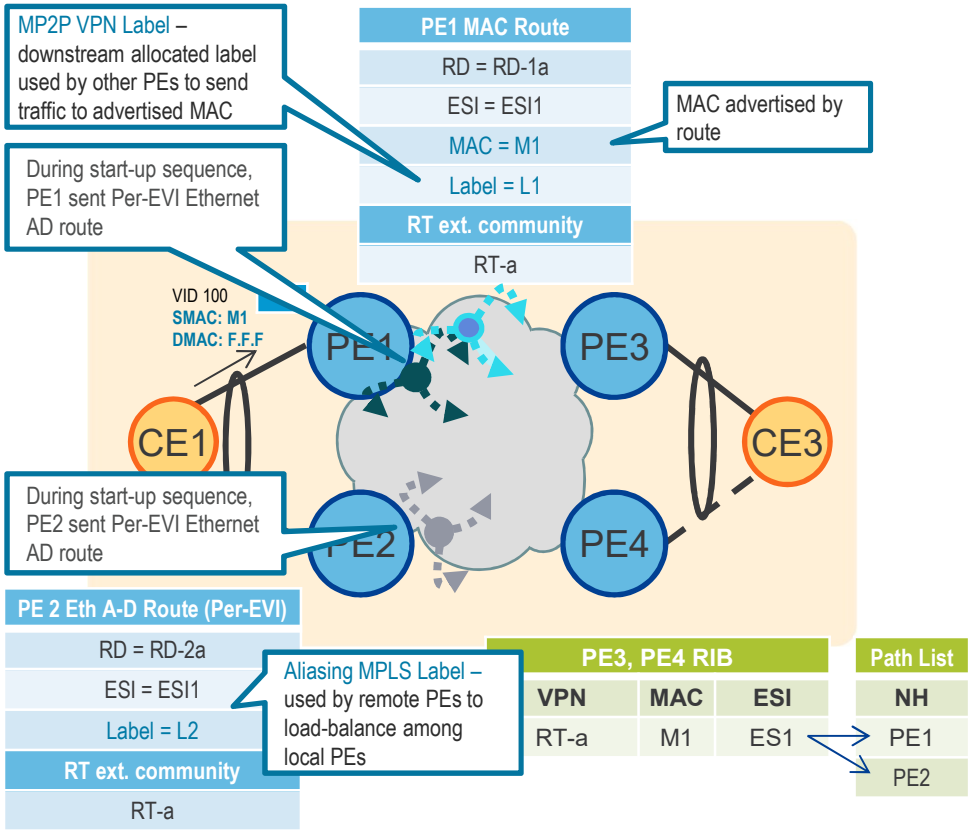
Unicast Traffic Forwarding

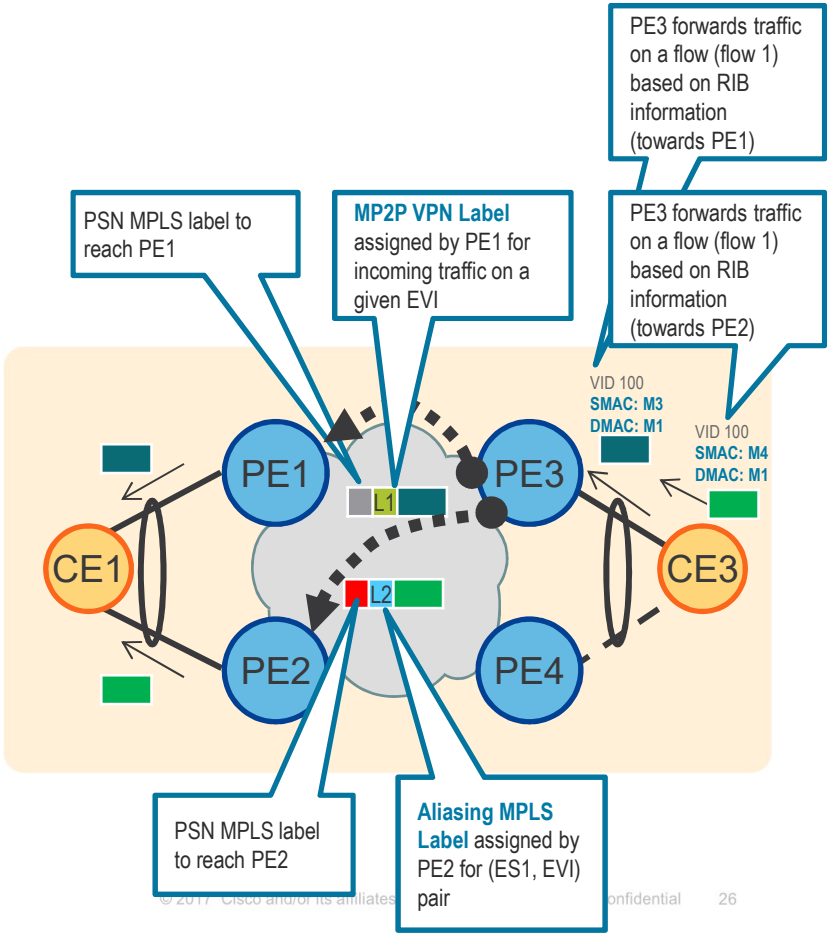




Life of a Packet

Unicast Forwarding and Aliasing





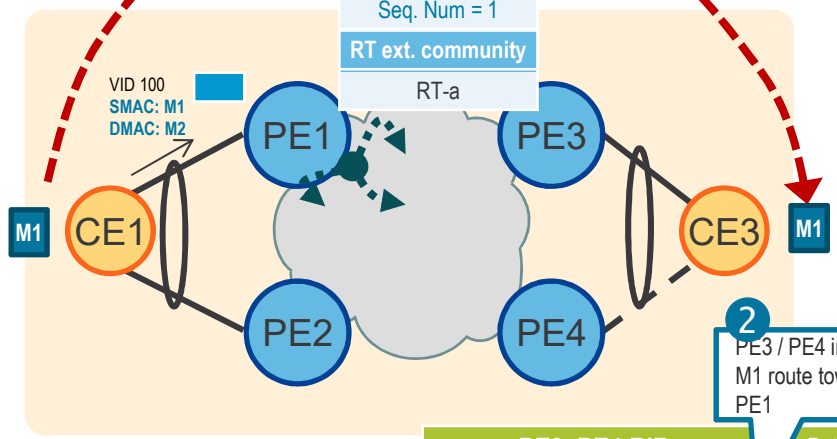
E-VPN Operational Scenarios

MAC Mobility

PE1 advertises MAC route for M1. Route may include MAC mobility community

PE1 MAC Route
RD = RD-1a
ESI = ESI1
MAC = M1
Label = L1
MAC Mobility ext. community
Seq. Num = 1
RT ext. community
RT-a

3 Host M1 moves from CE1 to CE3's location

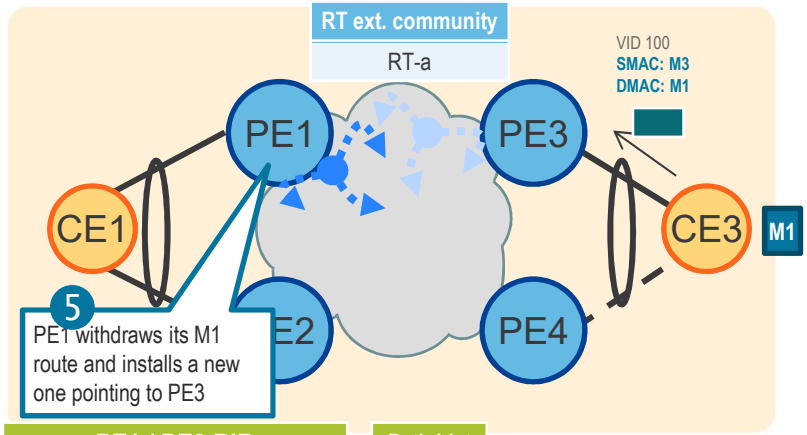


2 PE3 / PE4 install M1 route towards PE1

PE3, PE4 RIB			Path List
VPN	MAC	ESI	NH
RT-a	M1	ES1 →	PE1

PE3 MAC Route
RD = RD-3a
ESI = ESI2
MAC = M1
Label = L3
MAC Mobility ext. community.
Seq. Num = 2
RT ext. community
RT-a

4 After host sends traffic at new location, PE2 now adv MAC route for M1 incrementing sequence # in MAC mobility community

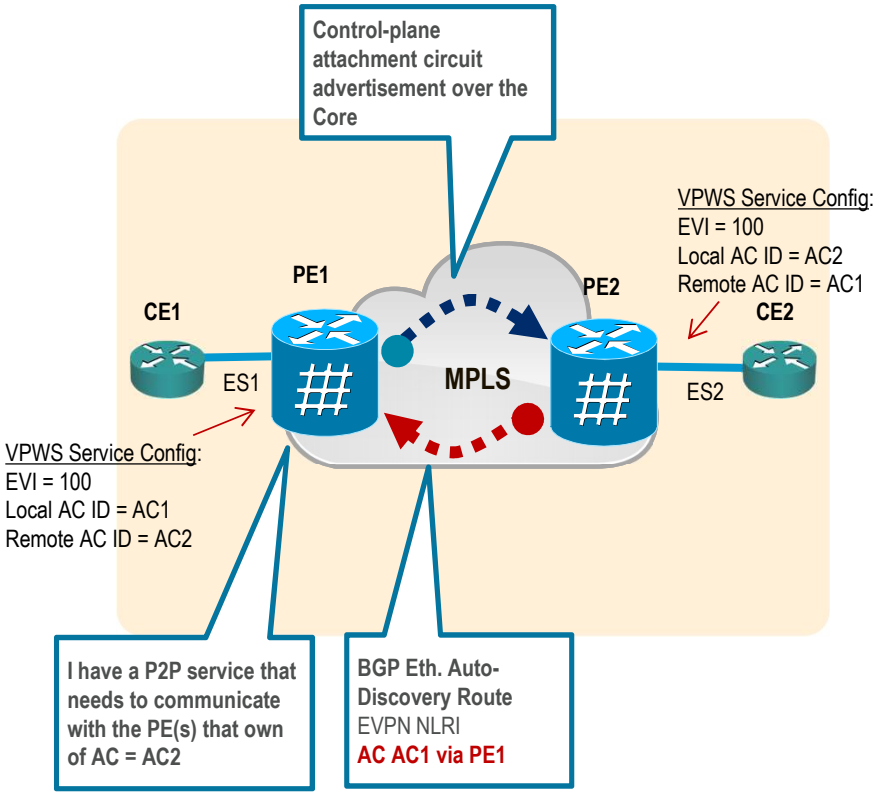


5 PE1 withdraws its M1 route and installs a new one pointing to PE3

PE1 / PE2 RIB			Path List
VPN	MAC	ESI	NH
RT-a	M1	ES2 →	PE3

EVPN VPWS

- Benefits of EVPN applied to point-to-point services
  - No signaling of PWs. Instead signals MP2P LSPs instead (ala L3VPN)
  - All-active CE multi-homing (per-flow LB)
  - Single-active CE multi-homing (per-service LB)
- Relies on a sub-set of EVPN routes to advertise Ethernet Segment and AC reachability
  - PE discovery & signaling via a single protocol – BGP
  - Per-EVI Ethernet Auto-Discovery route
  - Handles double-sided provisioning with remote PE auto-discovery
- Under standardization: draft-ietf-bess-evpn-vpws



EVPN VPWS Operation – All-active

ESI – 10 bytes ESI as specify by EVPN Ethernet segment IETF draft

Eth.Tag ID – 4-bytes local AC-ID

MPLS Label – (downstream assigned) used by remote PEs to reach segment

RD – RD unique per adv. PE per EVI

RT – RT associated with a given EVI

PE 1 Eth A-D Route
RD = RD-1a
ESI = ES1
Eth.Tag ID = AC1
Label (e.g. X)
RT ext. community
RT-a

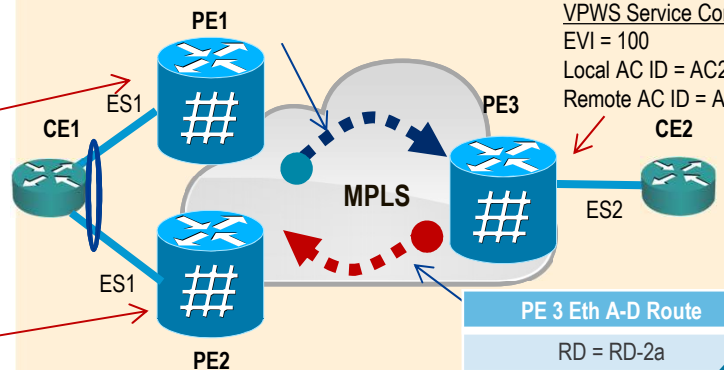
3

ALL-Active == per-flow load-baancing CE-PEs
Single bundle on CE device

VPWS Service Config:
EVI = 100
Local AC ID = AC1
Remote AC ID = AC2

1

VPWS Service Config:
EVI = 100
Local AC ID = AC1
Remote AC ID = AC2



VPWS Service Config:
EVI = 100
Local AC ID = AC2
Remote AC ID = AC1

2

ES2 – Since CE2 is single homed to PE2, ES2 = 0

Both PEs (PE1/PE2) shows as next hop for the remote AC

5

PE1 & PE2 RIB			
VPN	MAC	ESI	Eth.TAG
RT-a	-	0	AC2

Path List
NH
PE3

4

PE 3 Eth A-D Route
RD = RD-2a
ESI = ES2 (0)
Eth.Tag ID = AC2
Label (e.g. Y)
RT ext. community
RT-a

6

PE3 RIB			
VPN	MAC	ESI	Eth.TAG
RT-a	-	ES1	
RT-a	-	ES1	
RT-a	-	ES1	AC1

Path List
NH
PE1
PE2
PE1,PE2

EVPN VPWS Operation – Single-active

ESI – 10 bytes ESI as specify by EVPN Ethernet segment IETF draft

Eth.Tag ID – 4-bytes local AC-ID

MPLS Label – (downstream assigned) used by remote PEs to reach segment

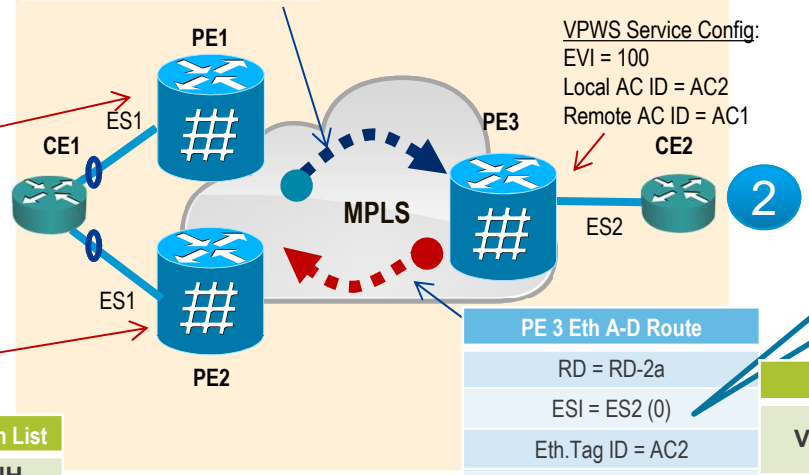
RD – RD unique per adv. PE per EVI

RT – RT associated with a given EVI

Single-Active == per-vlan load-balancing CE-PEs
Two bundles on CE device

PE 1 Eth A-D Route
RD = RD-1a
ESI = ES1
Eth.Tag ID = AC1
Label (e.g. X)
RT ext. community
RT-a

3



VPWS Service Config:
EVI = 100
Local AC ID = AC1
Remote AC ID = AC2

1

VPWS Service Config:
EVI = 100
Local AC ID = AC1
Remote AC ID = AC2

VPWS Service Config:
EVI = 100
Local AC ID = AC2
Remote AC ID = AC1

2

ES2 – Since CE2 is single homed to PE2, ES2 = 0

Only one PE (PE1) shows as next hop for the remote AC

PE 3 Eth A-D Route
RD = RD-2a
ESI = ES2 (0)
Eth.Tag ID = AC2
Label (e.g. Y)
RT ext. community
RT-a

4

5

PE1 & PE2 RIB			
VPN	MAC	ESI	Eth.TAG
RT-a	-	0	AC2

Path List
NH
PE3

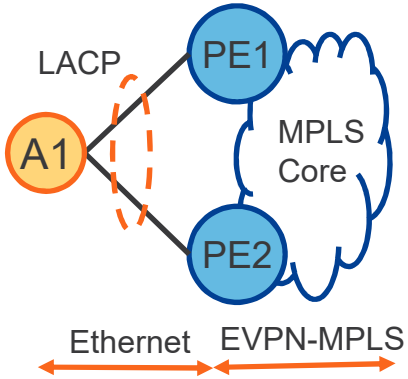
6

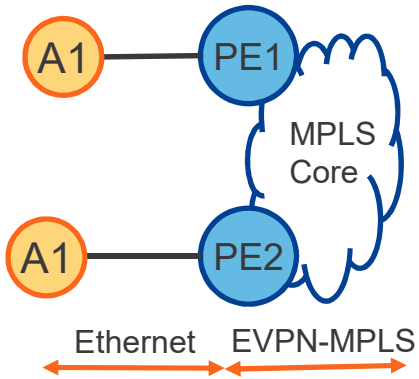
PE3 RIB			
VPN	MAC	ESI	Eth.TAG
RT-a	-	ES1	
RT-a	-	ES1	
RT-a	-	ES1	AC1

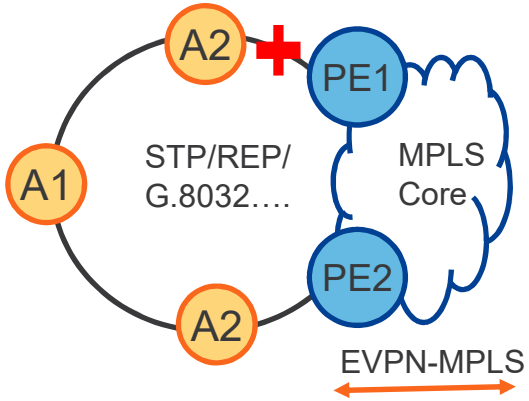
Path List
NH
PE1
PE2
PE1

EVPN Ethernet access

Single/Dual Homed Solution, Legacy L2 access

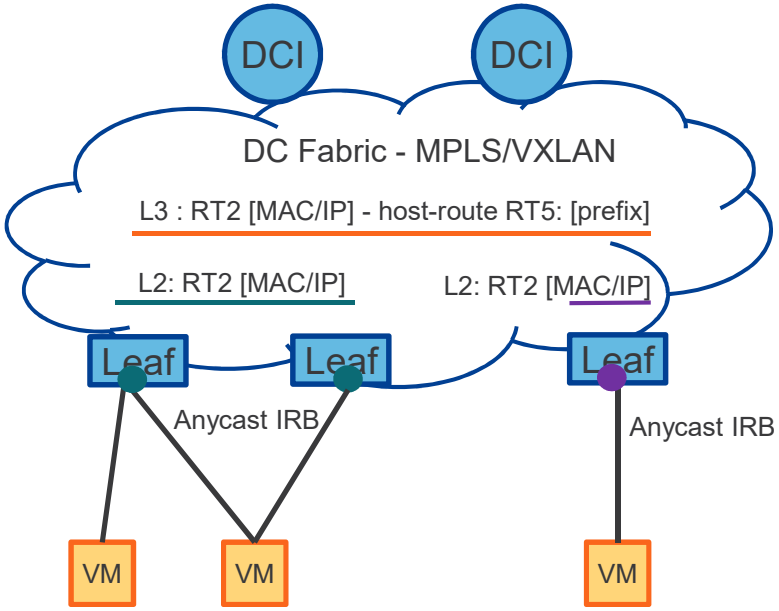




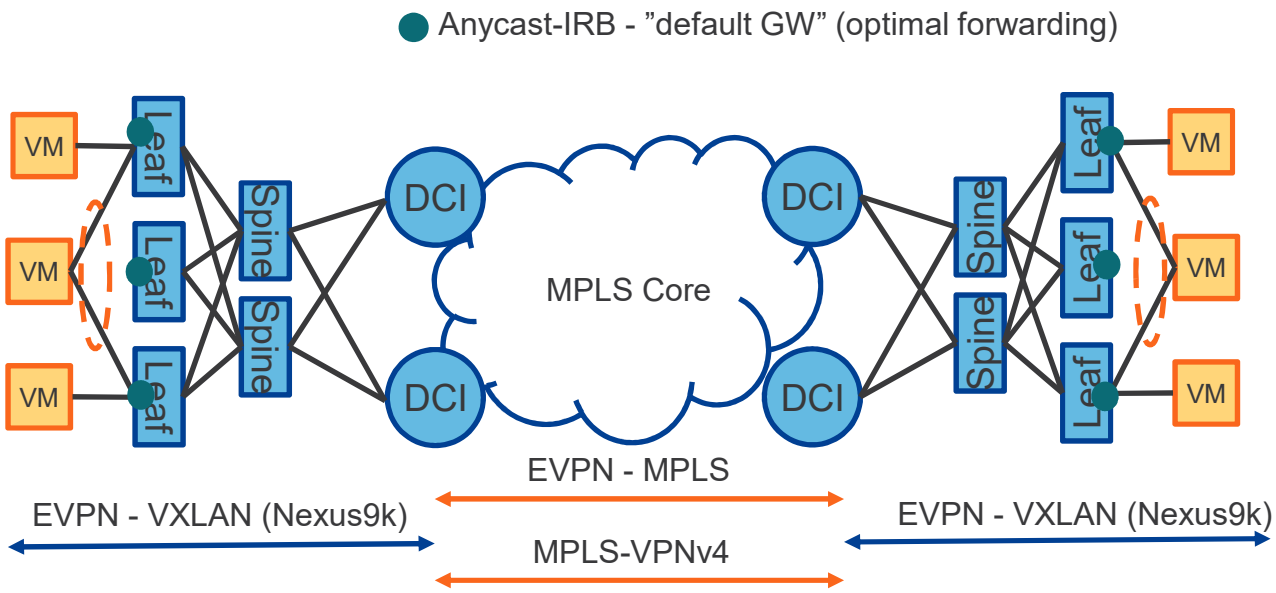


Symmetric Anycast IRB

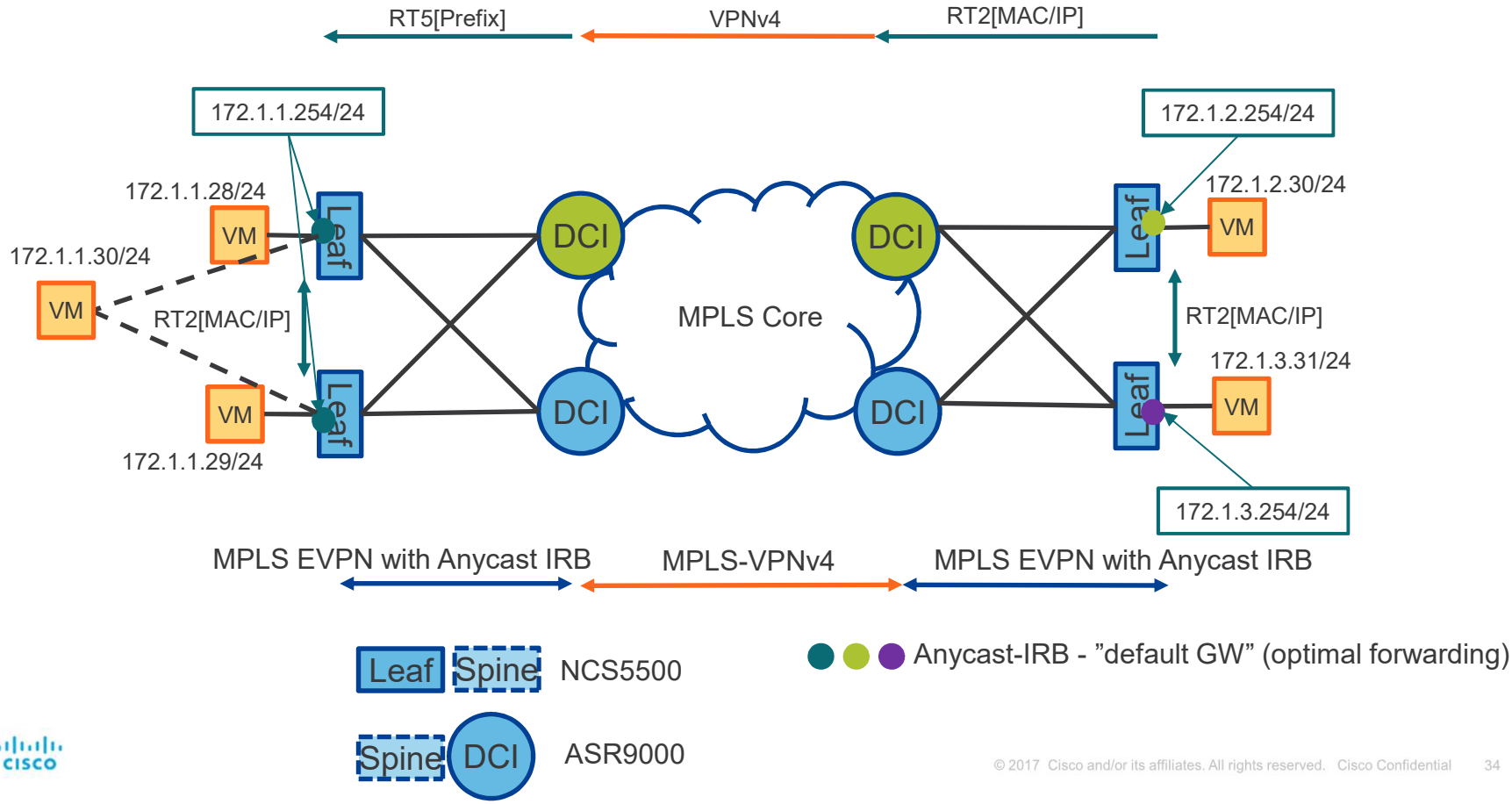
- Routing and Bridging in the same instance
- All-Active Multi-homed Access WITHOUT:
  - mLAG (mLACP)
  - VSS/vPCE…



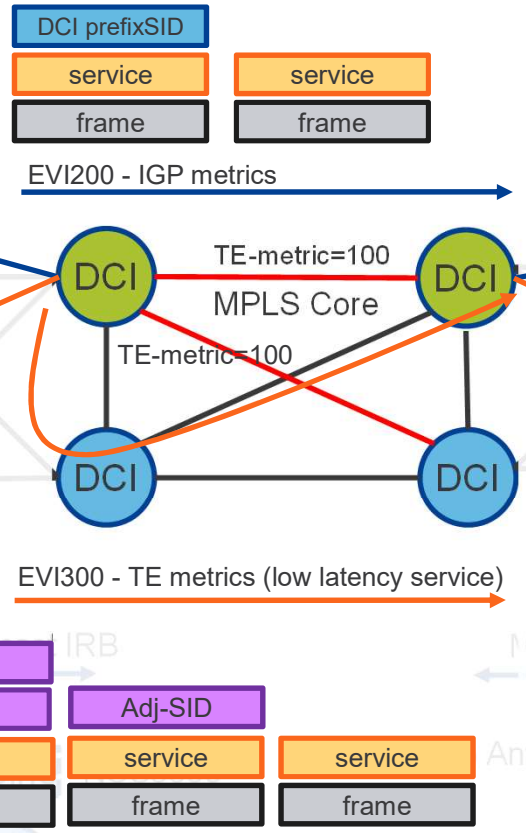
EVPN DCI Layer 2/3 gateway



SR & EVPN DC Fabric with Anycast IRB - Demo



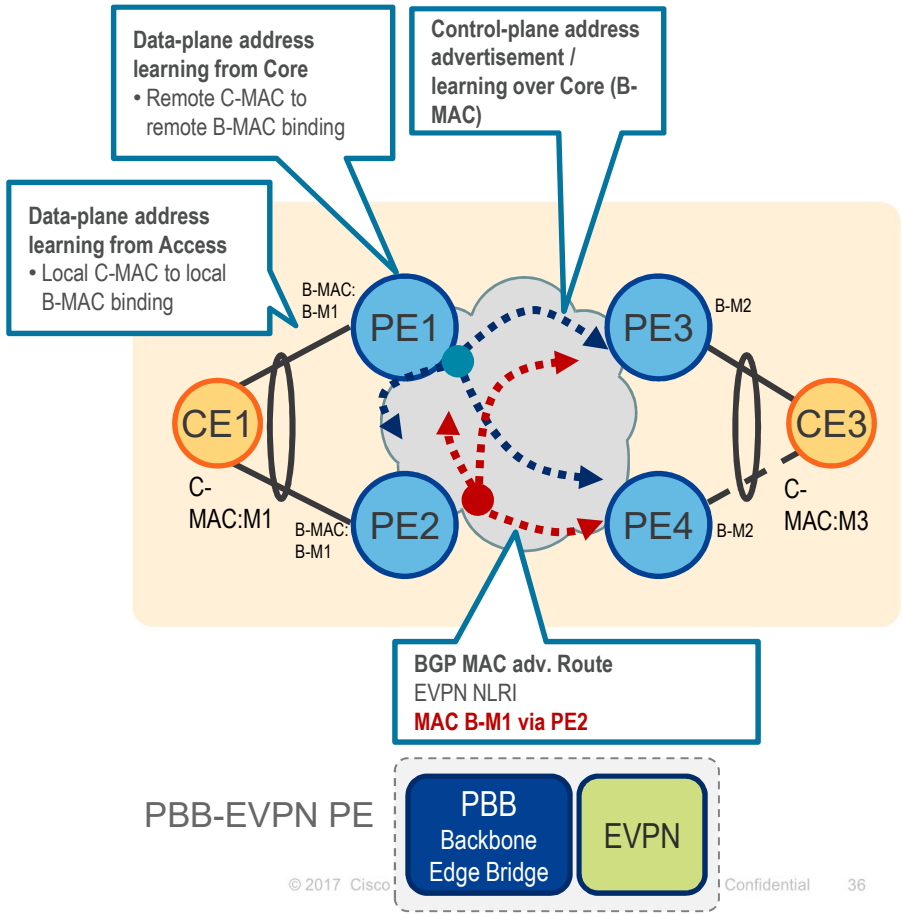
SR & EVPN-VPWS On-Demand Next-Hop



PBB Ethernet VPN

Highlights

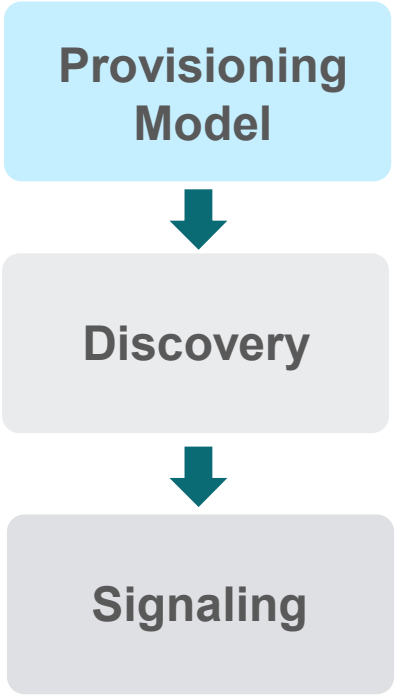
- Next generation solution for Ethernet multipoint (E-LAN) services by combining Provider Backbone Bridging (PBB - IEEE 802.1ah) and Ethernet VPN
- Data-plane learning of local C-MACs and remote C-MAC to B-MAC binding
- PEs run Multi-Protocol BGP to advertise local Backbone MAC addresses (B-MACs) & learn remote B-MACs
  - Takes advantage of PBB encapsulation to simplify BGP control plane operation – faster convergence
  - Lowers BGP resource usage (CPU, memory) on deployed infrastructure (PEs and RRs)
    - RFC7623



VPWS / VPLS

An abstraction

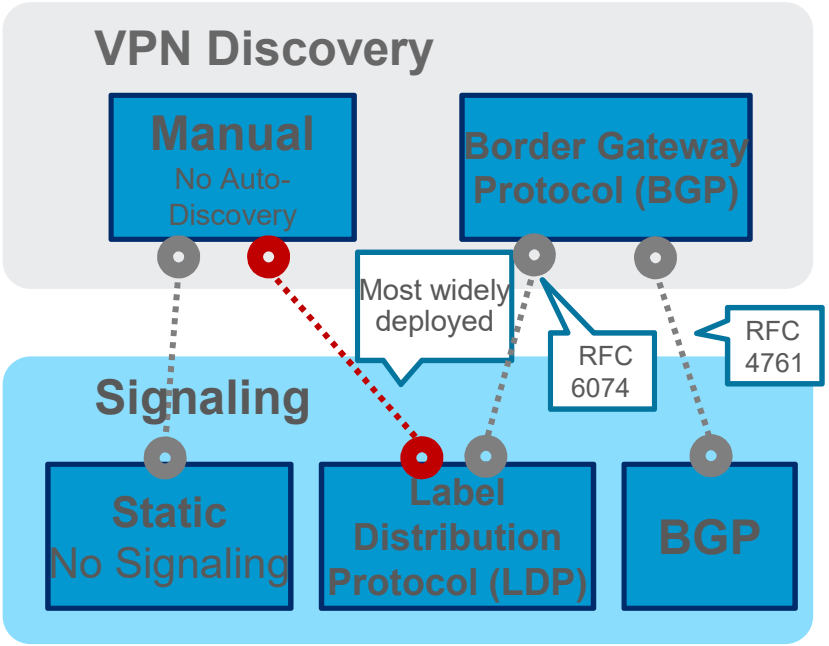
- Provisioning Model
  - What information needs to be configured and in what entities
  - Semantic structure of the endpoint identifiers (e.g. VC ID, VPN ID)
- Discovery
  - Provisioning information is distributed by a "discovery process“
  - Distribution of endpoint identifiers
- Signaling
  - When discovery process is complete, a signaling protocol is automatically invoked to set up pseudowires (PWs)



VPLS

Discovery and Signaling Alternatives

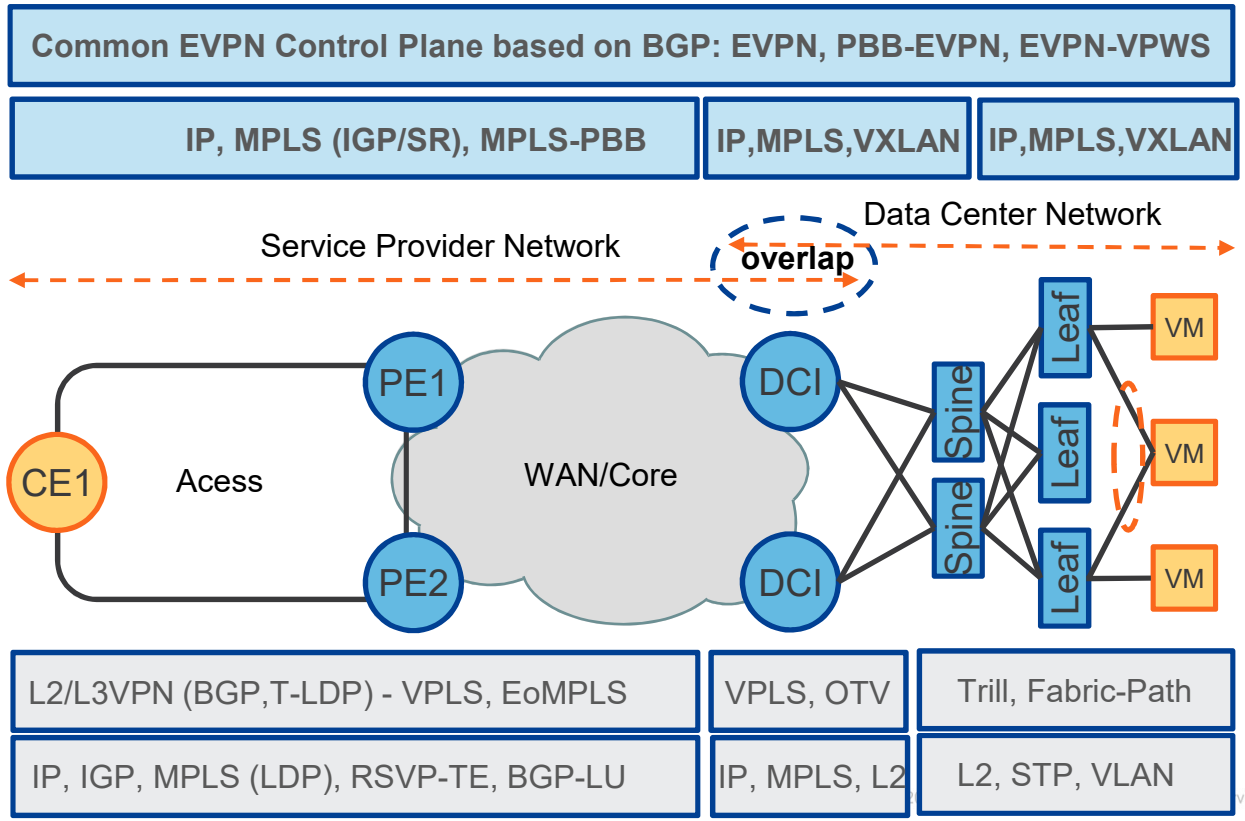
- VPLS Signaling
  - LDP-based (RFC 4762)
  - BGP-based (RFC 4761)
- VPLS with LDP-signaling and No auto-discovery
  - Most widely deployed solution
  - Operational complexity for larger deployments
- BGP-based Auto-Discovery (BGP-AD) (RFC 6074)
  - Enables discovery of PE devices in a VPLS instance



EVPN - End-to-End Control-Plane

Evolution:

Existing Solution:



Acknowledgment: Jose Liste

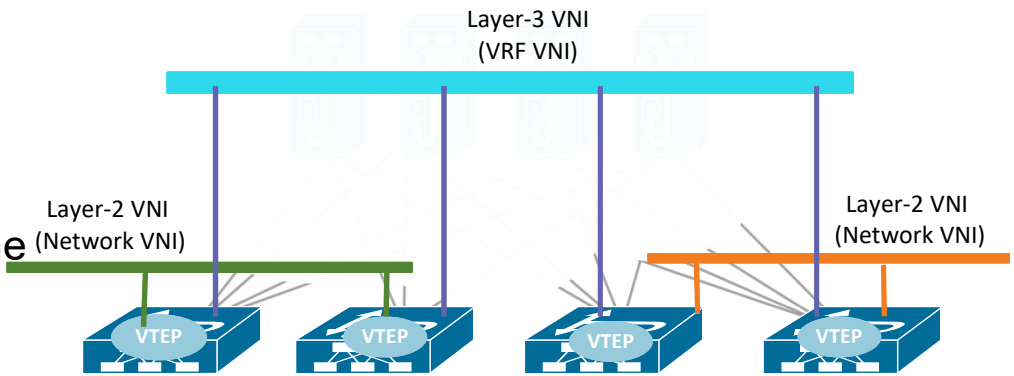
More Information

- RFC 7209: Requirements for Ethernet VPN (EVPN)
- RFC 7432: BGP MPLS-based Ethernet VPN
- RFC 7623: PBB-EVPN
- draft-ietf-bess-evpn-overlay: NVO solutions for EVPN
- draft-ietf-bess-evpn-vpws: VPWS support in EVPN
- draft-ietf-bess-evpn-inter-subnet-forwarding: IRB in EVPN
- draft-ietf-bess-evpn-ip-prefix-advertisement: IP prefixes in EVPN

CISCO

Symmetric EVPN IRB (1)

- Routing on both ingress and egress VTEPs
- Layer-3 VNI
  - Tenant VPN indicator
  - One per tenant VRF
- VTEP Router MAC
- Ingress VTEP routes packets onto the Layer-3 VNI
- Egress VTEP routes packets to the destination Layer-2 VNI



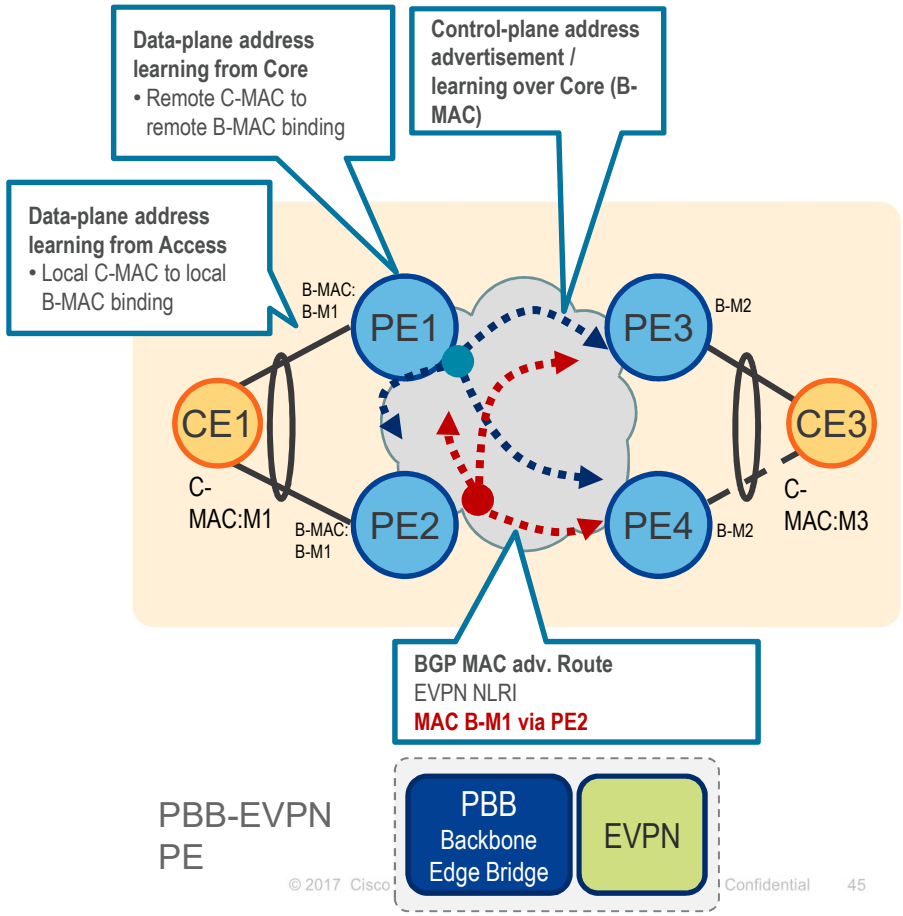
To PBB or not to PBB?

- What is the value of combining PBB and EVPN functions?
- Lower control-plane overhead than EVPN alone
  - PBB-EVPN uses only a sub-set of EVPN routes
  - Simpler and Faster failure convergence for all-active multi-homing scenarios
  - Faster MAC move convergence handled in data-plane
- Lower control-plane scale requirements than EVPN alone
  - BGP MAC advertisements for smaller Backbone MAC (B-MAC) address space
  - Requires less resources (CPU, memory) on deployed infrastructure (PEs / RRs)

PBB Ethernet VPN

Highlights

- Next generation solution for Ethernet multipoint (E-LAN) services by combining Provider Backbone Bridging (PBB - IEEE 802.1ah) and Ethernet VPN
- Data-plane learning of local C-MACs and remote C-MAC to B-MAC binding
- PEs run Multi-Protocol BGP to advertise local Backbone MAC addresses (B-MACs) & learn remote B-MACs
  - Takes advantage of PBB encapsulation to simplify BGP control plane operation – faster convergence
  - Lowers BGP resource usage (CPU, memory) on deployed infrastructure (PEs and RRs)
- RFC7623



EVPN Advantages:

Integrated Services

- Integrated Layer 2 and Layer 3 VPN services
- L3VPN-like principals and operational experience for scalability and control

Network Efficiency

- All-active Multi-homing & PE load-balancing (ECMP)
- Fast convergence (link, node, MAC moves)
- Control-Place (BGP) learning. PWs are no longer used.
- Optimized Broadcast, Unknown-unicast, Multicast traffic delivery

Service Flexibility

- Choice of MPLS, VxLAN or PBB data plane encapsulation
- Support existing and new services types (E-LAN, E-Line, E-TREE)
- Peer PE auto-discovery. Redundancy group auto-sensing

Investment Protection

- Operational consistency with L3 IP VPN
- Fully support IPv4 and IPv6 in the data plane and control plane
- Open-Standard and Multi-vendor support