NOKIA

EVPN

A tutorial

- Paresh Khatri
- 2019

Agenda

EVPN Background and Motivation

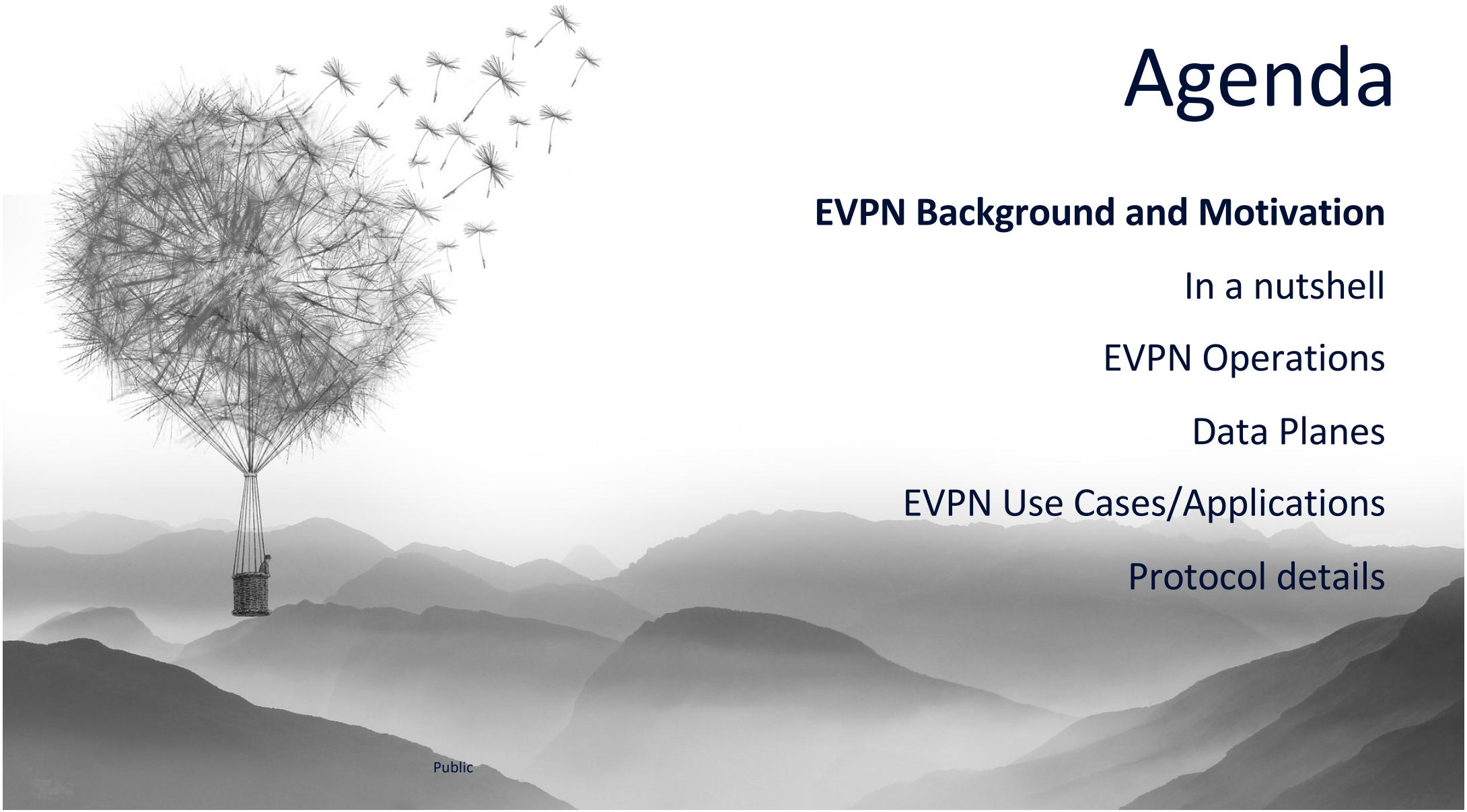
In a nutshell

EVPN Operations

Data Planes

EVPN Use Cases/Applications

Protocol details



Agenda

EVPN Background and Motivation

In a nutshell

EVPN Operations

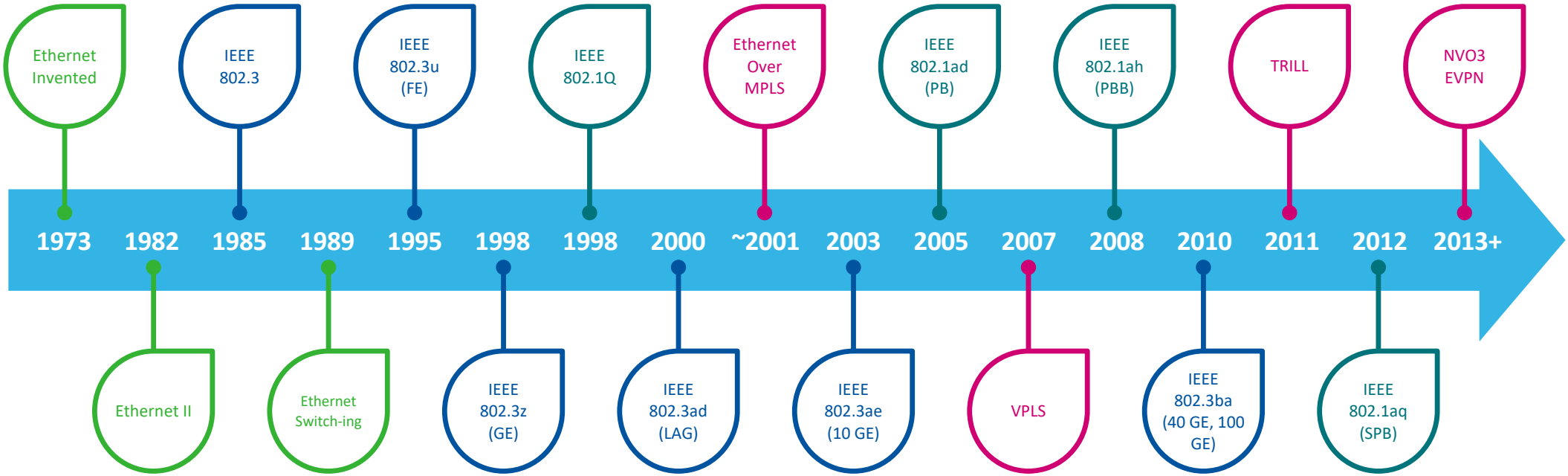
Data Planes

EVPN Use Cases/Applications

Protocol details

Ethernet Services Technology Continues to Evolve

Higher Speeds and Advanced Carrier-Grade Services



"The **widespread adoption of Ethernet L2VPN** services and the advent of **new applications** for the technology (e.g., data center interconnect) have culminated in a **new set of requirements** that are **not readily addressable** by the current Virtual Private LAN Service (VPLS) solution." — RFC7209

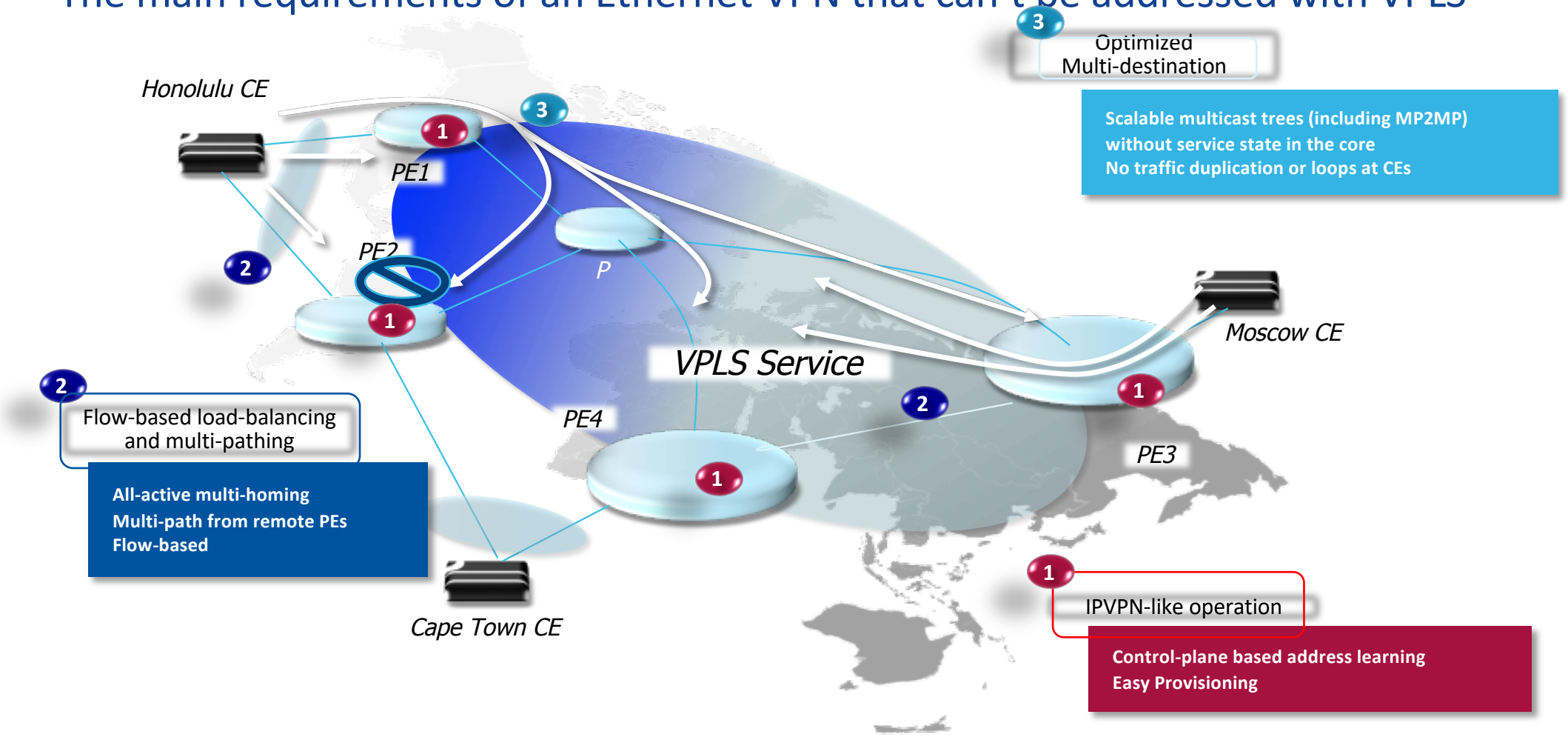
EVPN and the opportunity to make it right

- What have we learnt about VPNs
  - IP-VPN (RFC4364) is successfully deployed in SP networks without interop issues, easy to provision, supports all-active MH but only IP traffic
  - VPLS (RFC4761/4762/6074) has control plane interop issues, provisioning vs efficiency trade-offs, flood-and-learn is not optimum, but works for any Ethernet traffic

- Why another VPN technology
  - Cloud and NFV are shifting the way networks must behave
  - EVPN is an Ethernet VPN technology (provides L2 and L3) that provides the required flexibility, it is future-proof and inherits over a decade of VPN experience

- Where can we use EVPN
  - Cloud and virtualization services
  - Data Center Interconnect (DCI)
  - Integrated Layer-2 and Layer-3 VPN services
  - Overlay technologies that simplify topologies and protocols

The main requirements of an Ethernet VPN that can't be addressed with VPLS



EVPN Requirements and Benefits

	VPN Requirements	VPLS	EVPN	What does it do for me?
Address Learning	Control Plane Address Learning in the Core	✘	✓	Greater Scalability and Control
Provisioning	L3VPN-Like Operation	✘	✓	Simpler Provisioning and Automation
	Auto Discovery and Configuration	PEs Only	✓	Simpler Provisioning and Automation
Resiliency	Active-Standby Multihoming (Service-Based Load Balancing)	✓	✓	Standby Redundancy
	All-Active Multihoming (Flow-Based Load Balancing)	✘	✓	Active Redundancy and Link Utilization
Services	VLAN Based Service Interfaces	✓	✓	Virtualization and Advanced Services
	VLAN Aware Bundling Service Interfaces	✘	✓	Virtualization and Advanced Services
	Inter-Subnet Forwarding	✘	✓	Layer 2 and Layer 3 Over the Same Interface
Flow Optimization	ARP/ND Proxy	✘	✓	Security and MAC Provisioning
	MAC Mobility	✘	✓	Virtualization and Advanced Services

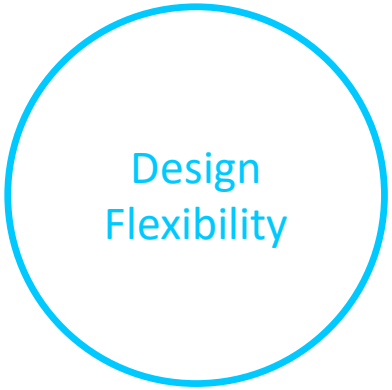
Overview: Next-Generation Ethernet VPN (EVPN)

- EVPN overview:
  - EVPN (RFC7432) is a BGP MPLS-Based Ethernet VPN, that uses a new MP-BGP address family to support MAC learning. It allows VPLS services to be operated as the L3VPN-like for better scalability and flexibility
  - EVPN is used to fill the gaps of other L2VPN technologies such as VPLS. The main objective of the EVPN is to build ELAN services in a similar way to RFC4364 IP-VPNs, while supporting MAC learning within the control plane (distributed by MP-BGP).
  - Multi-homing with all-active forwarding; Optimizing the delivery of multi-destination traffic (BUM)
  - Efficient hybrid services over a single VLAN; Delivering both L2 and L3 services over the same interface
  - EVPN used as control plane with multiple data plane encapsulations (VXLAN and MPLS)

EVPN key benefits

- Uniform control plane (MP-BGP) for L2/L3, p2p/p2mp service.
- L3VPN-like operation for scalability and control
- Seamless integration with existing L2 services.

- A/S, A/A Multi-homing with per flow redundancy and load balancing.
- Massive scale, with efficient BUM handling.
- Mass withdrawal
- More efficient hybrid service delivery over a single interface or VLAN



- MPLS or IP data plane encapsulation choices
- VXLAN or MPLSoUDP encapsulations enable EVPN over a simple IP network
- Simpler provisioning and management with a single VPN technology

- MAC/IP provisioning enables programmatic network control
- Consistent signaled FDB in control plane vs. flood-and-learn FDB in data plane
- Proxy ARP/ND to reduce/suppress BUM traffic
- Improved security (MAC/ARP/ND)

NOKIA

EVPN in the Standards Organizations

EVPN Application/Service	Standard document
ELAN	RFC7432 (EVPN) RFC7623 (PBB-EVPN)
ELINE	RFC8214 (EVPN-VPWS)
ETREE	RFC8317 (EVPN and PBB-EVPN E-Tree)
L3 VPN (Inter-subnet-forwarding)	draft-ietf-bess-evpn-inter-subnet-forwarding draft-ietf-bess-evpn-prefix-advertisement
EVPN for DC	RFC8365 draft-ietf-bess-evpn-optimized-ir
EVPN for DCI	draft-ietf-bess-dci-evpn-overlay draft-ietf-bess-evpn-vpls-seamless-integ
New applications - Multi-homing improvements - Proxy-ARP/ND and security - BUM optimizations - EVPN to IP-VPN interworking - Mcast	draft-ietf-bess-evpn-df-election-framework 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 draft-ietf-bess-evpn-irb-mcast draft-rabadan-sajassi-bess-evpn-ipvpn-interworking draft-ietf-bess-evpn-pim-proxy

Agenda

EVPN Background and Motivation

In a nutshell

EVPN Operations

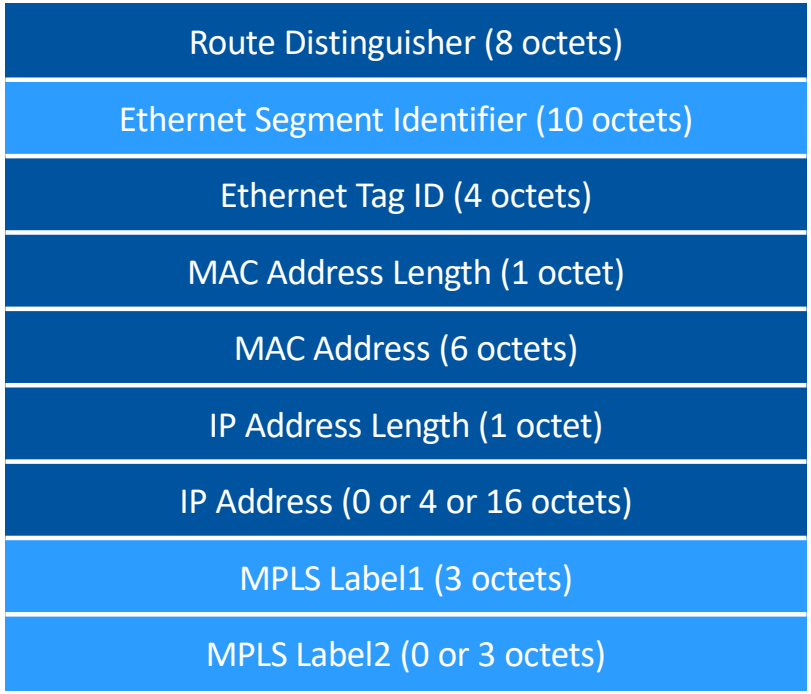
Data Planes

EVPN Use Cases/Applications

Protocol details

EVPN Control Plane Learning with MP-BGP

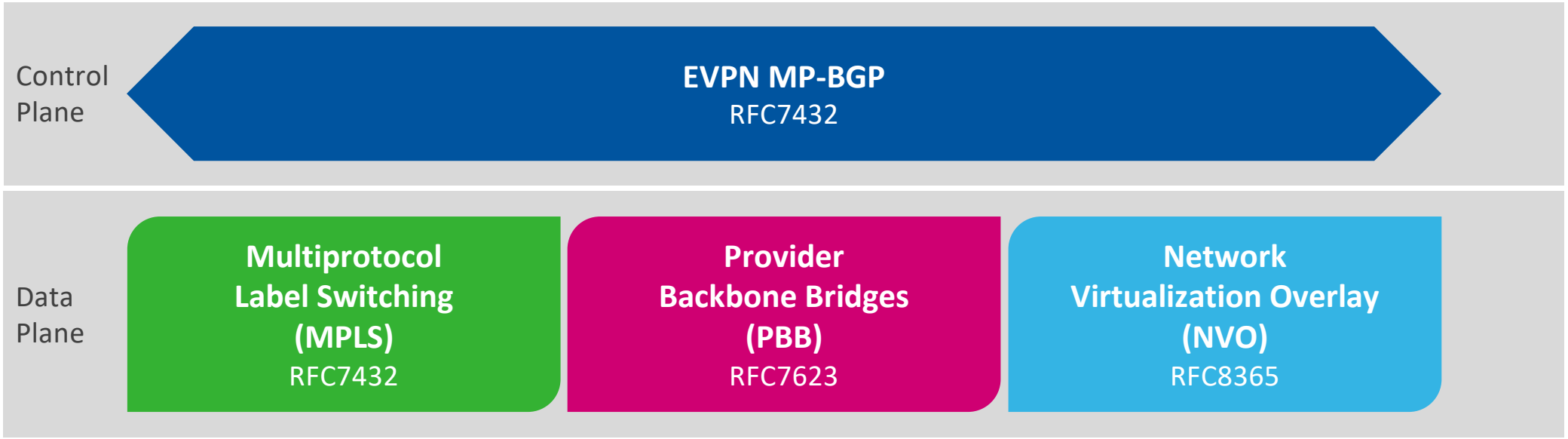
- Brings proven and inherent BGP control plane scalability to MAC routes
  - Consistent signaled FDB in any size network instead of flooding
  - Even more scalability and hierarchy with route reflectors
- BGP advertises MACs and IPs for next hop resolution with EVPN NLRI
  - AFI = 25 (L2VPN) and SAFI = 70 (EVPN)
  - Fully supports IPv4 and IPv6 in the control and data plane
- Offers greater control over MAC learning
  - What is signaled, from where and to whom
  - Ability to apply MAC learning policies
- Maintains virtualization and isolation of EVPN instances
- Enables traffic load balancing for multihomed CEs with ECMP MAC routes



MAC Advertisement Route
(Light Blue Fields are Optional)

EVPN Data Planes

One EVPN Control Plane with Multiple Data Plane Options



- EVPN over MPLS for VLL, VPLS and E-Tree services
- All-active multihoming for VPWS
- RSVP-TE or LDP MPLS protocols

- EVPN with PBB PE functionality for scaling very large networks over MPLS
- All-active multihoming for PBB-VPLS

- EVPN over NVO tunnels (VXLAN, NVGRE, MPLSoGRE) for data center fabric encapsulations
- Provides Layer 2 and Layer 3 DCI

Terminology (from RFCs)

EVPN: Ethernet VPN

EVI: An EVPN instance spanning the Provider Edge (PE) devices participating in that EVPN.

MAC-VRF: A Virtual Routing and Forwarding table for Media Access Control (MAC) addresses on a PE.

Ethernet Segment (ES): When a customer site (device or network) is connected to one or more PEs via a set of Ethernet links, then that set of links is referred to as an 'Ethernet segment'.

Ethernet Segment Identifier (ESI): A unique non-zero identifier that identifies an Ethernet segment is called an 'Ethernet Segment Identifier'.

Ethernet Tag: An Ethernet tag identifies a particular broadcast domain, e.g., a VLAN. An EVPN instance consists of one or more broadcast domains.

Terminology (from RFCs)

Single-Active Redundancy Mode: When a device or a network is multihomed to a group of two or more PEs and when only a single PE in such a redundancy group can forward traffic to/from the multihomed device or network for a given VLAN, such multihoming is referred to as "Single-Active".

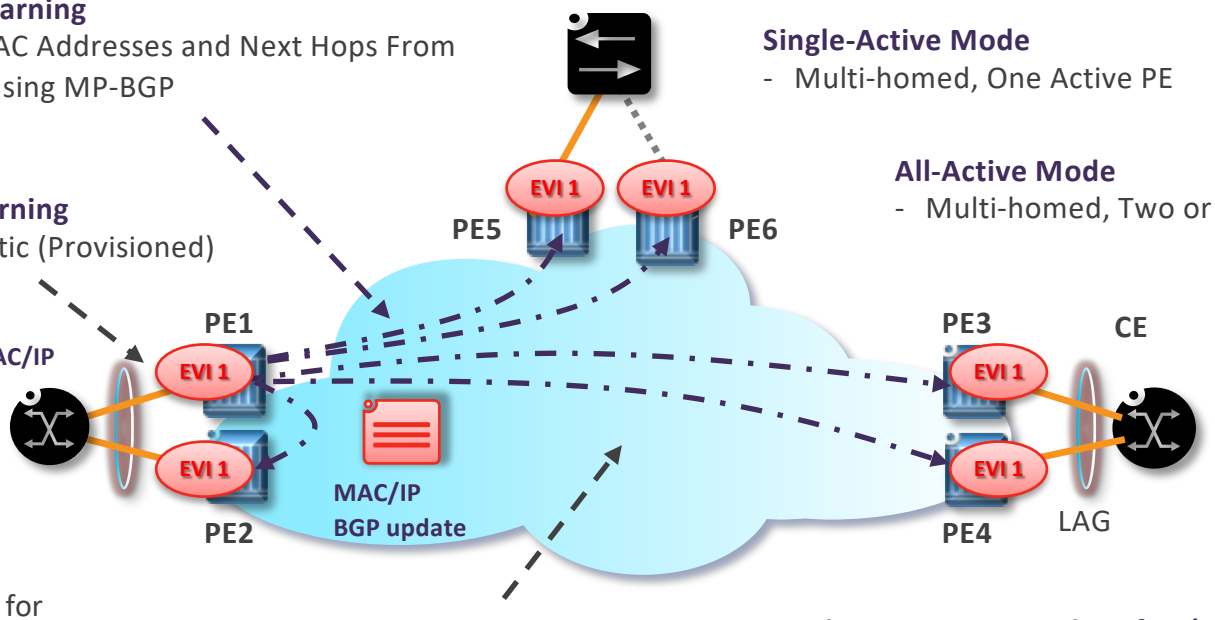
All-Active Redundancy Mode: When a device is multihomed to a group of two or more PEs and when all PEs in such redundancy group can forward traffic to/from the multihomed device or network for a given VLAN, such multihoming is referred to as "All-Active".

EVPN fundamental concepts

Control Plane Learning
PEs Advertise MAC Addresses and Next Hops From Connected CEs Using MP-BGP

Data Plane Learning
Dynamic or Static (Provisioned)

Customer Edge (CE)
Host, VM, Router or Switch



Single-Active Mode
- Multi-homed, One Active PE

All-Active Mode
- Multi-homed, Two or More Active PEs (4)

MAC-VRF
Virtual Routing Forwarding table for MACs

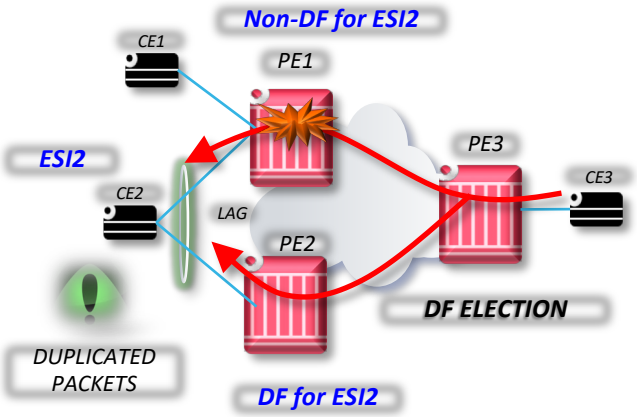
EVPN Instance (EVI)
Group of MAC-VRFs forming a VPN

Data Plane Encapsulation
RSVP-TE, LDP, BGP 3107
Segment Routing, IP fabric
(VXLAN or MPLSoUDP)

Ethernet Segment Identifier (ESI)
- Link(s) that Connect the CE to PEs (ESIs are Unique Across the Network)

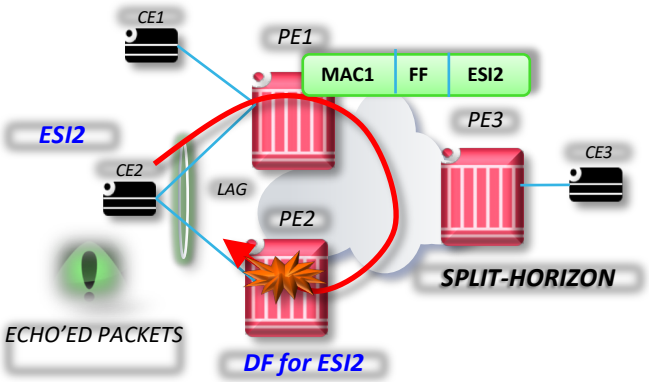
NOKIA

The three all-active MH challenges solved... by EVPN



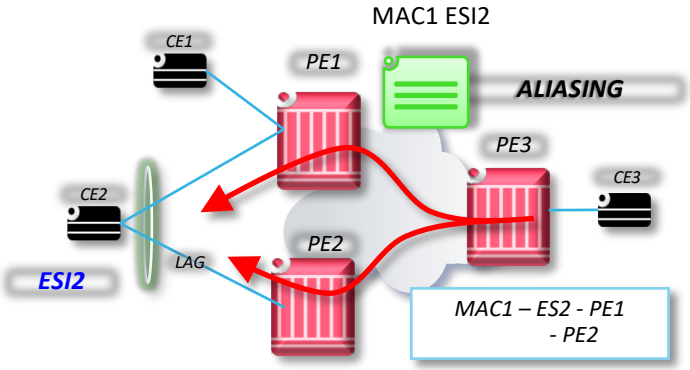
The DF election avoids duplicate BUM flooding to all-active CEs

- PE1/PE2 advertise their ESI
- EVPN elects a DF per ESI per service
- DF is responsible for BUM flooding into the Ethernet Segment



Split-horizon ensures that BUM traffic sent to the non-DF is not replicated back to the ESI

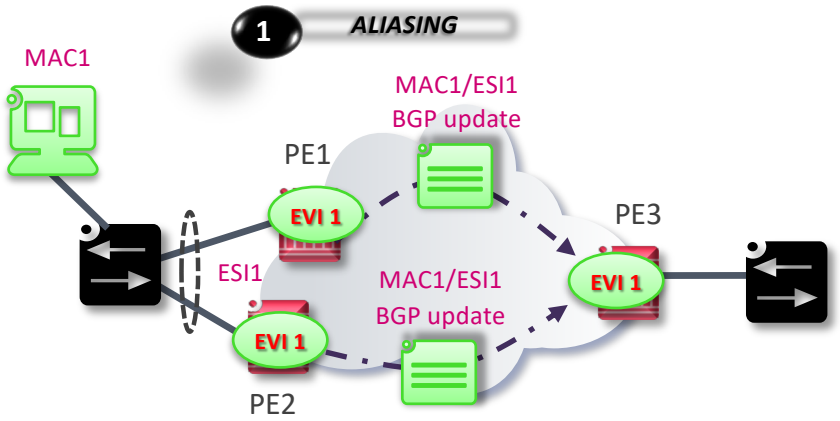
- The DF signals an ESI label that the non-DF uses to send BUM traffic to the DF
- The DF uses the ESI label to suppress the BUM to the ESI identified by the label



Aliasing allows load-balancing to the PEs part of the ESI

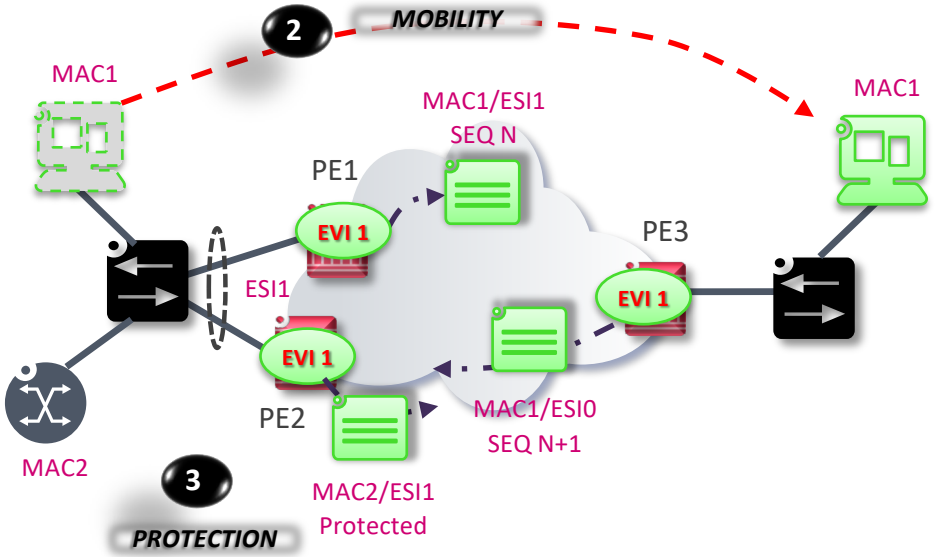
- EVPN advertises what PEs are part of the ESI
- PE3 does ECMP to all the ESI owners

EVPN supports MAC mobility, duplication and protection



A MAC advertised by two PEs using the same ESI is interpreted by the remote PEs as a multihomed MAC

- This function is used for aliasing
- It can also be used for “anycast” forwarding (if ecmp=1)

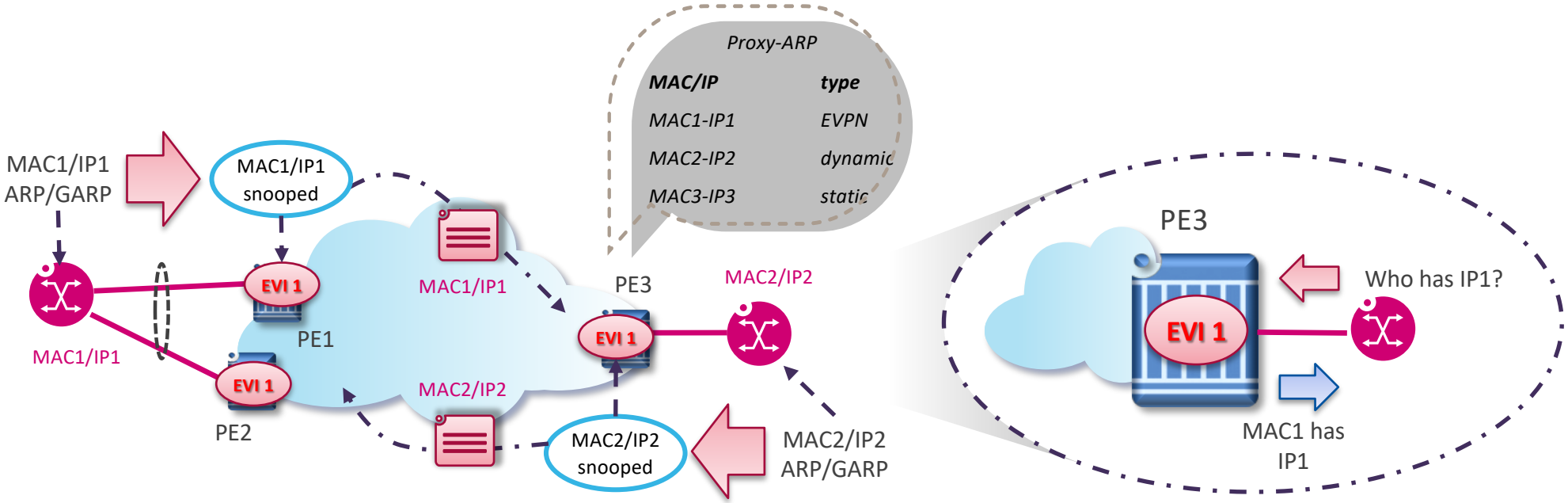


A MAC advertised by two PEs using different ESI is interpreted as mobility (until a threshold is reached)

- A SEQ number is incremented each time the MAC is advertised from a different ESI
- If MAC1 moves X times in Y minutes (configurable) mac-duplication is triggered

A MAC advertised as protected will not be overridden by the default PEs, and offending packets will be dropped

EVPN proxy-ARP/ND and unknown flooding suppression

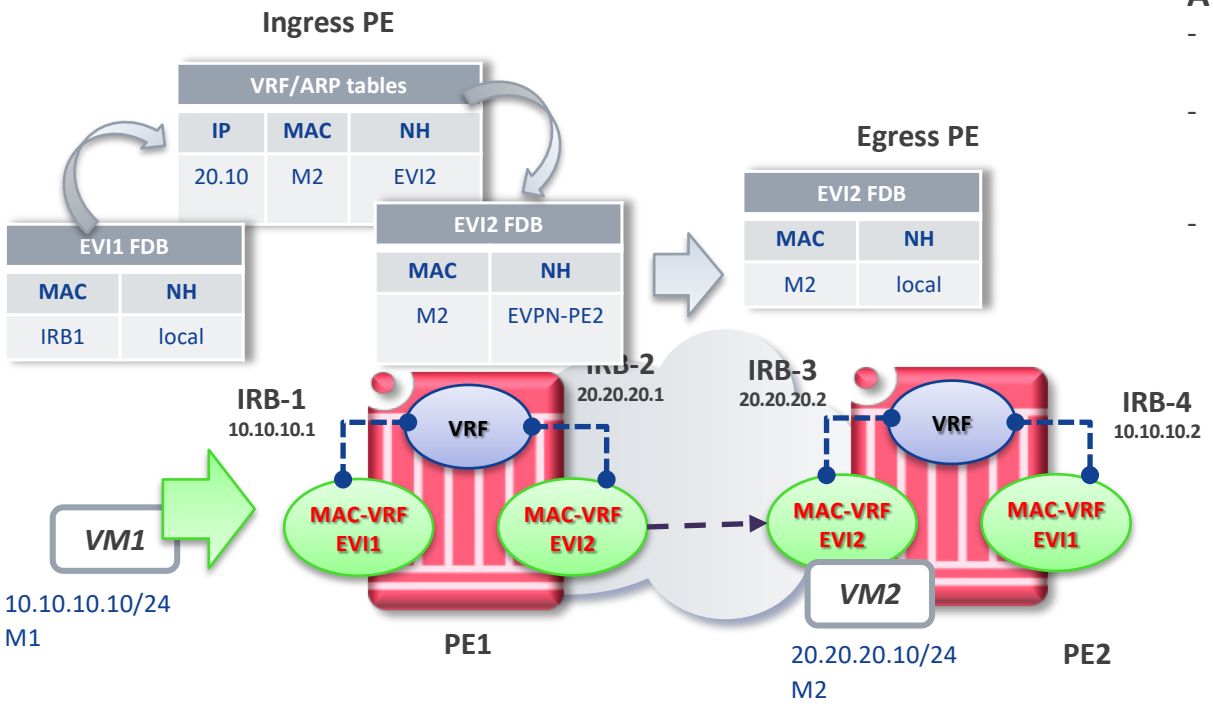


ARP/ND flooding and unknown flooding is a security issue and scalability issue in large L2 networks
- Especially in large DCs and IXP networks

EVPN can reduce/suppress unknown and ARP/ND flooding if all the hosts advertise their presence
- PEs build proxy-ARP/ND tables per EVI with MAC/IP info received from local snooped ARP/GARP/ND or EVPN or static entries
- PEs reply to ARP/ND requests/solicitation messages received on local ACs

EVPN provides integrated L2 and L3 forwarding

Asymmetric IRB model (draft-sajassi-l2vpn-evpn-inter-subnet-forwarding)



A customer (or tenant) is given:

- An EVI per subnet which exists in all the PEs in the network
- A VRF on each PE that has IRBs to all the MAC-VRFs for the customer and can forward traffic among all the subnets
- EVPN advertises the IRB MAC/IPs and learnt host MAC/IPs

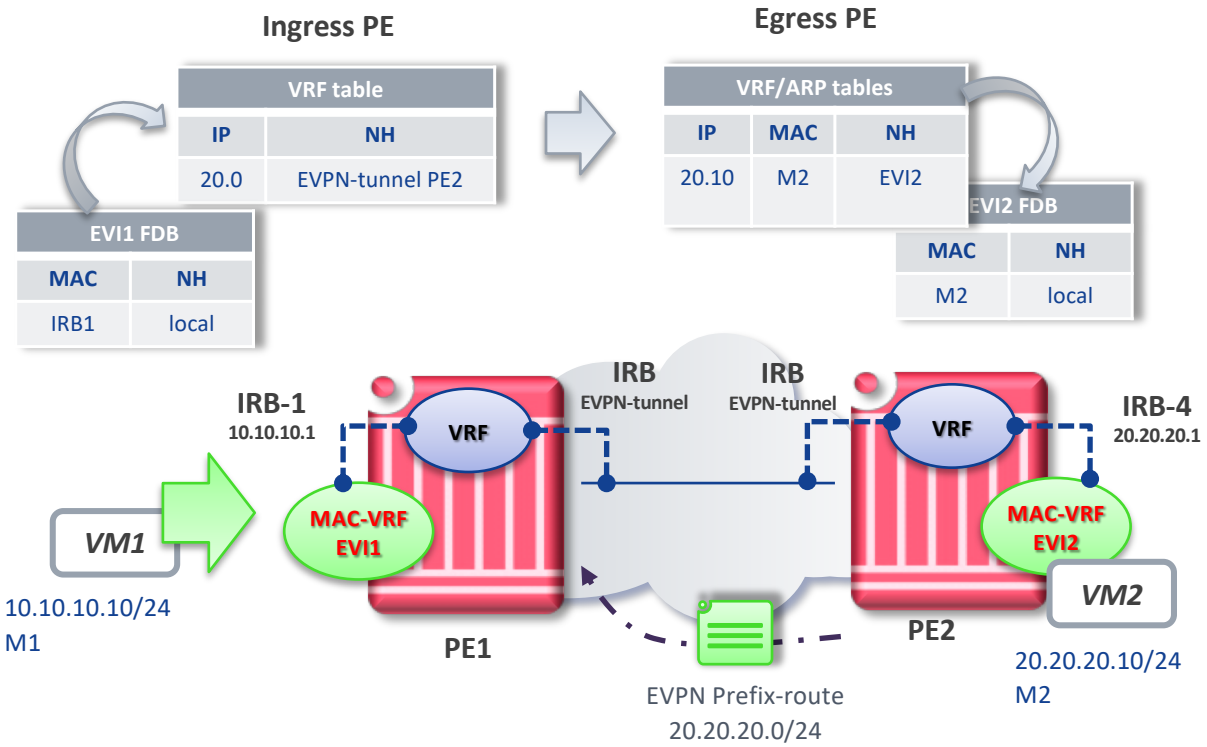
When a host sends traffic to a remote subnet:

- At the ingress PE
  - FDB lookup yields IRB interface
  - Routing/ARP lookup yields local EVI and remote MAC/PE
- At the egress PE
  - Only FDB lookup is required

NOTE: MAC-VRF is an EVI instance in a given PE

EVPN provides integrated L2 and L3 forwarding

Symmetric IRB model (draft-rabadan-l2vpn-evpn-prefix-advertisement)



A customer (or tenant) is given:

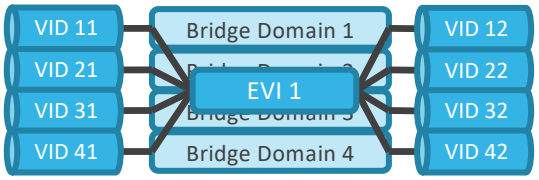
- An EVI per subnet which exists ONLY where there are hosts for that subnet
- A VRF on each PE that has IRBs to the local MAC-VRFs and a EVPN-tunnel IRB (no IP)
- Host MAC/IPs in one EVI are not imported by the remote PEs if the EVI is not local
- EVPN advertises IP prefixes that are imported in the VRF routing table

When a host sends traffic to a remote subnet:

- At the ingress PE
  - FDB lookup yields IRB interface
  - Routing lookup yields remote PE and MAC DA
- At the egress PE
  - Routing/ARP lookup yields MAC and local EVI
  - FDB lookup yields the local AC

The symmetric model saves ARP and FDB entries

EVPN Service Interfaces Overview

	VLAN Based Service Interface	VLAN Bundle Service Interface	VLAN Aware Bundle Service Interface
Service Interface Diagram			
Mapping Between VLAN ID and EVI	1:1	N:1	N:1
Broadcast Domains (VLAN) per EVI	Single	Multiple	Multiple
Bridge Domains per EVI	Single	Single	Multiple
Overlapping MACs Across VLANs	✓	✗	✓
VLAN Translation	✓	✗	✓

Agenda

EVPN Background and Motivation

In a nutshell

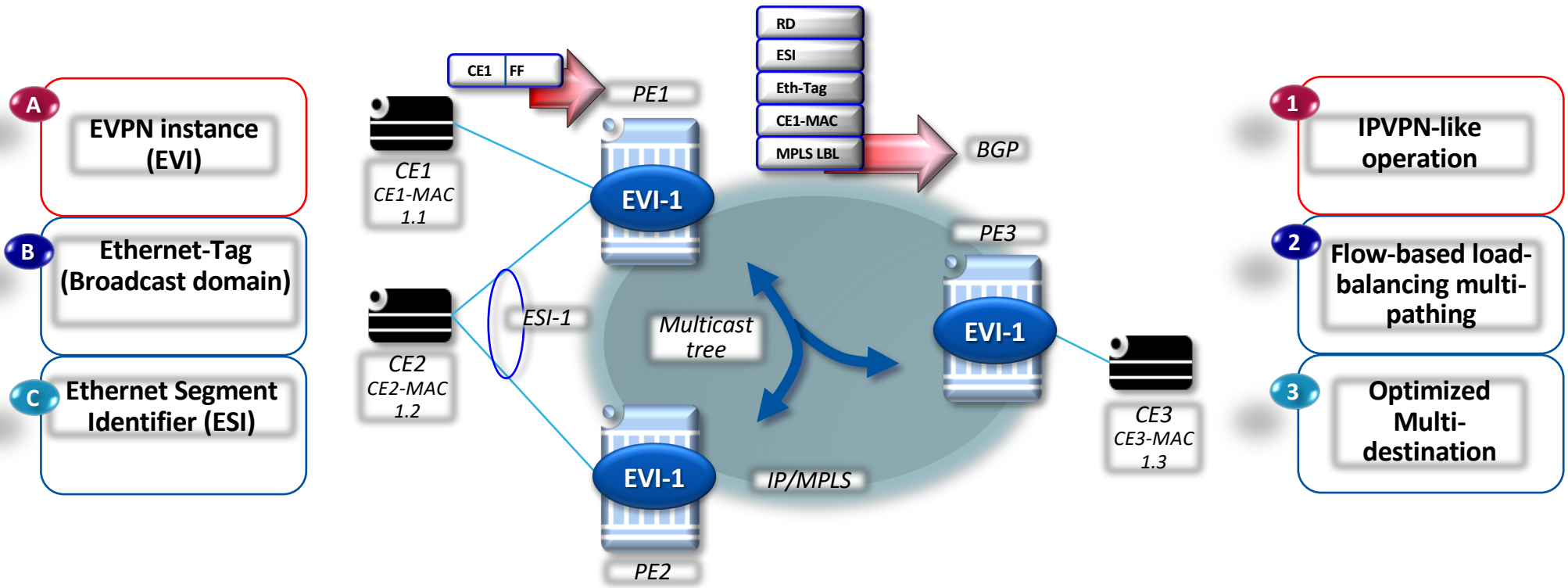
EVPN Operations

Data Planes

EVPN Use Cases/Applications

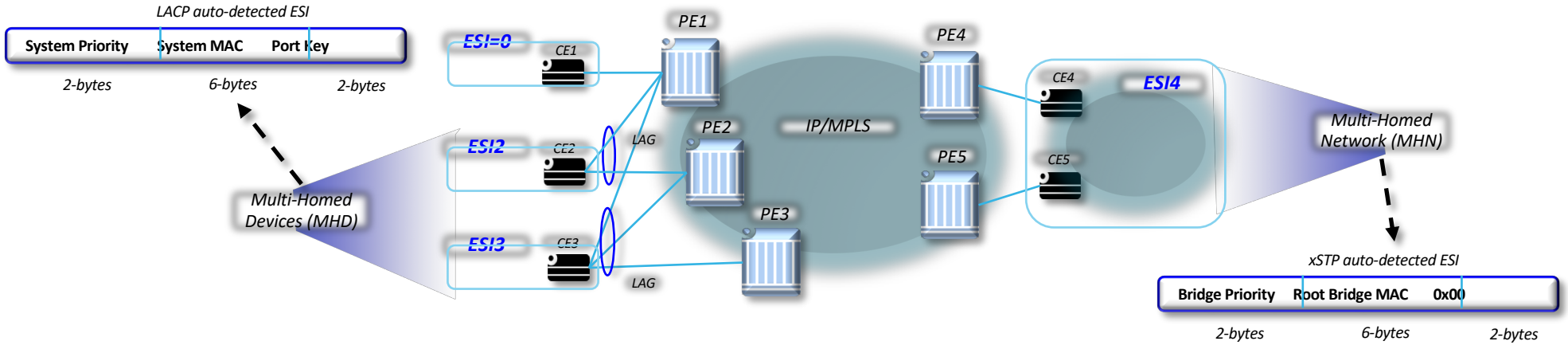
Protocol details

E-VPN in three concepts and three operational properties



E-VPN provides an IP-VPN-like operation with flow-based multi-homing/multi-pathing, optimal multi-destination and fast convergence

A key E-VPN concept: Ethernet Segments and ESIs

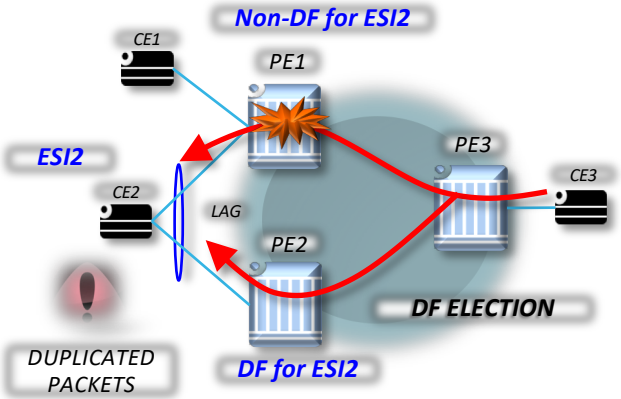


- Ethernet Segment (ES) is a site
- Ethernet Segment Identifier is unique 10-byte identifier
- Reserved ESIs
  - ESI=0 for SHD
  - MAX-ESI=0xFF (10x times); for PBB-EVPN

Two types of multi-homed ES:
- Single-active
- All-active (only if LAG is used at the CE)

ESI provisioning:
- Manually provisioned (if CEs are not managed by the Service Provider)
- Auto-detected (if CEs are managed by the Service Provider)

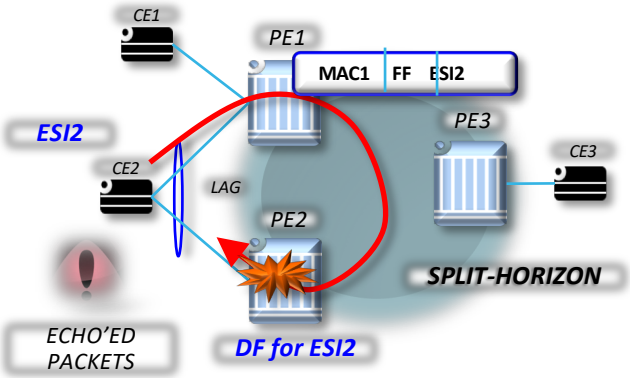
E-VPN optimal forwarding for all-active MH is provided by the Designated Forwarder (DF), split-horizon and aliasing functions



Issue: how to avoid duplicated BUM from PE3 on ESI2 (all-active MH)

Solution: DF election (non-DF will block all the BUM to ESI2, unicast can still follow both paths for all-active MH)

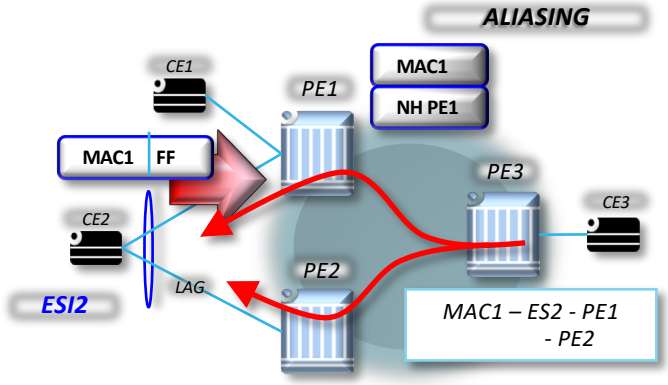
Optimization: Service carving can provide a DF per EVI as opposed to a DF for all the EVIs



Issue: how to prevent CE2 BUM from echoing back to ESI2 (all-active MH)

Solution: Split-horizon label
- PEs advertise an ESI label per ES
- PE1 adds ESI2 label to the MPLS stack
- PE2 does not send any packet with ESI2-label to ESI2

Optimization: Only the non-DF uses the split-horizon label and only for BUM
NOTE: core split-horizon is observed too

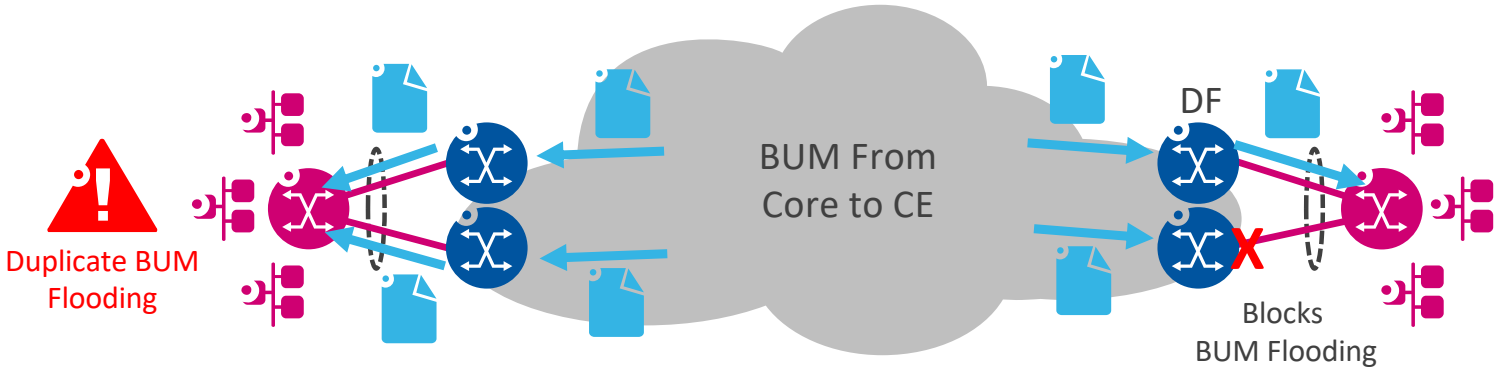


Issue: how to provide unicast flow-based load-balance to PE1/PE2 if MAC1 is only advertised from PE1

Solution: Aliasing
- PE2 will advertise an ESI2 A-D route
- PE3 will add PE2 to the list of NHs for MAC1 since it shares the same ESI as MAC1

EVPN Operation

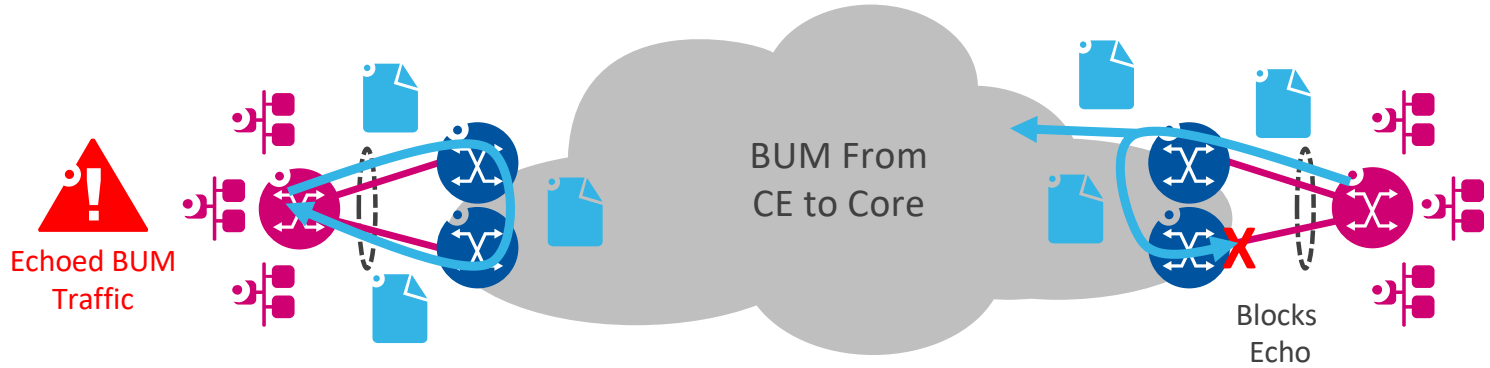
All-Active Multihoming and Designated Forwarder Election



- Avoids duplicate BUM flooding to all-active CEs
- PEs connected to multihomed CEs know about each other through ESI routes
- Elects a designated forwarder (DF) responsible for BUM flooding to the Ethernet segment
- Non-DF PEs block BUM flooding to the CE
- Flexible DF election and functionality
  - Same DF for all ESIs
  - Different DF per ESI
- Unicast still follows all-active paths

EVPN Operation

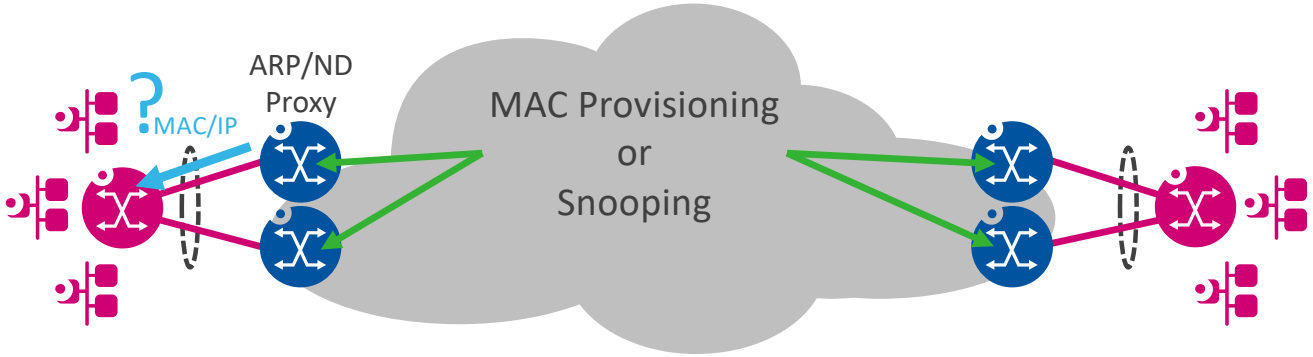
All-Active Multihoming and Split Horizon



- Ensures that BUM traffic from an ESI is not replicated back to the same ESI to an all-active CE
- PE advertises a split horizon label for each all-active Ethernet segment
- When an ingress PE floods BUM traffic, it pushes the split horizon label to identify the source Ethernet segment
- Egress PEs use this label for split horizon filtering and drop packets with the label destined to the Ethernet segment
- Implicit split horizon for core, since PEs won't flood received BUM traffic back into core

EVPN Operation

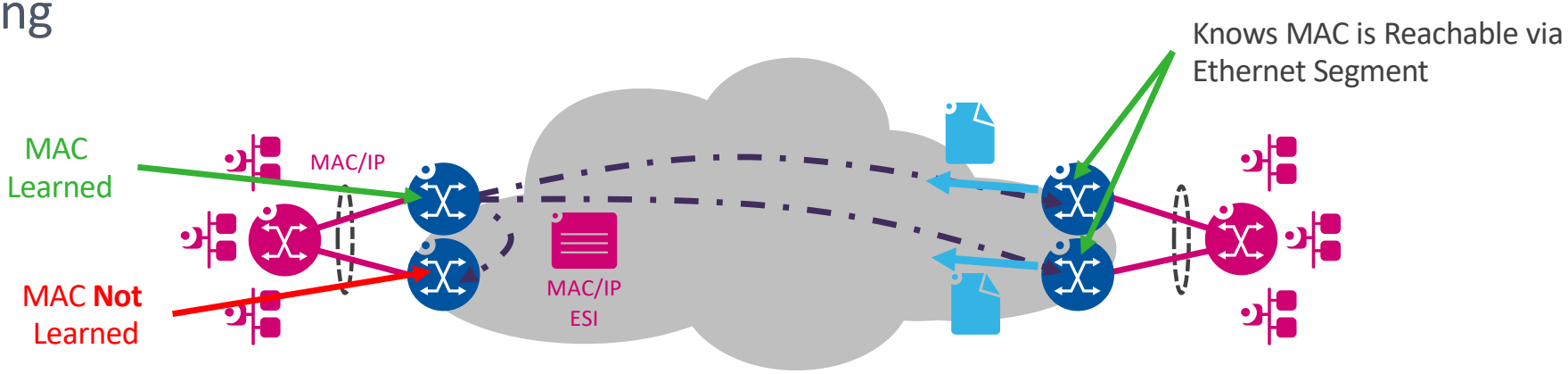
ARP/ND Proxy and Unknown Unicast Flooding Suppression



- ARP/ND is a security issue and a scalability issue in large networks
  - Unknown unicast traffic levels, especially in large data center and IXP networks
- We really don't need it anymore in orchestrated or provisioned networks where all MACs/IPs are known

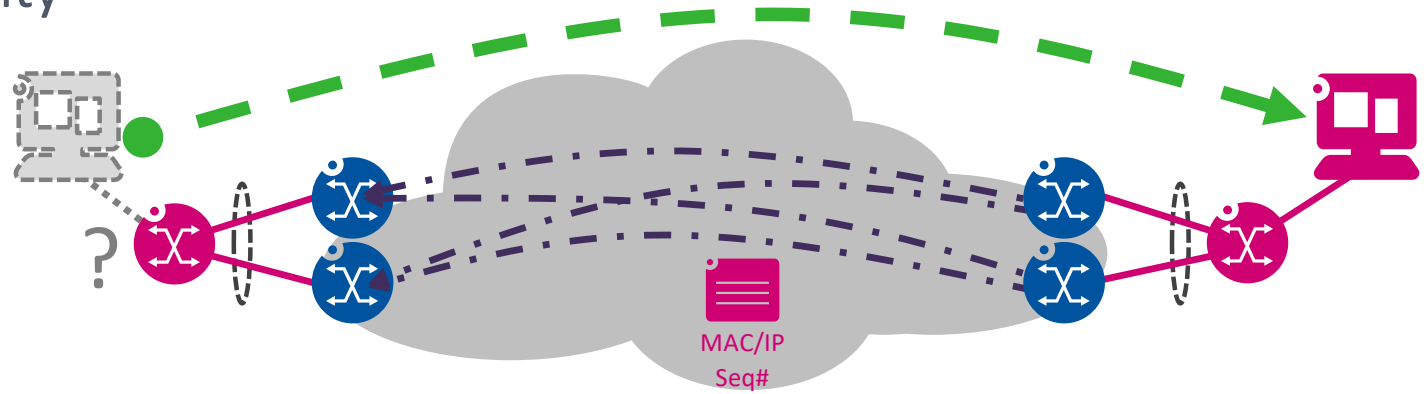
- EVPN can reduce or suppress unknown unicast flooding since all active MACs and IPs are advertised by PEs
  - PEs proxy ARP/ND based on MAC route table to CEs
  - ARP/ND/DHCP snooping optimizes and reduces unknown unicast flooding, useful in dynamic data center networks
  - Provisioning MAC addresses can reduce or eliminate unknown unicast flooding entirely
  - Can disable learning and snooping for programmatic network control

EVPN Operation
Aliasing



- Provides load-balancing to all-active CE when the MAC address is only learned by one PE
  - First MAC learning by PE is usually from a Layer 2 broadcast (ARP/ND/DHCP)
  - Broadcasts are sent on the primary link in a LAG
  - Can have periods of time when the MAC is only learned by the PE connected to the primary link
- PEs advertise the ESI in MAC routes with all-active mode
- Remote PEs can load-balance traffic across all PEs advertising the same ESI
  - Multipathing to CE always works, does not depend on random learning situations or hashing at CE
- Can also be used for a backup path in single-active mode with a standby link

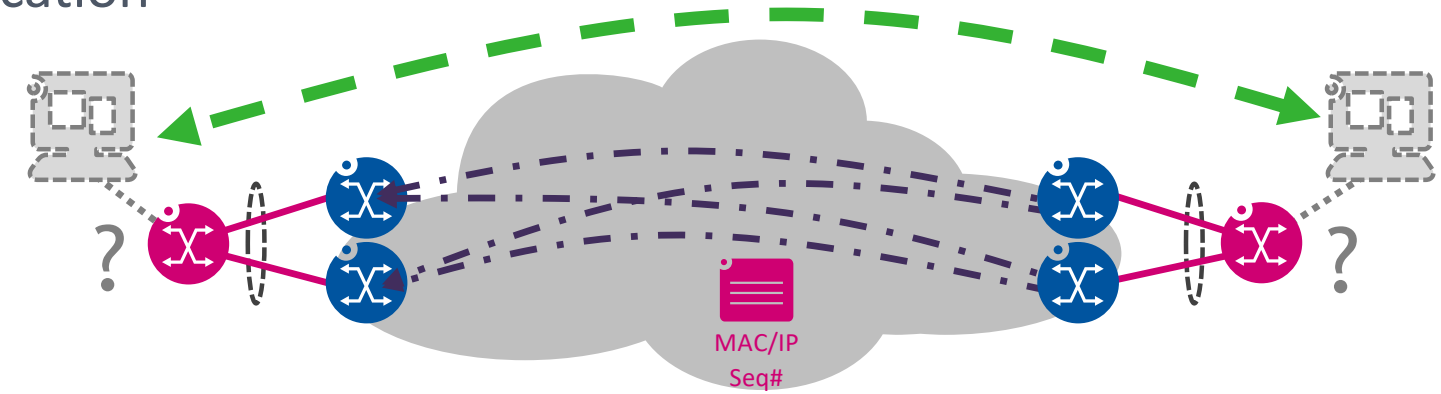
EVPN Operation
MAC Mobility



- MAC addresses may move between ESIs
- If local learning is used, the PE may not detect that a MAC address has moved and won't send a withdraw for it
- New PE sends a new MAC route
- Now there are two routes for the MAC address: an old wrong one and a new correct one

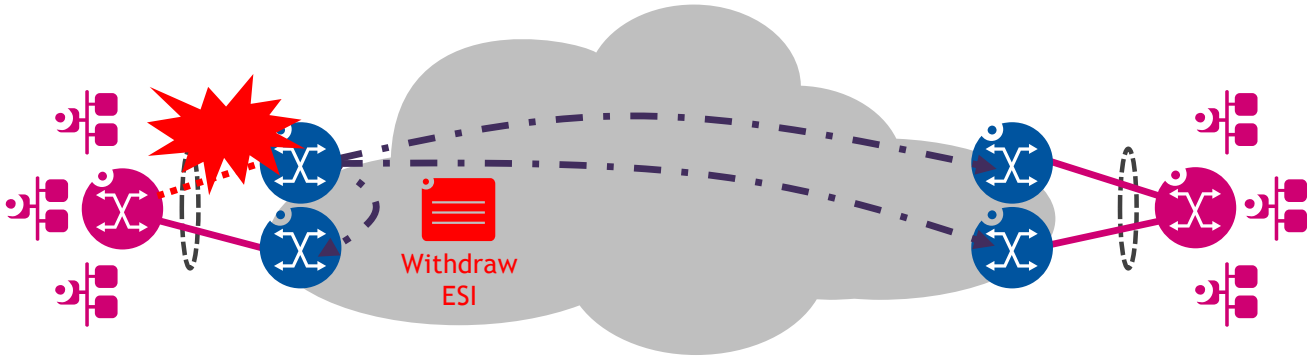
- Each MAC is advertised with a MAC mobility sequence number in an extended community with the MAC route
  - PE selects the MAC route with the highest sequence number
  - Triggers withdraw from PE advertising MAC route with the lower sequence number
  - Lowest PE IP address is used as the tie breaker if the sequence number is the same

EVPN Operation
MAC Duplication



- In certain bad situations, the same MAC could be learned by two PEs
  - MAC duplication
  - Rapid movement
  - Loops
- MAC duplication detection mechanism uses a configurable timer and move counter
  - Provides per-MAC duplication control vs. per-port control in Layer 2 bridging
- If five (*N*) moves (*M*) are detected in 180 s, then the MAC is considered duplicated (default timers)
- PEs stop advertising its route, PEs will use the route with the highest sequence number for forwarding
- Condition can be cleared manually or by implementing a retry timer to clear it automagically

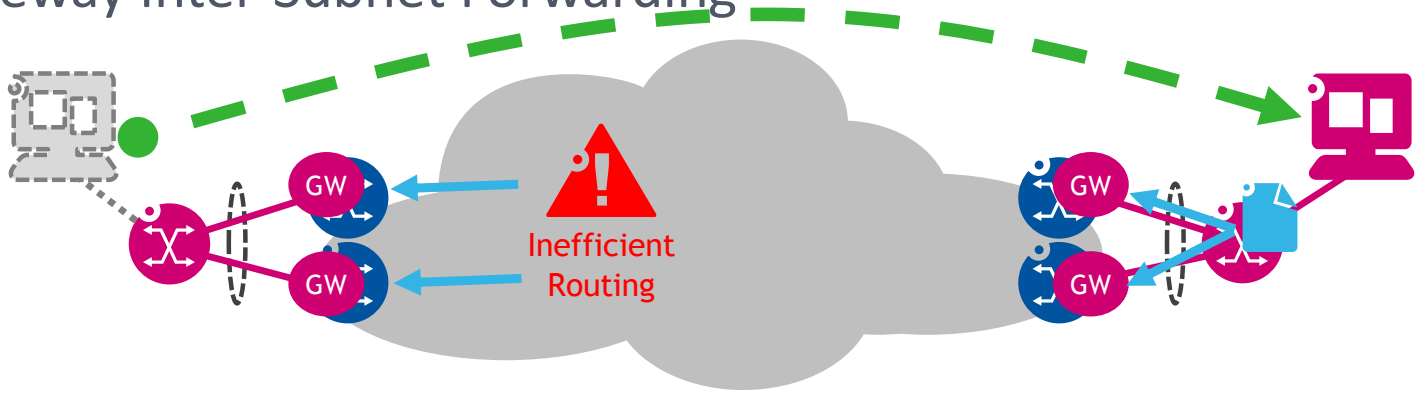
EVPN Operation
MAC Mass-withdraw



- Provides rapid convergence when a link failure affects many MAC addresses
- PEs advertise two routes
  - MAC/IP address and its ESI
  - Connectivity to ESIs
- If a failure affects an ESI, the PE simply withdraws the route for the ESI
- Remote PEs remove failed PE from the path for all MAC addresses associated with an ESI
- Functions as a MAC mass-withdraw and speeds convergence during link failures
- No need to wait for individual MAC addresses to be withdrawn

EVPN Operation

Default Gateway Inter-Subnet Forwarding



- EVPN supports inter-subnet forwarding when IP routing is required
- No additional separate L3VPN functionality is needed, uses EVPN default gateway
- One or more PEs is configured as the default gateway, 0.0.0.0 or :: MAC route is advertised with default gateway extended community
- Local PEs respond to ARP/ND requests for default gateway
- Enables efficient routing at local PE
- Avoids tromboning traffic across remote PEs to be routed after a MAC moves, if all default gateways use the same MAC address

BGP E-VPN routes

Ethernet A-D route

1

Route Distinguisher (8 byte)
Ethernet Segment Id (10 bytes)
Ethernet Tag ID (4 bytes)
MPLS label (3 bytes)

EVPN NLRI encoded in
MP_REACH_NLRI/ MP_UNREACH_NLRI
AFI=25 SAFI=70 (EVPN)

Route Type (1 byte)
Length (1 byte)
Route Type specific (variable)

3 Inclusive Multicast Ethernet Tag Route

Route Distinguisher (8 byte)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP addr (4 or 16 bytes)

PMSI Tunnel Attribute

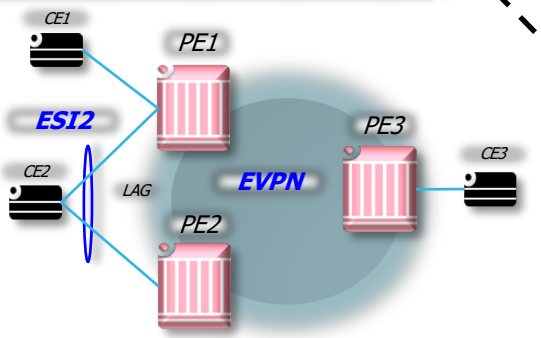
Flags (1 byte)
Tunnel Type (1 byte)
MPLS label (3 bytes)
Tunnel Identifier (variable)

0x06 | Sub-type | Flags (1 byte) | Rsvd = 0
Rsvd = 0 | ESI label

ESI Label Extended community

Second Low order bit defined as Root-leaf bit (1=leaf)

Low order bit of the flags is defined as single-active(0=A-A)



2 MAC Advertisement route

Route Distinguisher (8 byte)
Ethernet Segment Id (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (4 or 16 bytes)
MPLS label (3 bytes)

Ethernet Segment Route

4

Route Distinguisher (8 byte)
Ethernet Segment Id (10 bytes)

0x06 | Sub-type | ES-Import
ES-Import Cont'd

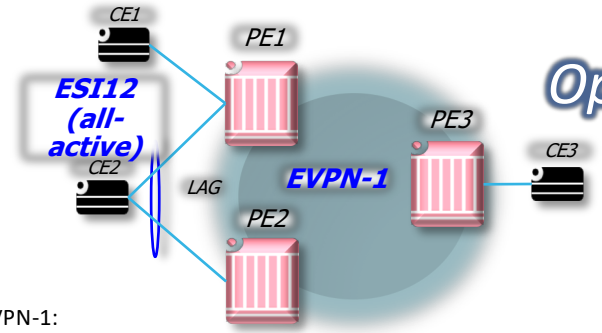
ES-Import route target

EVPN routes	PE1-PE2	PE3	Usage
ES	*Export/import*	*N/A*	*ES discovery and DF election*
A-D per ESI	*Export/import*	*Import*	*Mass-withdraw and ESI labels*
A-D per EVI	*Export/import*	*Import*	*Aliasing advertisement*
MAC	*Export/import*	*Export/import*	*MAC/IP advertisement*
Inc. mcast	*Export/import*	*Export/import*	*Flooding tree setup*

MAC Mobility Extended community

0x06 | Sub-type | Rsvd = 0
Sequence Number

E-VPN usage example



E-VPN-1:
- All-active MH for CE2
- Ingress replication
- VLAN-based service interfaces
- MAC forwarding model, label per EVI
- Data plane learning from CEs

Operational Procedures

Network startup procedures

- Infrastructure setup
- ES discovery and DF election

Service startup procedures

- Service provisioning
- Flooding tree setup
- Split-Horizon setup
- Fast convergence setup
- Aliasing setup

Functional State Procedures

- BUM packet from CE1
- BUM packet from CE2
- Unicast packet from CE3 to CE1
- Unicast packet from CE3 to CE2
- CE2-PE2 link failure

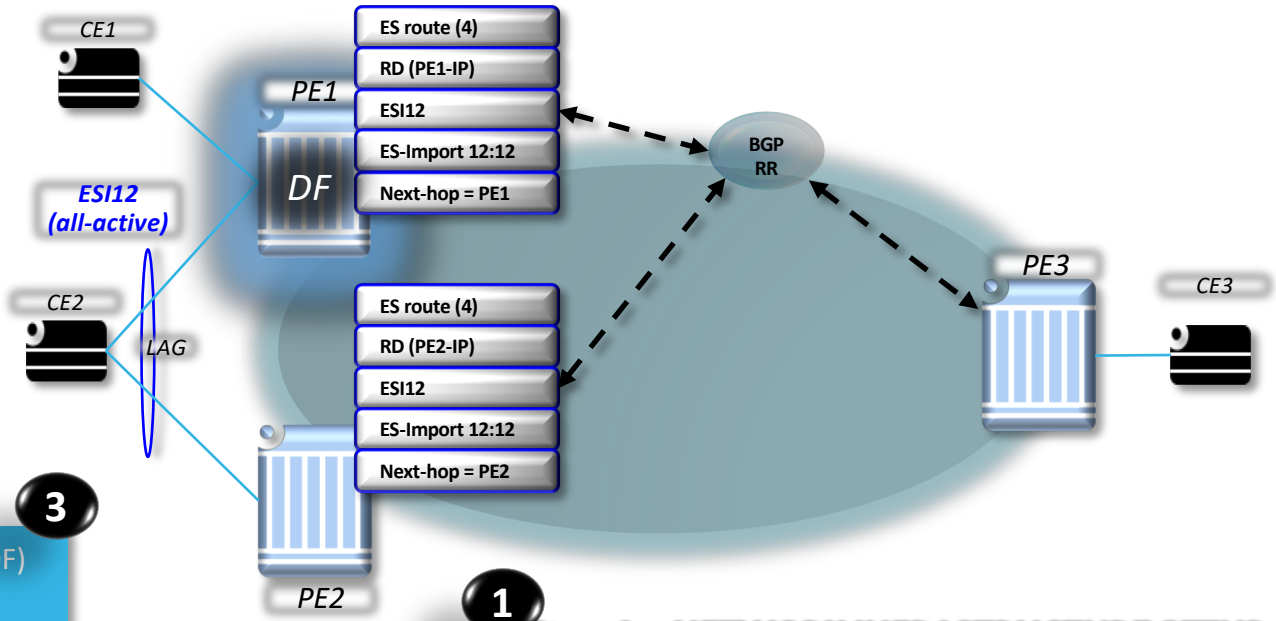
E-VPN network startup procedures

…common for all the E-VPN services

Step-2 – ES AUTO-DISCOVERY

2

- PEs with unreserved ESIs generate one ES route (type 4) per ESI:
  - .- Auto-derived RD (system IP)
  - .- Auto-derived/provisioned ESI
  - .- Auto-derived ES-Import ext-comm (MAC portion of the ESI)
- PE1 and PE2 import the route
- PE3 discard the route



Step-3 – DF ELECTION

3

- PE1 and PE2 elect a Designated Forwarder (DF) per <ESI, EVI>, e.g. PE1
  - .- DF for all EVIs: highest IP, the next-highest, etc.
  - .- Service-carving: DF per EVI
- DF is only considered for BUM and only in the PE1→CE2 direction

1

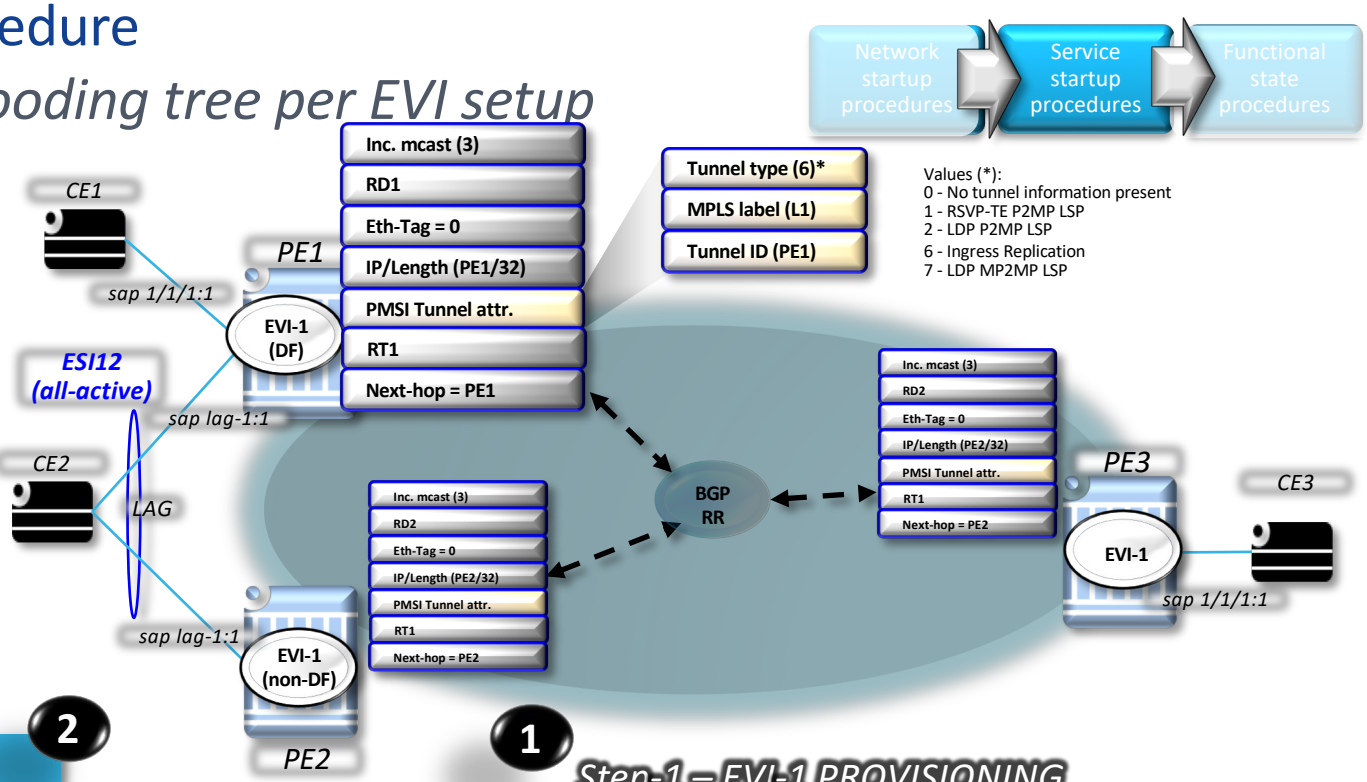
Step-1 – NETWORK INFRASTRUCTURE SETUP

- LDP or RSVP setup for p2p tunnels among the three PEs
- P2MP mLDP or P2MP RSVP capabilities enabled for the creation of flooding trees
- MP2MP mLDP possible too (not possible in VPLS)

NOKIA

E-VPN service startup procedure

Service provisioning and flooding tree per EVI setup



Step-3 – EVI Flooding Tree setup

Based on the received Incl. mcast routes, each PE establishes its flooding tree that will be used for BUM traffic

3

Step-2 – INCLUSIVE MCAST ROUTES EXCHANGE

2

PEs exchange Inclusive Multicast routes (per <EVI, Eth-Tag>) that include PMSI tunnel attributes:

- Tunnel type and ID
- For ingress replication the attribute carries an MPLS downstream allocated label

1

Step-1 – EVI-1 PROVISIONING

EVI-1 parameters to be provisioned on each PE:

- EVI RD and RT
- CE-VID (1) and port/lag binding to EVI (1)
- CE-VID to Ethernet Tag binding for VLAN-aware service interfaces

E-VPN service startup procedure

A-D routes per ESI for split-horizon and fast convergence procedures

Step-6 – FAST CONVERGENCE SETUP FOR ESI12

6

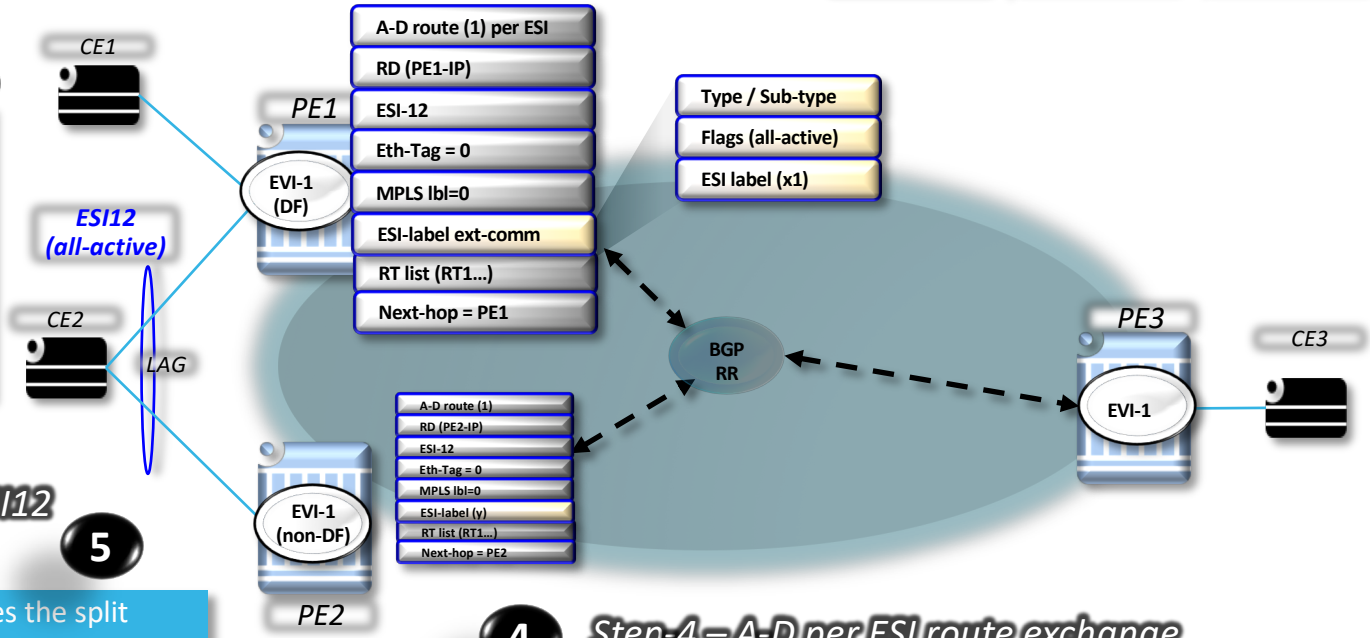
- PE3 will install the PE1 (A-D route) and PE2 (A-D route) for ESI12
- MAC addresses from PE1/PE2 will only be valid as long as the A-D route is active
- If CE2-PE1 fails, PE1 will withdraw its ESI12 A-D route and PE3 will provide a uniform failover for all the MACs from PE1

Step-5 – SPLIT-HORIZON SETUP FOR ESI12

5

The ESI-label (identifies the source ESI) enables the split horizon mechanism so that BUM packets sent by CE2 to the non-DF, are not be looped back

- Ingress replication: ESI-label x will be used by PE2 when sending traffic to PE1 (downstream allocation of "x")
- P2MP LSP: ESI-label y will be used by PE2 when sending traffic to the P2MP LSP (upstream allocation)



Step-4 – A-D per ESI route exchange

4

- PE1 & PE2 send an A-D route for ESI12
- PE3 does not since it does not have unreserved ESIs
- All the PEs import the received routes based on the RT (RT1)

NOKIA

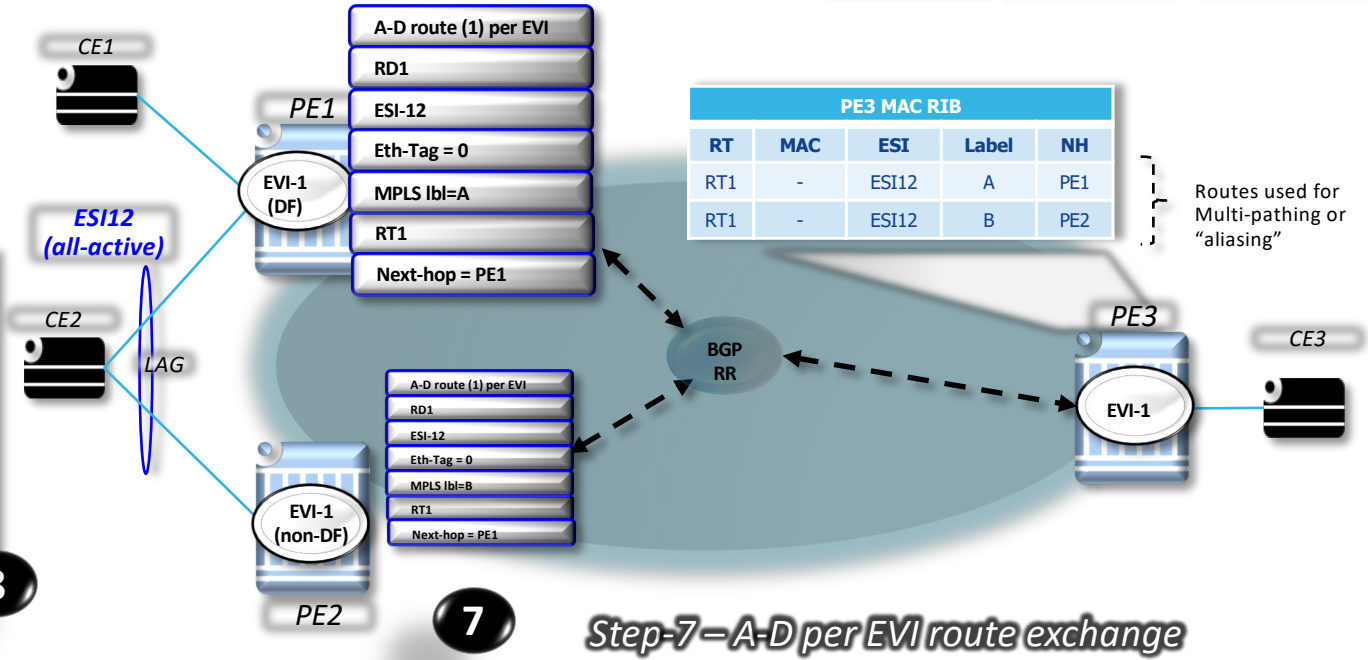
E-VPN service startup procedure

A-D routes per <ESI, EVI> for aliasing procedures

Step-8 – ALIASING SETUP FOR EVI1, ESI12

- PE3 will install the A-D routes per EVI for aliasing functions
- Any flow destined to a MAC learnt from ESI12 can be load-balanced to PE1 or PE2 as long as the A-D route for EVI1 from PE1 or PE2 is active
- The load-balancing will be based on an ingress hashing algorithm at PE3

8



7

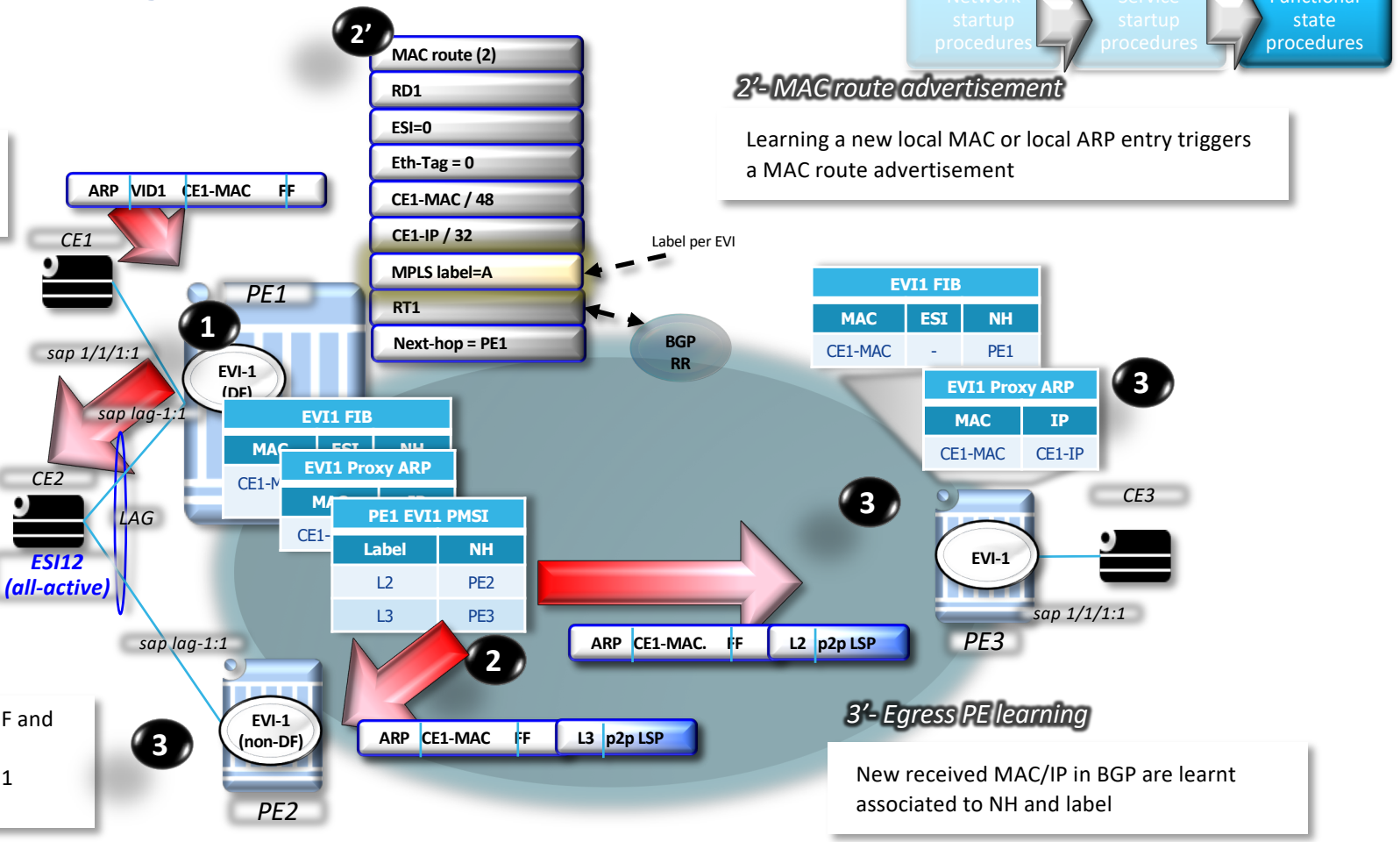
Step-7 – A-D per EVI route exchange

- PE1 & PE2 send an A-D route for <ESI12, EVI1>
  - For VLAN-aware services, PE1&2 would send an A-D route per <ESI, Eth-Tag> per EVI
- PE3 does not since it does not have unreserved ESIs
- All the PEs import the received routes based on the RT (RT1)

NOKIA

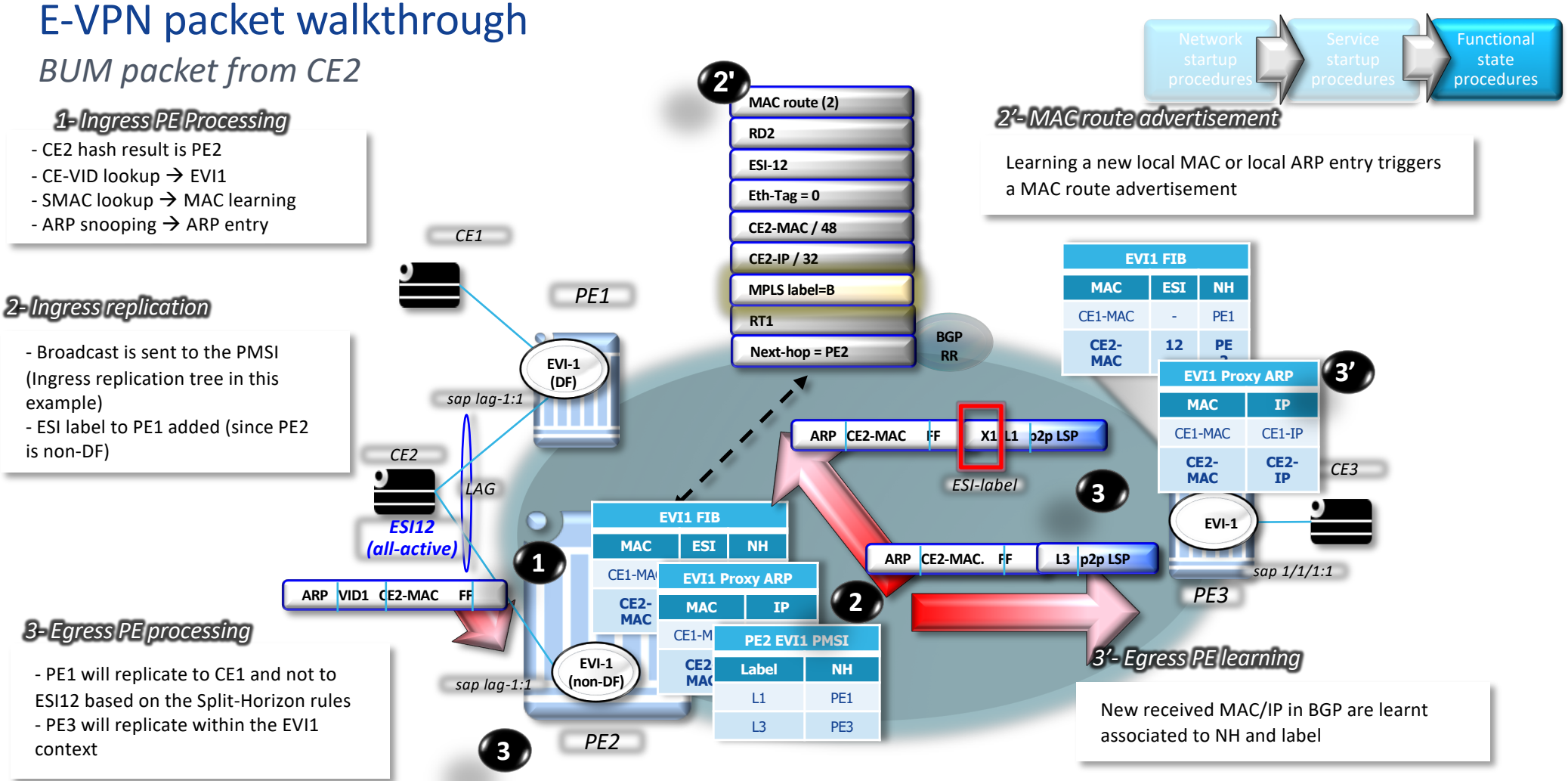
E-VPN packet walkthrough

BUM packet from CE1



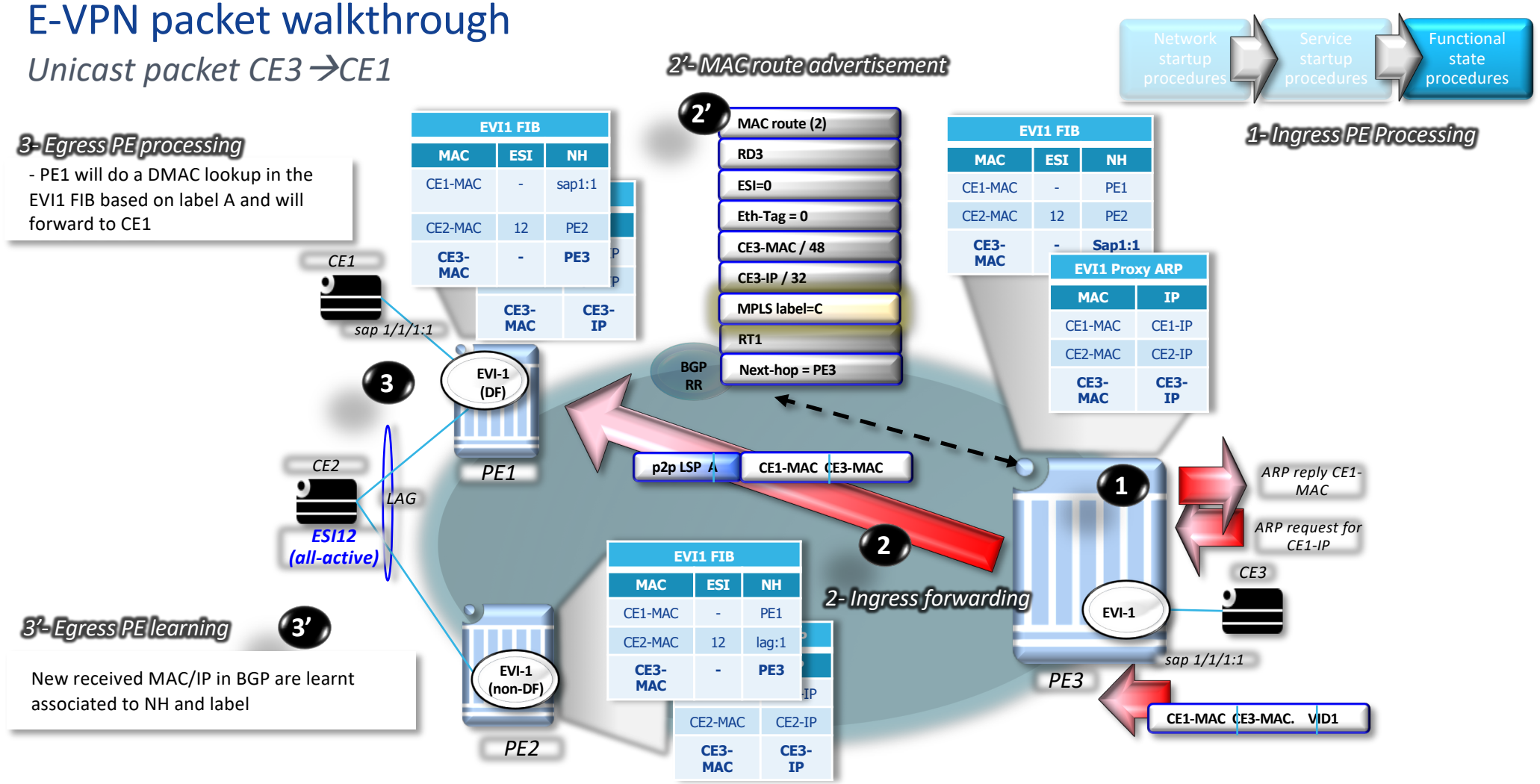
E-VPN packet walkthrough

BUM packet from CE2



E-VPN packet walkthrough

Unicast packet CE3 → CE1



E-VPN packet walkthrough

Unicast packet CE3 → CE2

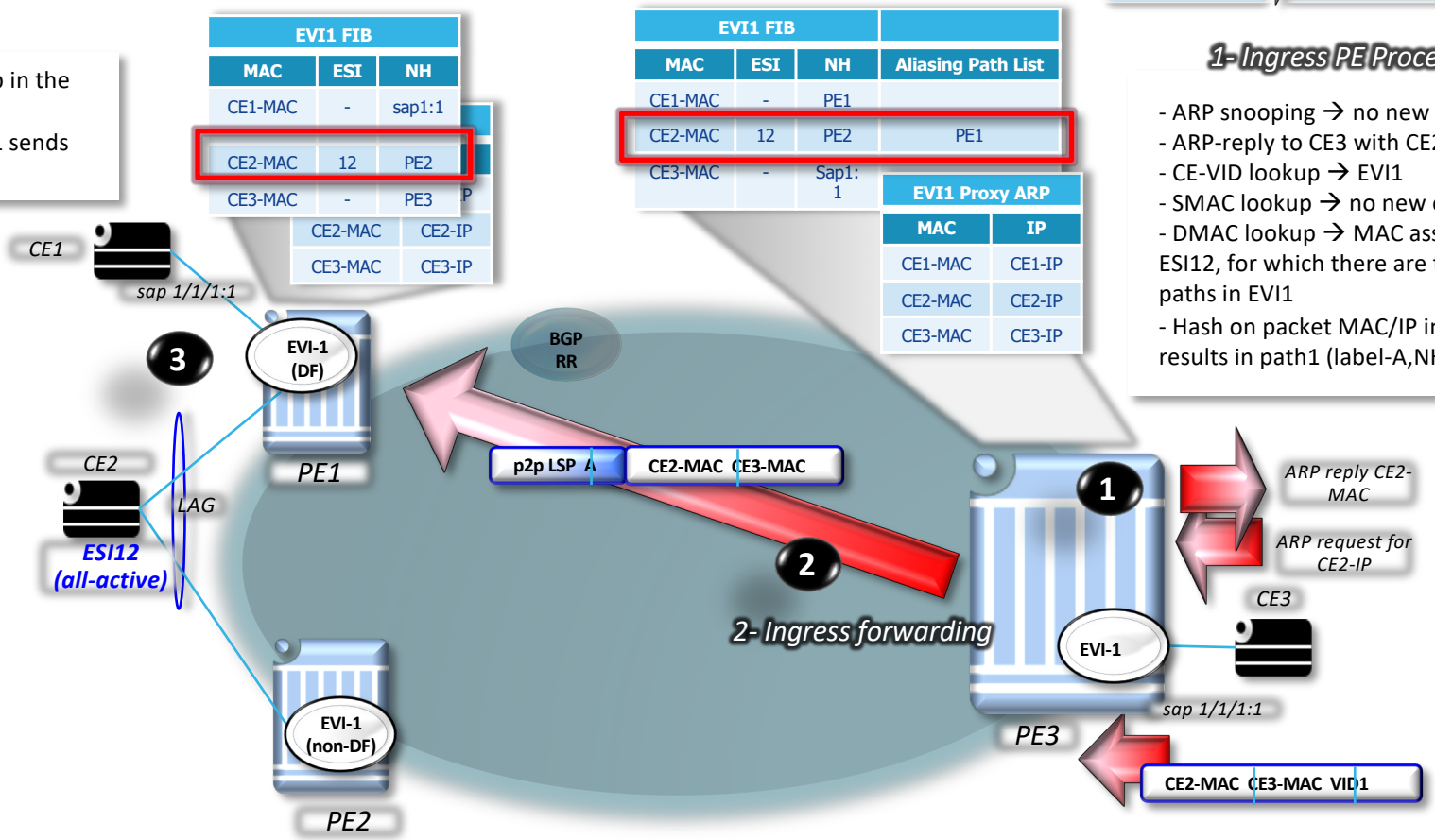
Network startup procedures → Service startup procedures → Functional state procedures

3- Egress PE processing

- PE1 will do a DMAC lookup in the EVI1 FIB based on label A
- Since ESI12 is local ESI, PE1 sends the packet directly to CE2

1- Ingress PE Processing

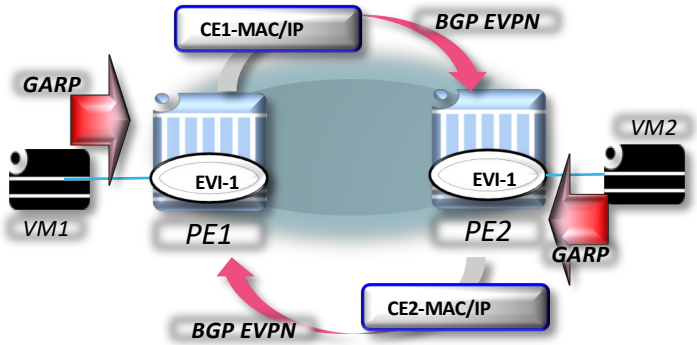
- ARP snooping → no new entry
- ARP-reply to CE3 with CE2-MAC
- CE-VID lookup → EVI1
- SMAC lookup → no new entry
- DMAC lookup → MAC associated to ESI12, for which there are two A-D active paths in EVI1
- Hash on packet MAC/IP information results in path1 (label-A,NH-PE1)



NOKIA

E-VPN traffic flow optimization procedures

Flooding suppression, inter-subnet forwarding and mac-mobility



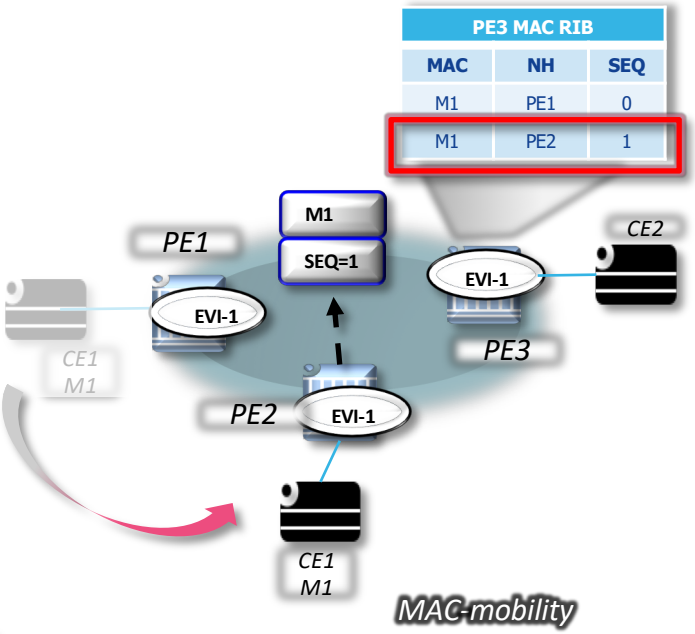
Inter-subnet forwarding

- EVPN and IRB to a local VRF allow for routing to hosts in other subnets

- Only EVPN (advertising MACs/IPs) is required, no IP-VPN is needed

- EVPN updates populate VRF routing table

- IRB interfaces advertised in EVPN with default GW ext-comm



MAC-mobility

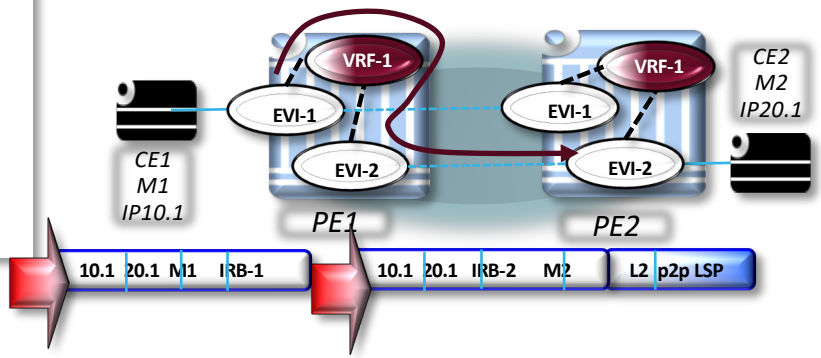
- Each MAC is advertised along with a mac-mobility ext-comm, including a sequence number

- A PE selects the MAC route with the highest SEQ. number

Unknown unicast flooding suppression

- EVPN allows flooding suppression in certain networks (DC)

- Assuming VMs always signal its presence (GARP/DHCP) unknown flooding can be suppressed since all the active MACs and IPs will always be distributed by EVPN



Agenda

EVPN Background and Motivation

In a nutshell

EVPN Operations

Data Planes

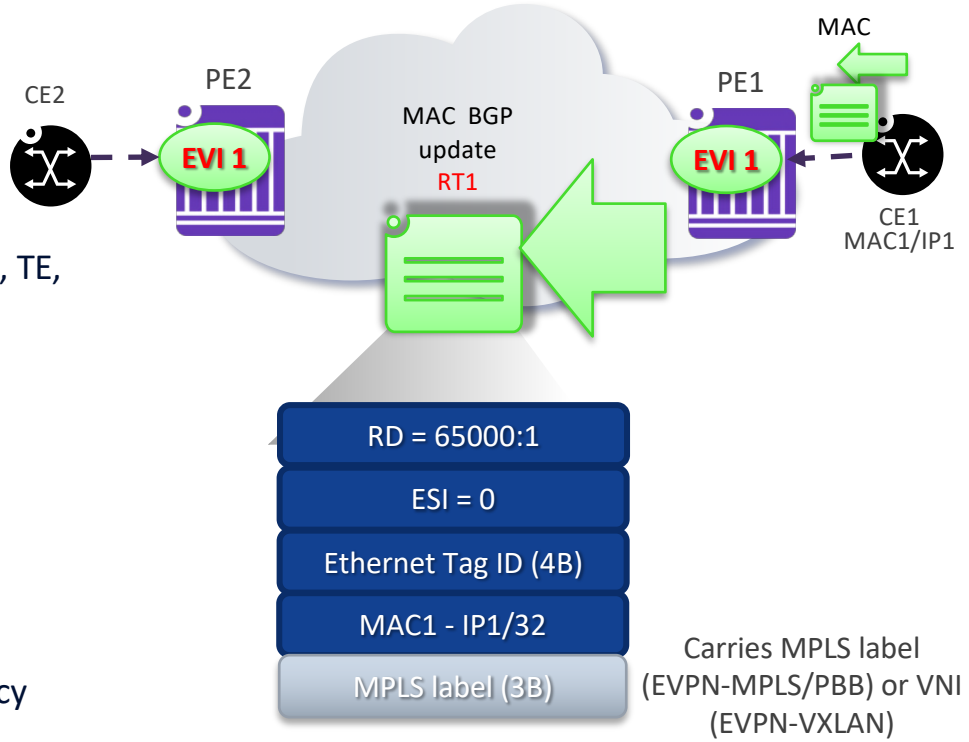
EVPN Use Cases/Applications

Protocol details

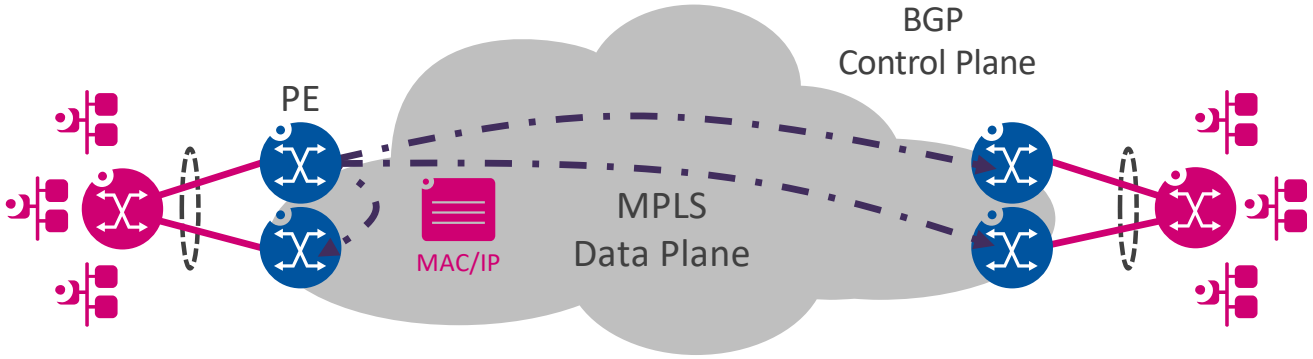
EVPN abstracts the control plane

To support current and future data plane encapsulations

- EVPN-MPLS /EVPN-PBB
  - Uses a 20-bit MPLS service label as MAC-VRF de-multiplexer
  - Transport tunnel: RSVP/LDP/BGP/SR-ISIS/SR-OSPF
  - Takes advantage of all the underlying MPLS capabilities (FRR, TE, etc.)
- EVPN-VXLAN
  - Typically used in Datacenter (Nuage)
  - Uses a 24-bit VNI as MAC-VRF de-multiplexer
  - The VNI is encapsulated as part of the VXLAN header
  - VXLAN is transported over UDP/IP
  - Takes full advantage of the VXLAN simplicity and transparency



EVPN Multiprotocol Label Switching (MPLS) Data Plane
RFC7432 (EVPN-MPLS)



- EVPN over an MPLS data plane is the original EVPN solution in the base specification
- Requires IGP, RSVP-TE or LDP, BGP
- No pseudowires
- MPLS runs in the core network's control plane and data plane
- Core network supports all the MPLS features we know and love, since EVPN uses MPLS as the data plane (TE, FRR, …)

EVPN for MPLS Tunnels in ELAN Services

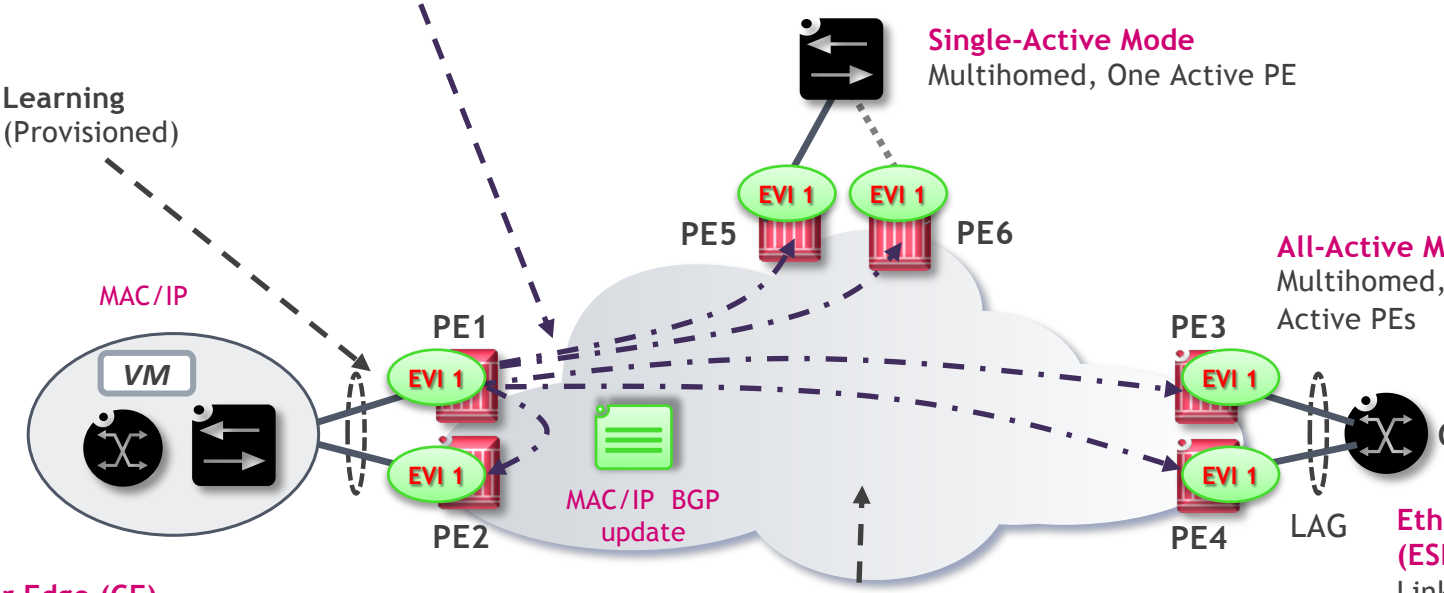
EVPN Instance (EVI)
Identifies a VPN

MAC-VRF
Virtual Routing Forwarding table for MACs

Ethernet Tag
Broadcast or Bridge Domain in the EVI

Control Plane Learning
PEs Advertise MAC Addresses and Next Hops From Connected CEs Using MP-BGP

Data/Mgmt Plane Learning
Dynamic or Static (Provisioned)



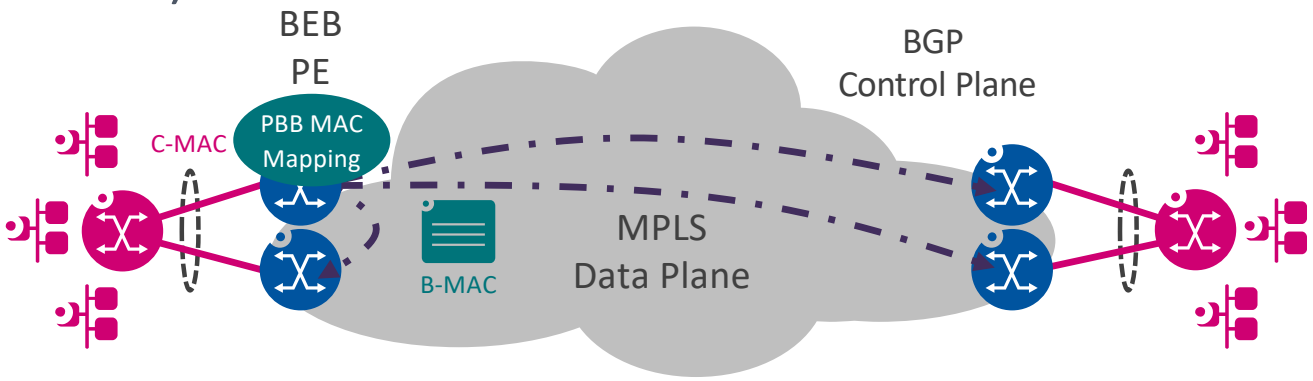
Customer Edge (CE)
Host, VM, Router or Switch

Data Plane Encapsulation
MPLS tunnel : RSVP-TE / LDP / SR

Ethernet Segment Identifier (ESI)
Link(s) that Connect the CE to PEs (ESIs are Unique Across the Network)

Provider Backbone Bridges (PBB) EVPN Data Plane

RFC7623 (PBB-EVPN)

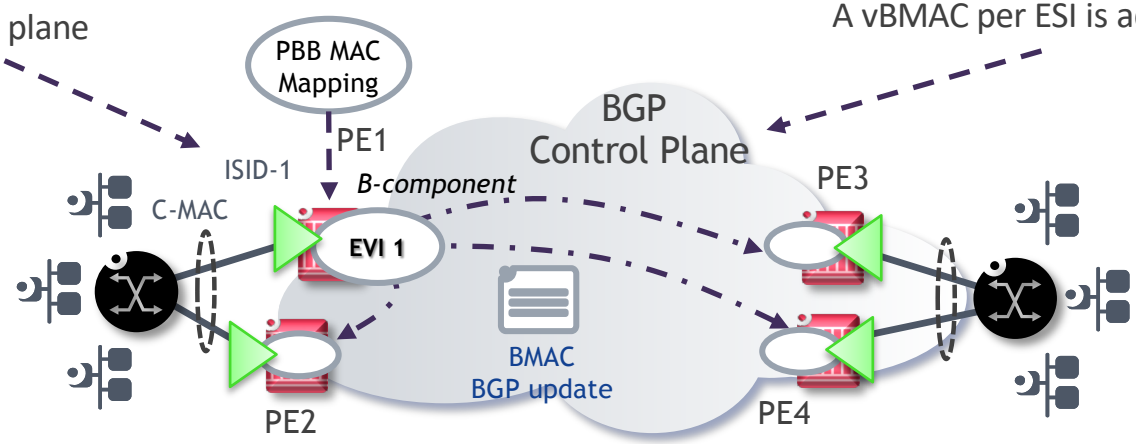


- PBB-EVPN combines IEEE 802.1ah PBB with EVPN
- PEs are PBB Backbone Edge Bridges (BEB)
- Reduces number of MACs in EVPN by aggregating customer MACs with backbone MACs
  - Same concept as route aggregation in IP
- Scales EVPN networks to a very large number of MACs
  - PEs only advertise backbone MACs with BGP
  - Customer MAC and backbone MAC mapping is learned in the data plane
  - Useful for providing services to networks where the MACs are not under your control
- MPLS runs in the control plane and data plane

EVPN for PBB for large L2 networks (RFC 7623)

CMAC Data plane learning
CMAC-local AC or CMAC to remote BMAC mapping is learnt in the data plane

BMAC Control plane learning
System BMACs are advertised by MP-BGP
A vBMAC per ESI is advertised by the active PE

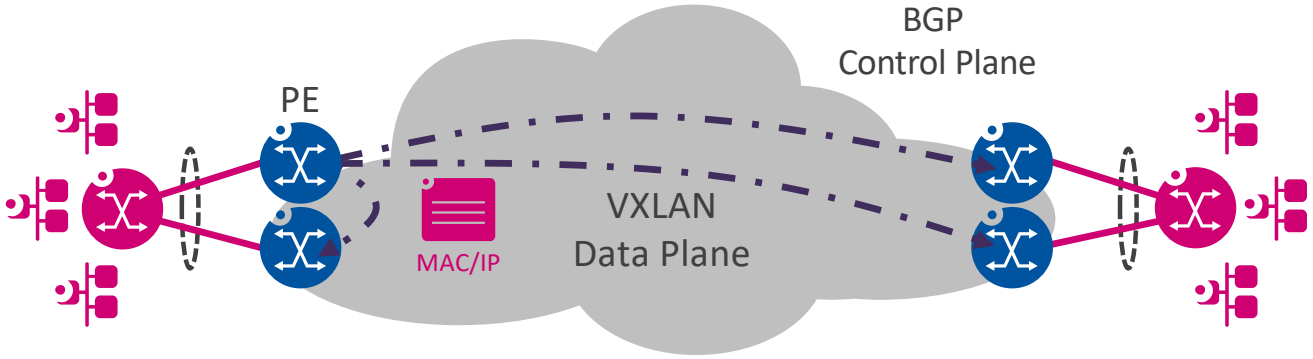


PBB-EVPN combines 802.1ah and EVPN
- PEs have I-components mapped to B-components (EVIs)
- BGP-EVPN is enabled in the B-VPLS domain
- Reduces the number of MACs in EVPN by aggregating CMACs with BMACs

Used to scale very large layer-2 EVPN networks
- Per-ISID flooding trees are supported
- The B-component EVI uses MPLS data plane

EVPN Virtual Extensible LAN (VXLAN) Data Plane

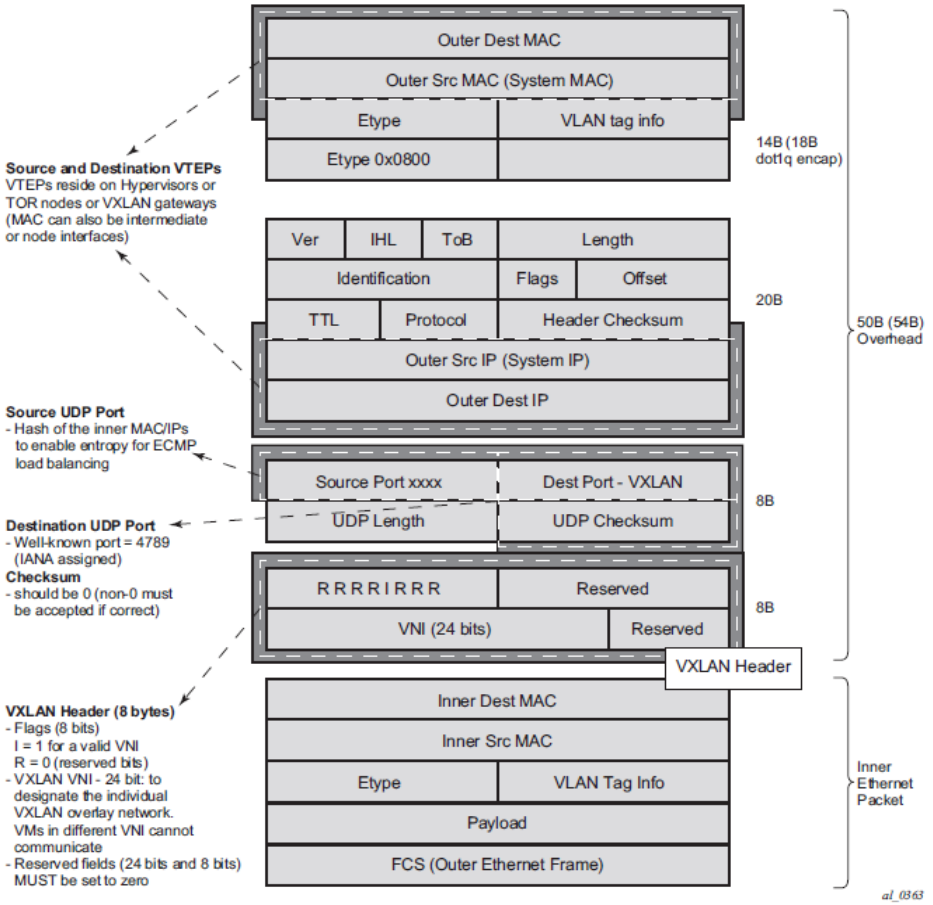
RFC8365 (EVPN-VXLAN)



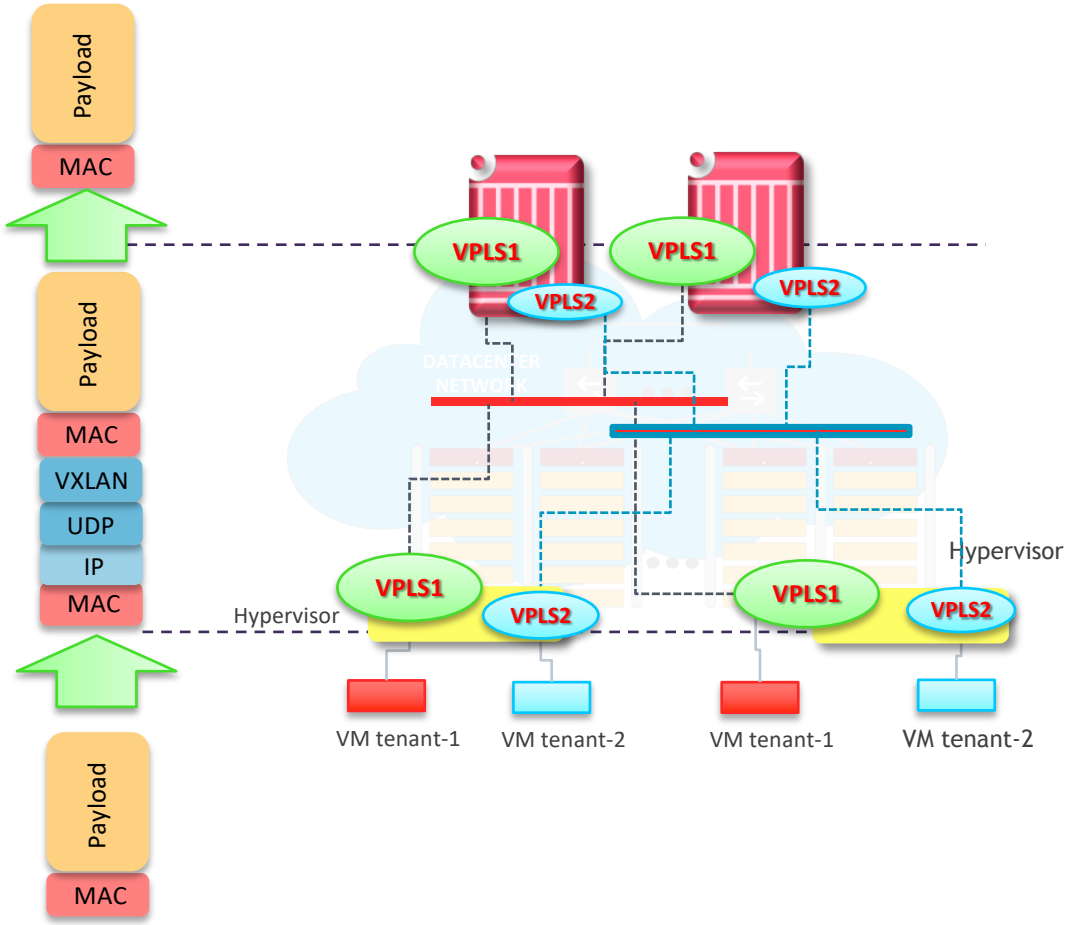
- EVPN-VXLAN uses EVPN over a VXLAN data plane
  - VXLAN is typically used for data center extension over WAN
  - Can also be used as an overlay in any IP network for IP/Ethernet services
  - Useful when MPLS is unavailable or unwanted
  - Alternative to NVGRE or MPLSoGRE (NVO3)
  - PIM is not needed with ingress BUM replication
- VXLAN provides the Layer 2 overlay over IP
  - IP reachability is required between PEs
  - EVPN uses BGP control plane for MAC route advertisements
  - VXLAN data plane uses UDP to encapsulate the VXLAN header and Layer 2 frame
- Provides all the benefits of EVPN for DCI and virtualized networks

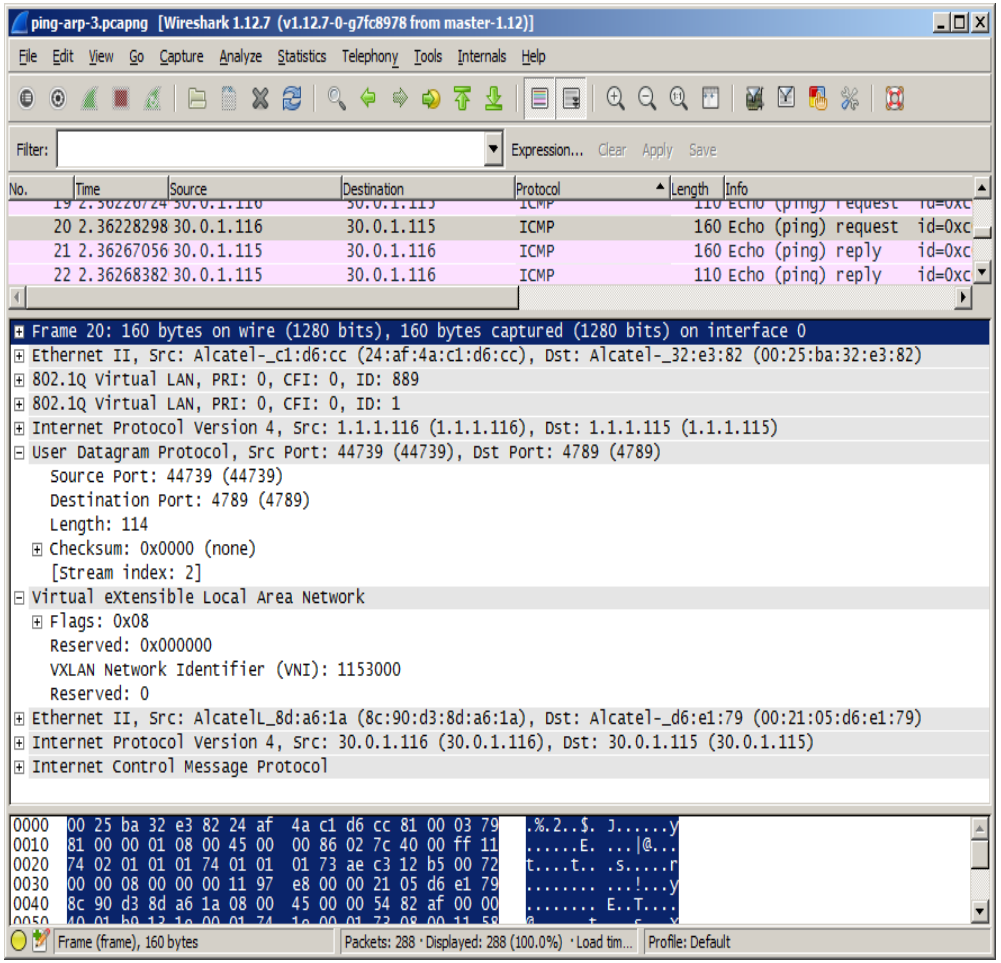
VXLAN (Virtual eXtended LAN): The DATA PLANE standard for DC IP fabrics

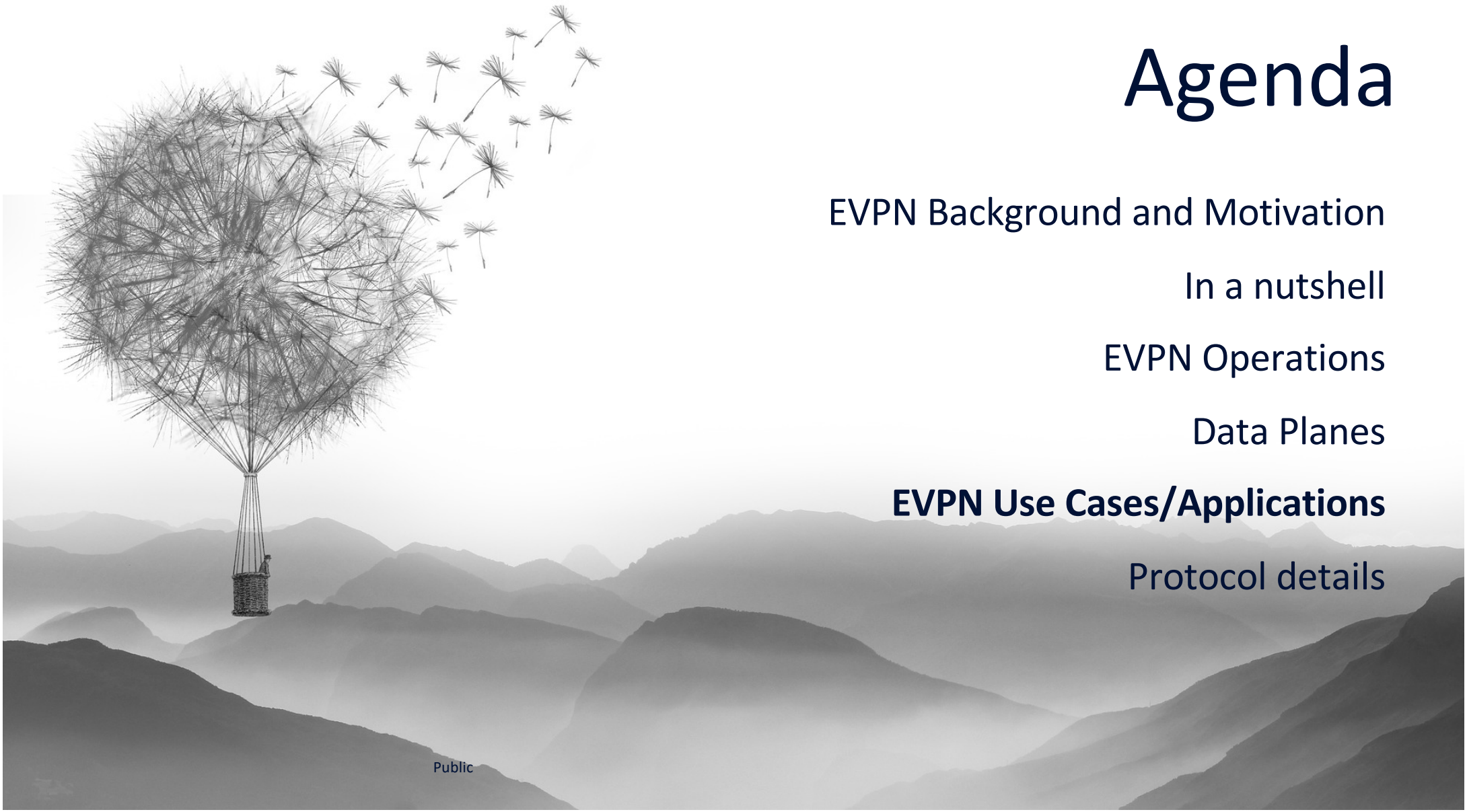
- VXLAN encapsulates Ethernet in IP over a logic L3 tunnel
  - Runs over IPv4 (or IPv6)
  - Uses UDP, source port is a hash of MAC or IPs to provide load balancing entropy
  - 8 byte VXLAN header provides 24 bit VXLAN Network Identifier (VNI) and flags
  - Total encapsulation overhead is ~50 bytes
- VXLAN is routable with IP, so the underlay network may be any network that uses existing resiliency and load balancing mechanisms
  - ECMP
  - IGPs/BGP
  - IP FRR
- VXLAN tunnel endpoints can be on network equipment or computing infrastructure
  - Deliver a VPN straight to a hypervisor



VXLAN Overlay Tunnels Frame Format







Agenda

EVPN Background and Motivation

In a nutshell

EVPN Operations

Data Planes

EVPN Use Cases/Applications

Protocol details

Where can we find Ethernet VPNs?

Carrier Ethernet

Cloud integration

IXP

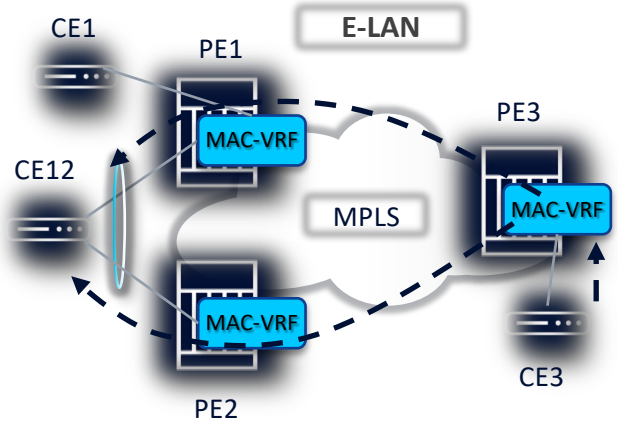
SD-WAN

5G

NOKIA

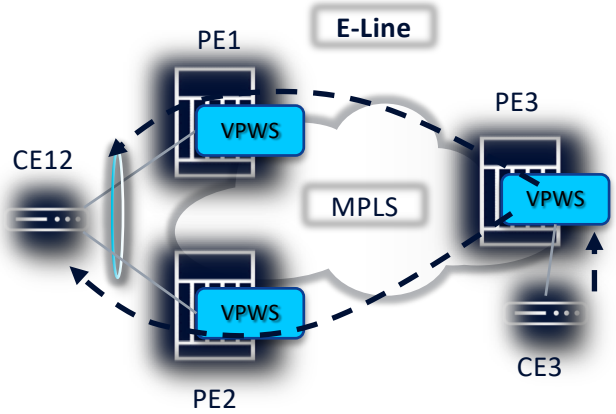
EVPN for Carrier Ethernet

Driven by multi-homing, simplification and advanced features



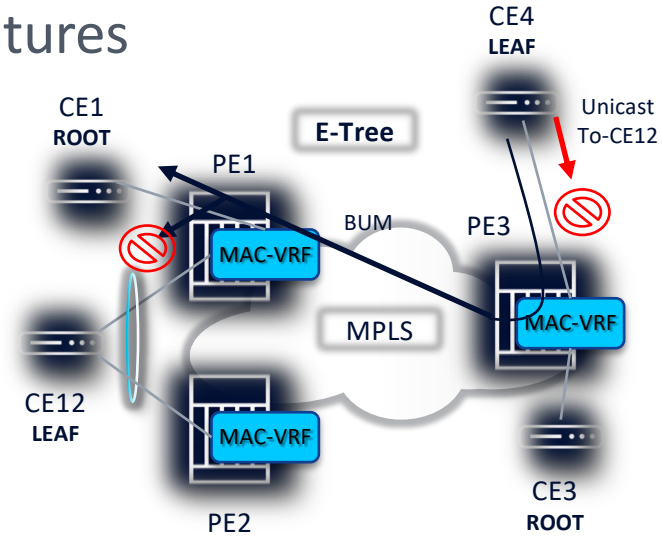
(PBB)EVPN-MPLS (RFC7432/7623)

- Uniform control plane
- A/A, A/S Multi-homing
- Load-balancing
- BUM reduction/suppression
- Security and loop protection
- Inter-As option-B/C for massive scale
- PBB: control plane simplification / MAC reduction



EVPN-VPWS (RFC8214)

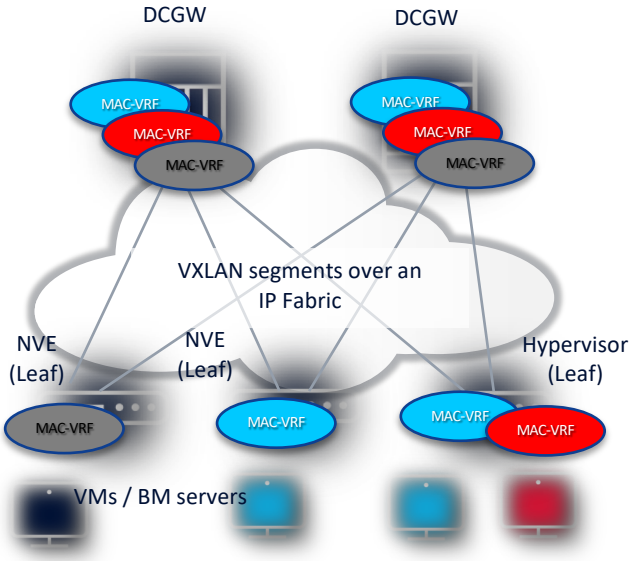
- Uniform control plane
- A/A, A/S Multi-homing

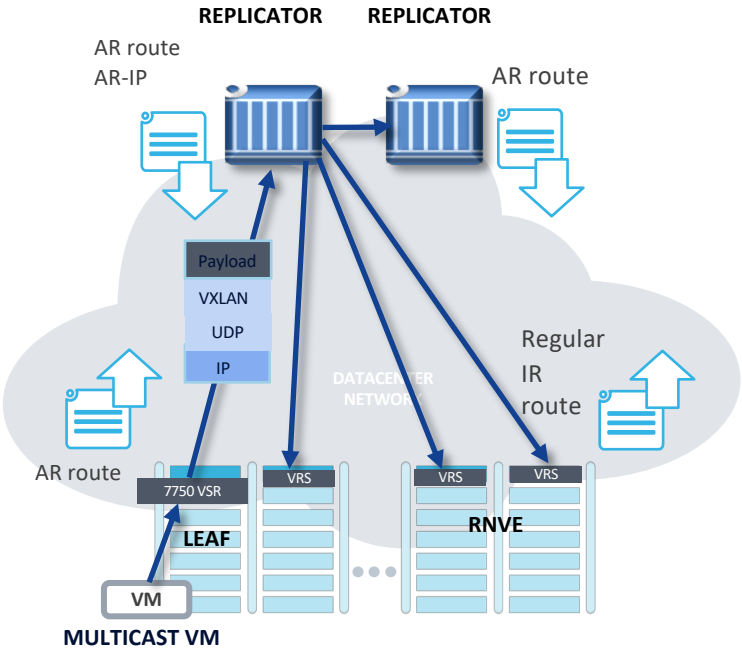


(PBB)EVPN-E-Tree (RFC8317)

- Leaf-to-leaf Unicast traffic filtered at ingress
- Eefficiency, simplification and multi-homing
- PBB: control plane simplification / MAC reduction

EVPN and Cloud Integration





EVPN VXLAN is the de-facto standard for Cloud and DCI

- auto-discovery of remote VTEPs via IMEA route
- Distribution of MAC/IP

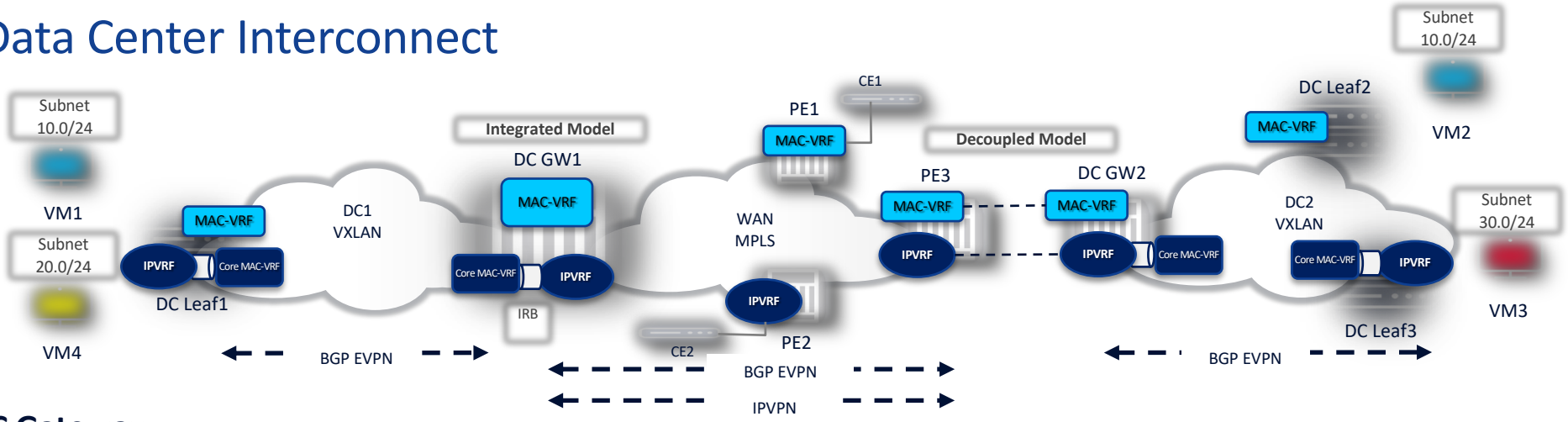
Advance options:

- Distribution of IP-Prefixes for inter-subnet forwrding
- MAC protection, duplication detection, loop protection
- Proxy-ARP/ND
- Service-chaining using MAC/IP routes, ESI

Multicast efficiency

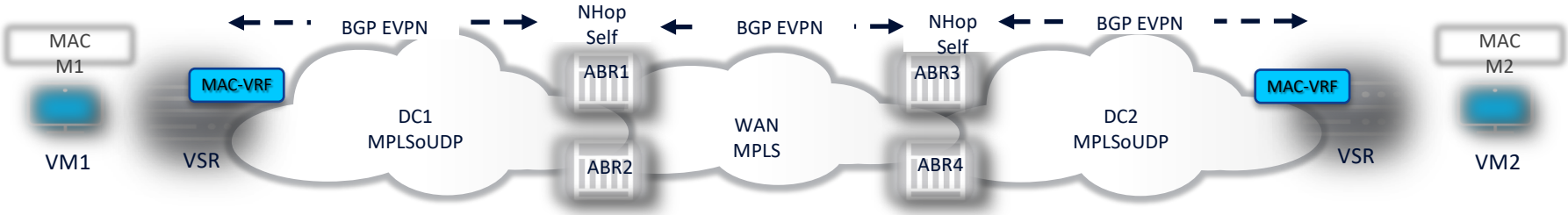
- Ingress replication (may be a challenge on VM for multicast)
- Assisted-Replication (draft-ietf-bess-evpn-optimized-ir)
  - Help software-based PE and NVEs with low-performance replication capabilities

Data Center Interconnect



DC Gateway

- Translation between VXLAN and EVPN-VPLS, IP prefixes and IPVPN
- Demarcation: QoS, Security, aggregation, isolation



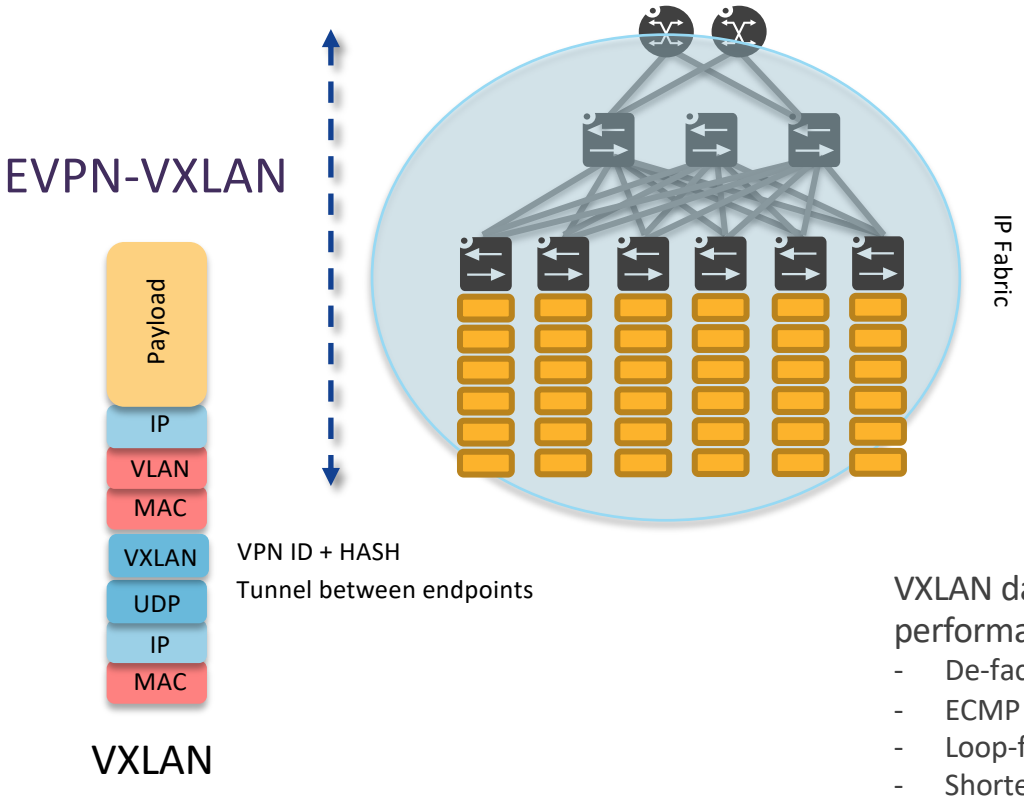
DC Gateway-less

- Inter-AS opt B or NHS RR and MPLSoUDP (NFIX), Inter/Intra-As opt C
- ASBR/ABR terminate DC tunnel, swap service label and push MPLS stack for WAN tunnel (and vice versa)
- Scalable DCI without service instantiation on the DC GW

Data Center use-case

Cloud computing and NFV are shifting DC networks to SDN-based DCs where only VXLAN and EVPN provide the required capabilities

- Legacy DC networks can't cope with 10,000s of dynamic hosts/VMs



Required EVPN features

- EVPN provides L2/L3 connectivity for 1,000s of tenants in the DC
- The IP fabric can also be extended to the WAN for DC interconnect
- MAC mobility, proxy-ARP/ND, MAC protection, unknown flooding suppression, inter-subnet forwarding

VXLAN data plane provides the required scalability, performance and simplicity

- De-facto standard with assisted hardware in servers
- ECMP and fast resiliency
- Loop-free forwarding for L2
- Shortest path between any 2 endpoints

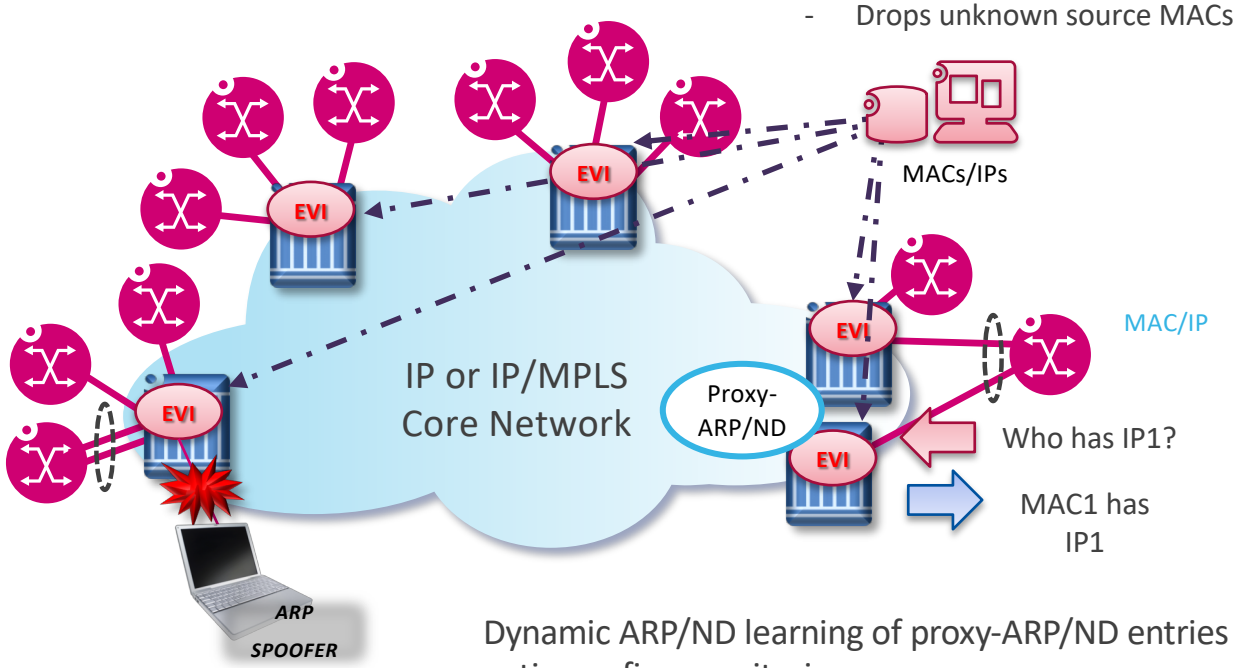
Ethernet VPN (EVPN)

Peering Fabric Architecture for IXPs

- Layer 2 protocols and VPLS are popular architecture choices for IXP peering fabrics
- EVPN is an emerging technology that enables a new architecture for IXP peering fabrics with some advantages
  - Multihoming to different PEs
  - Manage and program all IP/MACs with local proxy from PEs
  - Simpler topology with VXLAN instead of MPLS with less protocols to manage
- Let’s take a look at the EVPN technology and how it can be used as an IXP peering fabric architecture

Internet Exchange Points
Peering Fabric

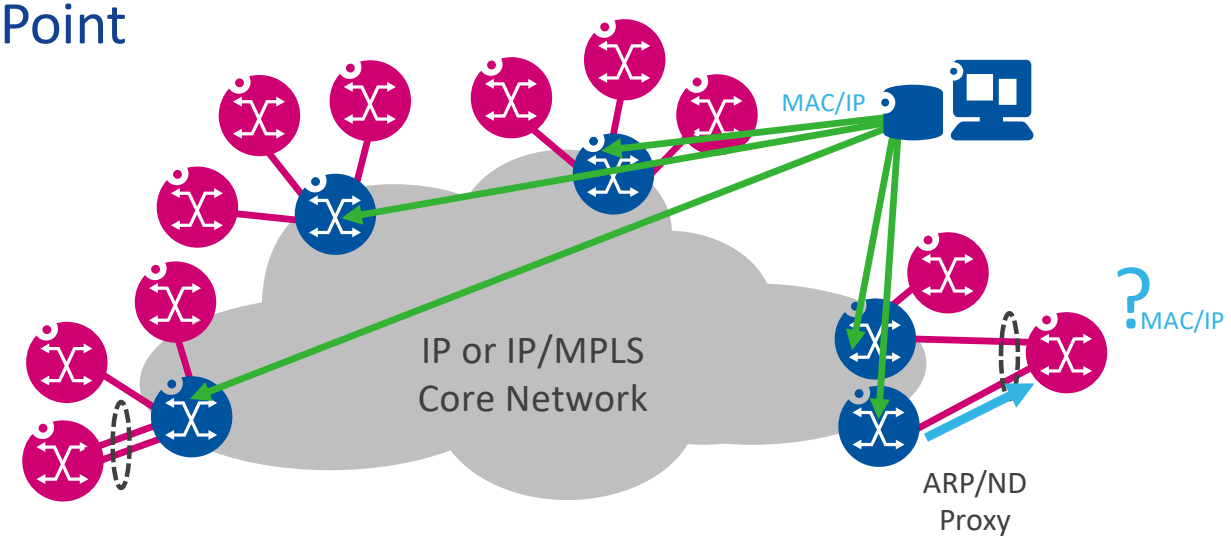
Static MAC/IP provisioning of the router interfaces for maximum security
- Suppresses unknown and ARP/ND flooding
- Drops unknown source MACs



EVPN required features
- L2 interconnection over a VXLAN or MPLS peering fabric
- Proxy-ARP/ND and unknown/ARP/ND suppression
- MAC duplication, MAC protection
- Anti-spoofing operation

Dynamic ARP/ND learning of proxy-ARP/ND entries for easy provisioning, minimum flooding and anti-spoofing monitoring
- Dynamic learning of ARP/ND entries is possible
- Anti-spoofing monitors hosts claiming the same IP
  - If a duplicate is detected, an alarm is triggered and MAC/IPs put in hold-down mode
  - An option to inject an anti-spoof mac is possible too

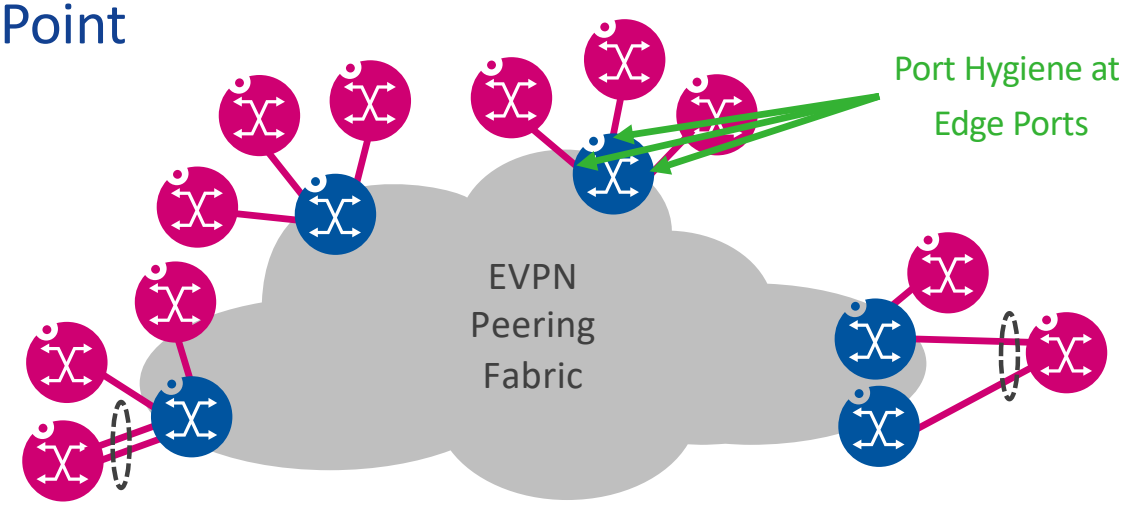
Internet Exchange Point
Peering Fabric



- Provides Layer 2 interconnection over an EVPN peering fabric
  - IP/MPLS core network with MPLS data plane
  - IP core network with VXLAN data plane
- Supports single or all-active multihoming to the peering fabric VLAN
- Supports PNIs and/or other overlay VLANs
- Enables precise fine-grained control over MAC addresses
  - Static MAC provisioning and ARP/ND proxy from PEs can reduce or eliminate unknown unicast
  - Per-MAC loop control vs per-port or per-VLAN isolates potential loops
  - Works together with edge port hygiene features to provide a clean and secure peering fabric

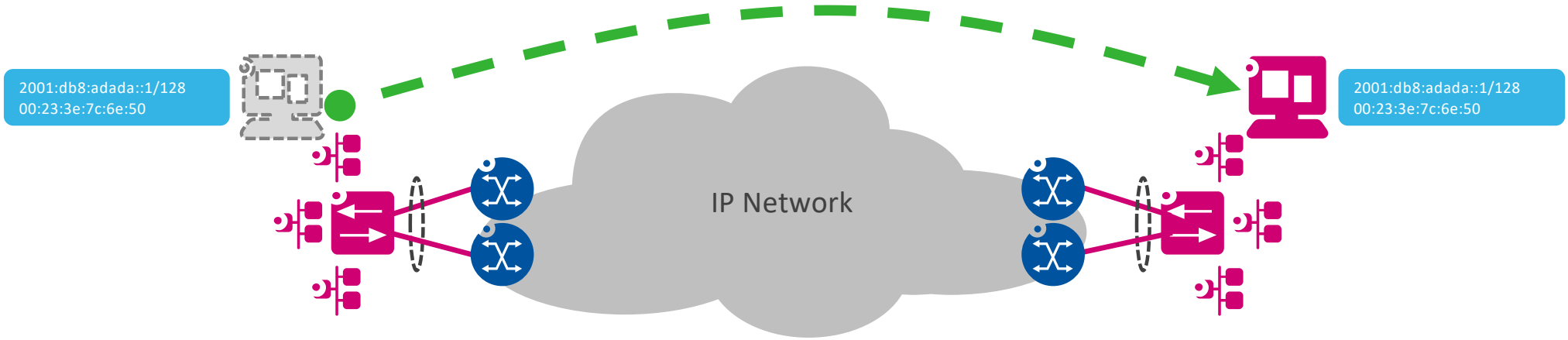
NOKIA

Internet Exchange Point
Peering Fabric



- EVPN provides the technology for the peering fabric and MAC/IP management over the core
- Still need to use existing port security mechanisms and follow BCPs for port hygiene and allowed traffic
  - Typically allow IPv4, IPv6, ARP and block unwanted traffic types
  - MAC address locking
  - BUM control

Layer 2 or Layer 3 Data Center Interconnect

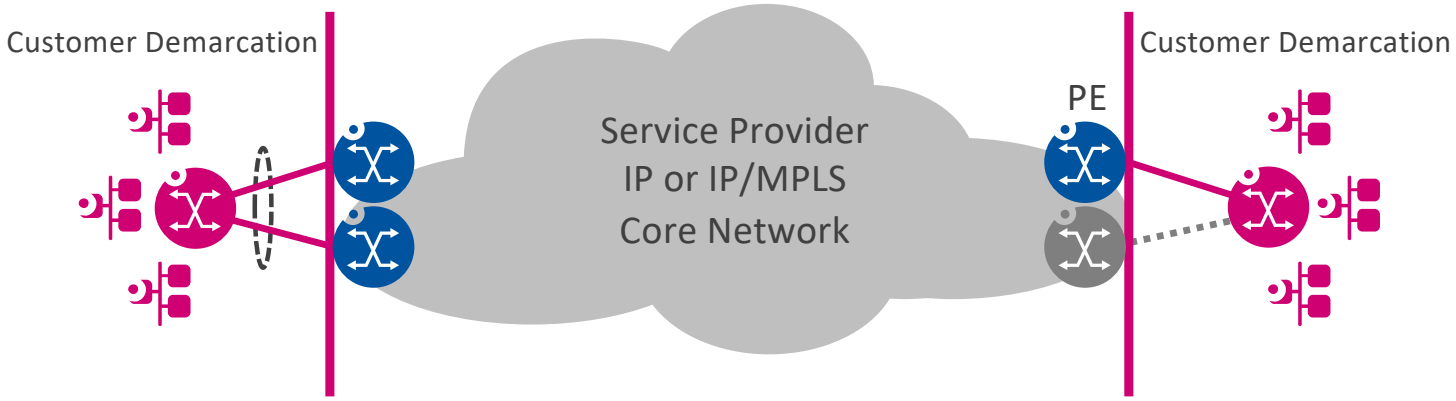


- Enables scalable Layer 2 or Layer 3 DCI services for virtualized data centers
- IP/MAC mobility for VMs that move between data centers
  - Faster moves while maintaining correct FDB on all routers
- Local IP gateway at each PE optimizes routing

- Provides all the benefits of EVPN for DCI and virtualized networks
  - No BUM for MAC learning
  - Integrated Layer 2 switching and Layer 3 routing over the same interface or VLAN

Overlay VPNs from Service Provider to Customer

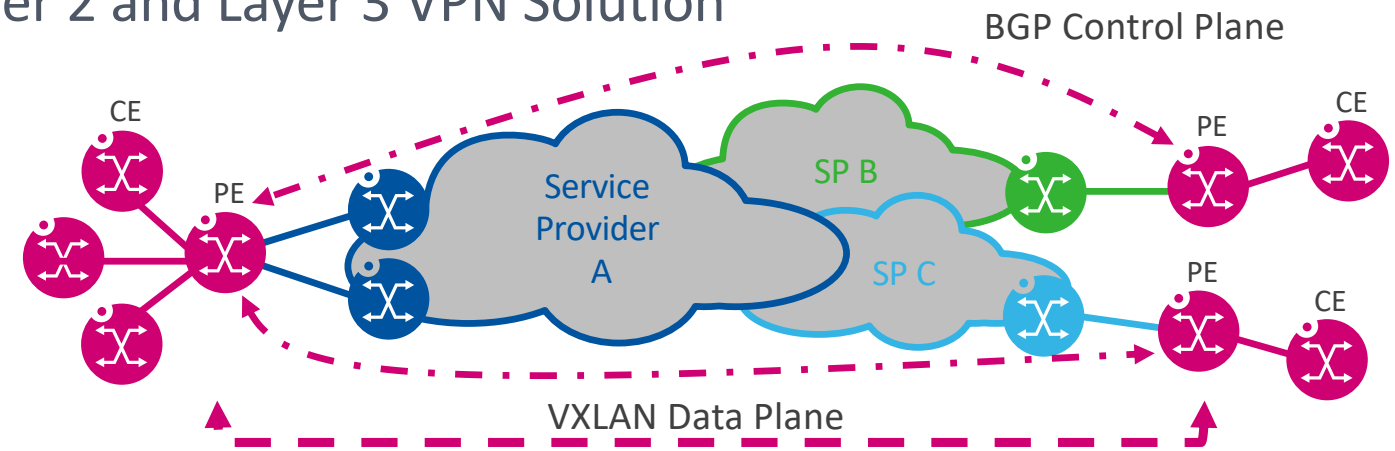
Layer 2 and Layer 3 Services



- EVPN provides Layer 2 and Layer 3 services
  - Single interface, single VLAN to customer
  - One VPN technology for both services, no need for multiple VPN protocols
  - All-active or single-active PE to CE connection
- EVPN service can be provided over any core network
  - MPLS core can use EVPN
  - IP core can use EVPN-VXLAN

Overlay VPNs over IP

Flexible Layer 2 and Layer 3 VPN Solution



- EVPN-VXLAN works over any IP service to provide a flexible Layer 2 and Layer 3 VPN
- Just requires IP connectivity between sites, no MPLS or any special configuration by IP service provider
  - Service provider network is transparent to EVPN
  - EVPN overlay is transparent to service providers
- VPN routing between endpoints can be controlled with BGP and routing policies to service providers
- Routing and MAC/IP advertisement within EVPN controlled via IBGP between PEs

Agenda

EVPN Background and Motivation

In a nutshell

EVPN Operations

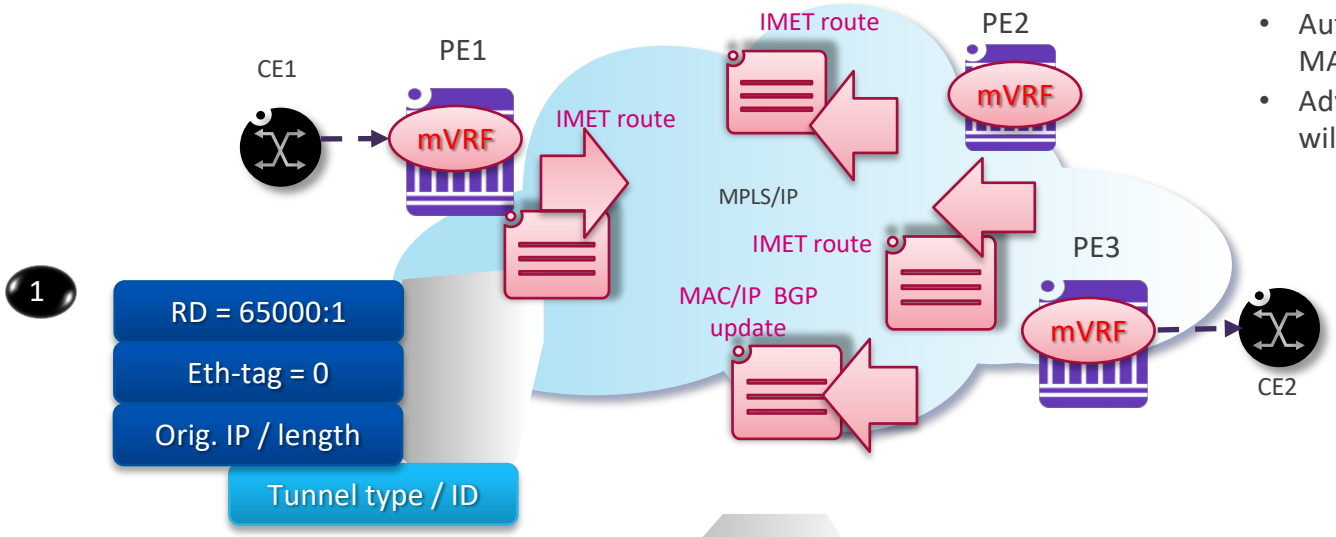
Data Planes

EVPN Use Cases/Applications

Protocol details

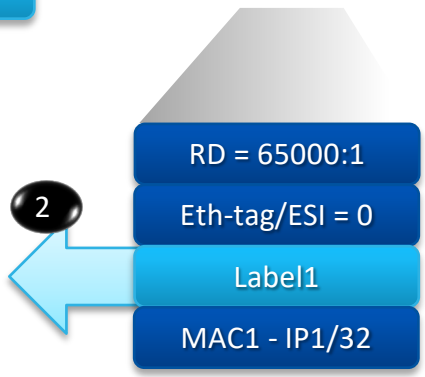
THE MP-BGP EVPN ROUTES

MAC and Inclusive Multicast routes (single home)



Indicates the tunnel type and identifier to be used to send BUM. Signals the BUM label for IR.

Carries the label for unicast traffic to MAC1



Inclusive Multicast Eth-Tag route (IMET) (type 3)

- Auto-discovers the EVPN 'bindings' from a MAC-VRF
- Advertises the provider tunnel that the PE will use for BUM
  - Ingress Replication (IR)
  - P2MP tunnel

MAC/IP Advertisement route (type 2)

Advertises learned MACs and (optionally) IP

- MACs populate remote FDBs
- MAC/IPs populate proxy-ARP tables

THE BASIC MP-BGP EVPN ROUTES: Type 2 and 3

RFC7432 Non-multi-homing routes

MAC Mobility extended community
Allows MAC mobility and MAC protection

0x06	0x03	Flags	Rsvd
Sequence Number			

EVPN NLRI encoded in MP_REACH_NLRI/ MP_UNREACH_NLRI AFI=25 SAFI=70 (EVPN)

Route Type (1 byte)
Length (1 byte)
Route Type specific (variable)

PMSI Tunnel Attribute (PTA)
Indicates tunnel type and ID for BUM traffic

Flags (1 byte)
Tunnel Type (1 byte)
MPLS label (3 bytes)
Tunnel Identifier (variable)

MAC/IP Advertisement route

Route Distinguisher (8 byte)
Ethernet Segment Id (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (4 or 16 bytes)
MPLS label1 (3 bytes)
MPLS label2 (0 or 3 bytes)

2

Eth-tag
Only in VLAN-aware bundle services

MAC Address
Used to populate remote FDBs
Length always 48

IP Address
Used to populate remote proxy-arp/nd tables
(also host routes if label2 is present)
Length always 32 or 128

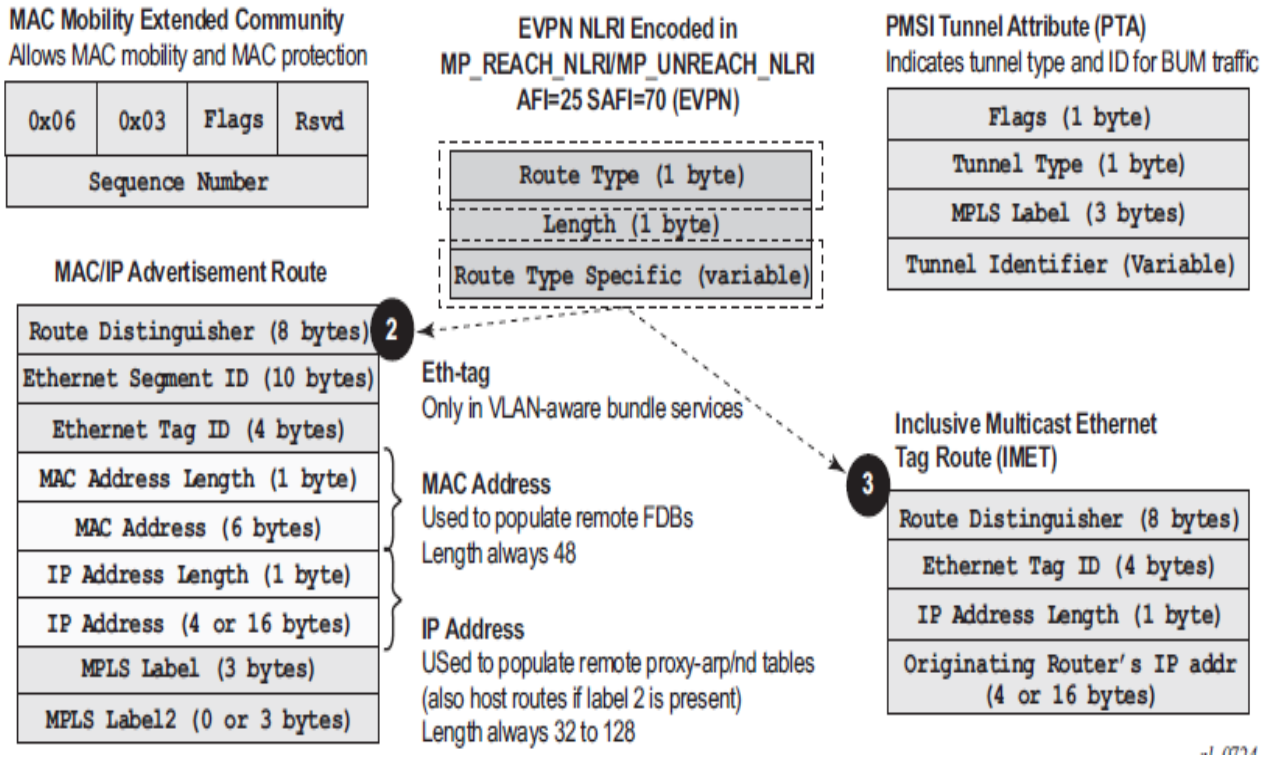
3

Inclusive Multicast Ethernet Tag Route (IMET)

Route Distinguisher (8 byte)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP addr (4 or 16 bytes)

BGP-EVPN Control Plane: Route Type 2 and 3

- Route Type 3: Inclusive Multicast Ethernet Tag Route
- Route Type 2: MAC / IP Advertisement Route

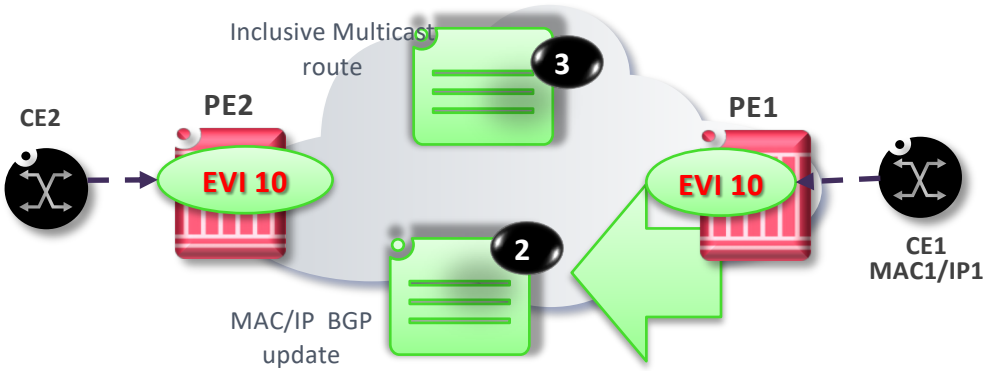


- Inclusive Multicast Ethernet Tag Route
  - Auto-discovers the VXLAN 'binds' for a service and set up the BUM tree
  - PMSI attribute:
    → Tunnel type = Ingress replication (6).
    → Flags = Leaf not required.
    → MPLS label = Carries the VNI configured in the VPLS service. Only one VNI can be configured per VPLS service.
    → Tunnel end-point = Equal to the originating IP address.

- MAC/IP Advertisement Route
  - Advertises learnt MACs or VM MAC/IPs
  - MACs populate remote FDBs
  - MAC/IPs populate proxy-ARP tables
  - ESI value = 0:0:0:0:0:0:0:0:0:0
  - MPLS Label 1: Carries the VNI configured in the VPLS service
  - MAC Address:
    → It will be 00:00:00:00:00:00 for the Unknown MAC route address.
    → It will be different from 00:…:00 for the rest of the advertised MACs.

BGP-EVPN Control Plane: Route Type 2 and 3

"Peer 1: 1.1.1.12: UPDATE
Peer 1: 1.1.1.12 - Received BGP UPDATE:
Withdrawn Length = 0
Total Path Attr Length = 88
Flag: 0x90 Type: 14 Len: 44 Multiprotocol Reachable NLRI:
Address Family EVPN
NextHop len 4 NextHop 1.1.1.12
Type: EVPN-MAC Len: 33 RD: 1.1.1.12:10 ESI: ESI-0, tag: 0, mac len: 48 mac: 52:54:00:d5:ef:60, IP len: 0, IP: NULL, label1: 10
Flag: 0x40 Type: 1 Len: 1 Origin: 0
Flag: 0x40 Type: 2 Len: 0 AS Path:
Flag: 0x80 Type: 4 Len: 4 MED: 0
Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
Flag: 0xc0 Type: 16 Len: 16 Extended Community:
target:65100:10
bgp-tunnel-encap:VXLAN



Peer 1: 1.1.1.12 - Send BGP UPDATE:
Withdrawn Length = 0
Total Path Attr Length = 84
Flag: 0x90 Type: 14 Len: 28 Multiprotocol Reachable NLRI:
Address Family EVPN
NextHop len 4 NextHop 1.1.1.11
Type: EVPN-Incl-mcast Len: 17 RD: 1.1.1.11:10, tag: 0, orig_addr len: 32, orig_addr: 1.1.1.11
Flag: 0x40 Type: 1 Len: 1 Origin: 0
Flag: 0x40 Type: 2 Len: 0 AS Path:
Flag: 0x80 Type: 4 Len: 4 MED: 0
Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
Flag: 0xc0 Type: 16 Len: 16 Extended Community:
target:65100:10
bgp-tunnel-encap:VXLAN
Flag: 0xc0 Type: 22 Len: 9 PMSI:
Tunnel-type Ingress Replication (6)
Flags: (0x0)[Type: None BM: 0 U: 0 Leaf: not required]
MPLS Label 10
Tunnel-Endpoint 1.1.1.11

BGP-EVPN: BGP Tunnel Encapsulation Extended Community

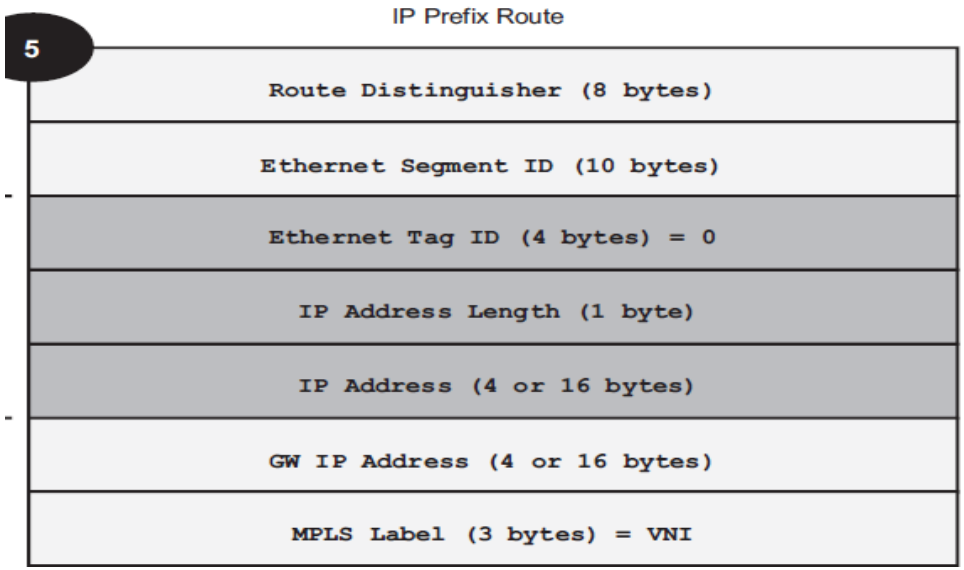
- The following routes are sent with the BGP Tunnel Encapsulation Extended Community (RFC5512):
  - → Inclusive Multicast Ethernet Tag routes.
    - → MAC / IP Advertisement routes.
    - → AD per-EVI routes.
- The Ethernet-Segment and the AD per-ESI routes don’t include this extended community.

BGP-EVPN Control Plane: Route Type 5

- Defined in draft-ietf-bess-evpn-prefix-advertisement
- Route Type 5: IP Prefix Route: Route-type 5 is used when IP prefixes advertisement is required in EVPN with an IRB backhaul R-VPLS + VPRN services
- GW IP address: Can carry two different values:

→ If different from zero, the route-type 5 will carry the primary IP interface address of the VPRN behind which the IP prefix is known. This is the case for the regular IRB backhaul R-VPLS model.

→ If 0.0.0.0, the route-type 5 will be sent along with a MAC next-hop extended community that will carry the VPRN interface MAC address. This is the case for the EVPN tunnel R-VPLS model.




```
Peer 1: 1.1.1.12 - Send BGP UPDATE:
   Withdrawn Length = 0
   Total Path Attr Length = 133
   Flag: 0x90 Type: 14 Len: 81 Multiprotocol Reachable NLRI:
      Address Family EVPN
      NextHop len 4 NextHop 1.1.1.11
      Type: EVPN-IP-Prefix Len: 34 RD: 65100:20, tag: 0, ip_prefix:
172.16.20.
1/32 gw_ip 0.0.0.0 Label: 20
      Type: EVPN-IP-Prefix Len: 34 RD: 65100:20, tag: 0, ip_prefix:
172.16.20.
0/24 gw_ip 0.0.0.0 Label: 20
   Flag: 0x40 Type: 1 Len: 1 Origin: 0
   Flag: 0x40 Type: 2 Len: 0 AS Path:
   Flag: 0x80 Type: 4 Len: 4 MED: 0
   Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
   Flag: 0xc0 Type: 16 Len: 24 Extended Community:
      target:65100:20
      mac-nh:00:0b:ff:ff:ff:51
      bgp-tunnel-encap:VXLAN
```

```
*A:>config>service>vpls# info
        allow-ip-int-bind
        .....
        bgp-evpn
          ip-route-advertisement
          vxlan
            no shutdown
```


EVPN Multi-Homing

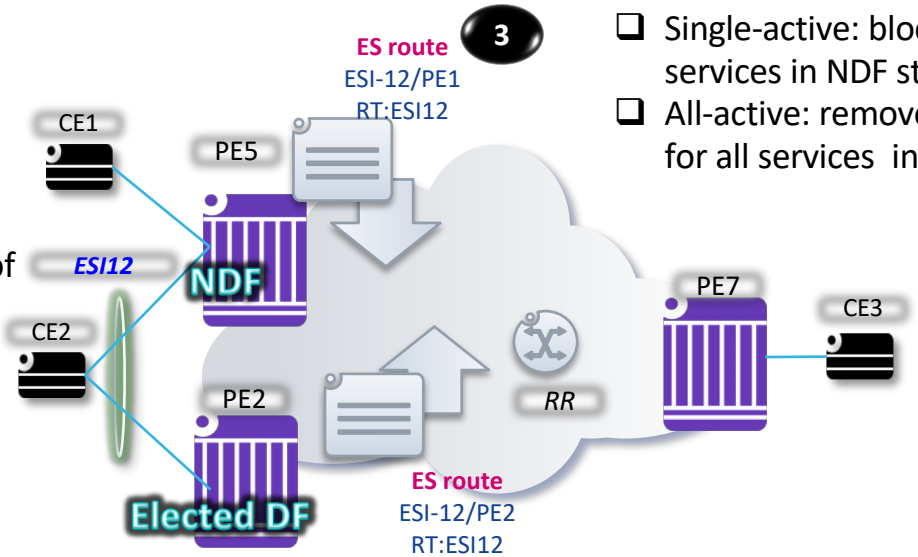
Ethernet Segment (ES) Discovery and Designated Forwarder (DF) Election

1 **ESI-12 is provisioned and advertised in an ES route**

- PE5 and PE2 advertise ESI-12 route along with an auto-derived ES Import Route Target
- PE5 and PE2 import each other's route because they have matching ES Import RT (ES is provisioned in them)
- PE7 will ignore the route as it has no matching import RT (the ES is not provisioned in it)

2 **DF Election**

- PE5/PE2 run the DF election for each service using ESI-12
- The candidate list is comprised of the IPs of the PEs that are part of the ES
- Election is based on 'service-carving'



DF and NDF service behavior

- Single-active: blocks Tx/Rx on ESI-12 for all the services in NDF state
- All-active: removes ESI-12 from the flooding list for all services in NDF state

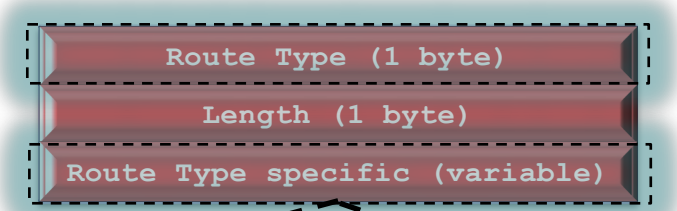
The Multi-homing EVPN routes: Route Type 1 and 4

AD per-ESI, AD per-EVI and ES routes

ESI-label extended community
Only in AD per-ESI routes

Low order bit of the flags is defined as single-active bit (0=AA)

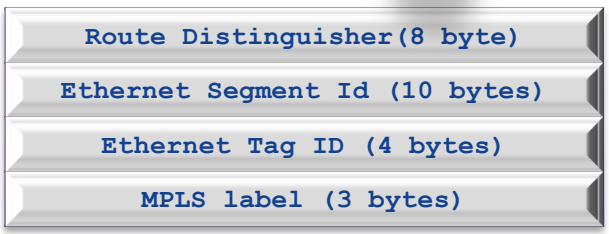
EVPN NLRI encoded in
MP_REACH_NLRI/ MP_UNREACH_NLRI
AFI=25 SAFI=70 (EVPN)



ES-Import route target
It is an auto-derived (from the MAC portion of the ESI) route-target that supports RT-Constraint

1

Ethernet AD route



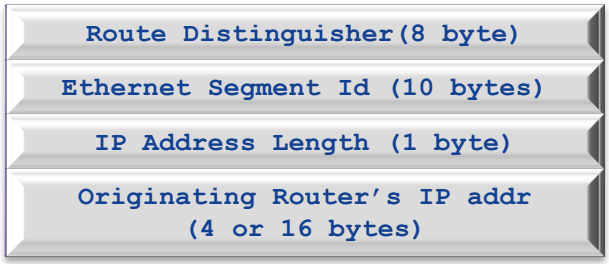
Two subtypes:

AD per-ESI
System route used to advertise the ES capabilities (mode and ESI label)
Responsible for **Mass Withdraw**

AD per-EVI
Service route used to advertise reachability to an ES
Responsible for **Aliasing**

4

Ethernet Segment Route
System route used for DF election for a given ESI



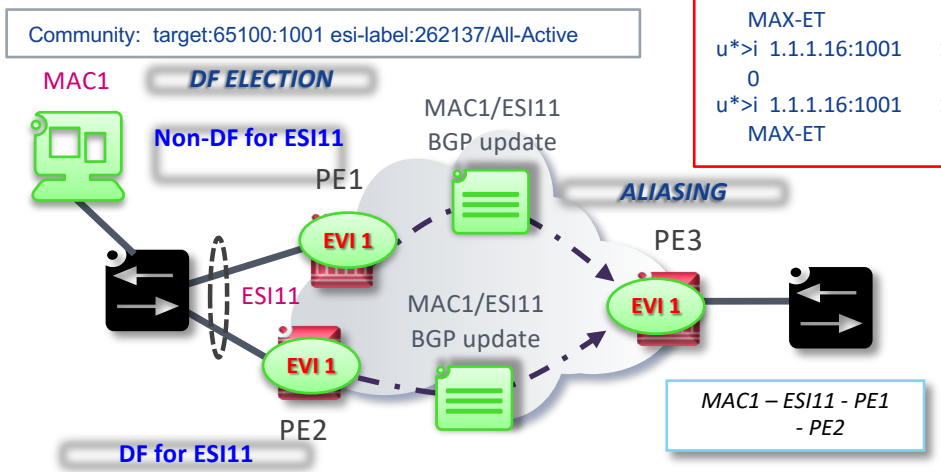
ES Advertisement Discovery Route: Route Type 1 (EVPN-VPLS Multihoming)

```
A:SR-2# show router bgp routes evpn auto-disc rd 1.1.1.13:1001
===============================================
 BGP Router ID:1.1.1.12       AS:65100      Local AS:65100
===============================================
BGP EVPN Auto-Disc Routes
===============================================
Flag  Route Dist.       ESI                    NextHop
      Tag                                      Label
-------------------------------------------------------------------------------
u*>i  1.1.1.13:1001     12:12:12:12:12:12:12:12:12:12 1.1.1.13
      0                                        LABEL 262136
u*>i  1.1.1.13:1001     12:12:12:12:12:12:12:12:12:12 1.1.1.13
      MAX-ET                                   LABEL 0
-------------------------------------------------------------------------------
Routes : 2
```

- AD per-ESI routes will announce the ethernet-segment capabilities, including the mode (single-active or all-active) as well as the ESI label for split-horizon.
- AD per-EVI routes are advertised so that PE3 knows what services (EVIs) are associated with the ESI. These routes are used by PE3 for its aliasing procedures.

```
A:SR-3# show router bgp routes evpn auto-disc
=======================================
 BGP Router ID:1.1.1.13       AS:65100      Local AS:65100
=======================================
BGP EVPN Auto-Disc Routes
=======================================
Flag  Route Dist.       ESI                    NextHop
      Tag                                      Label
-------------------------------------------------------------------------------
u*>i  1.1.1.12:1001     11:11:11:11:11:11:11:11:11:11 1.1.1.12
      0                                        LABEL 262136
u*>i  1.1.1.12:1001     11:11:11:11:11:11:11:11:11:11 1.1.1.12
      MAX-ET                                   LABEL 0
u*>i  1.1.1.16:1001     11:11:11:11:11:11:11:11:11:11 1.1.1.16
      0                                        LABEL 262136
u*>i  1.1.1.16:1001     11:11:11:11:11:11:11:11:11:11 1.1.1.16
      MAX-ET                                   LABEL 0
```

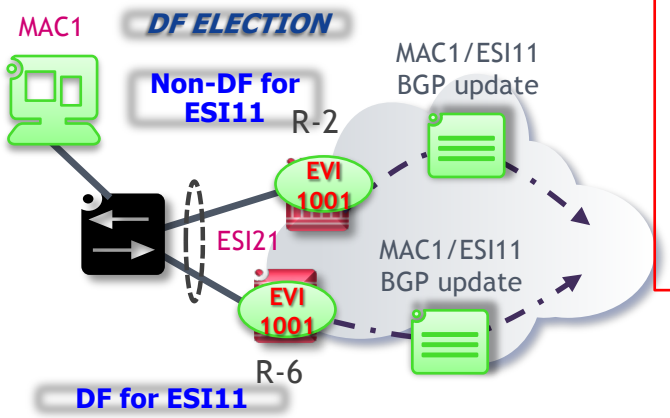
```
A:SR-3# show service id 1001 evpn-mpls esi 11-11-11-11-11-11-11-11-11-11
=======================================================
BGP EVPN-MPLS Ethernet Segment Dest
=======================================================
Eth SegId                Num. Macs          Last Change
-------------------------------------------------------------------------------
11:11:11:11:11:11:11:11:11:11  1            09/06/2016 20:55:51
=======================================================
BGP EVPN-MPLS Dest TEP Info
=======================================================
TEP Address          Egr Label          Last Change
                     Transport
-------------------------------------------------------------------------------
1.1.1.12             262136             09/06/2016 20:55:51
                     ldp
1.1.1.16             262136             09/06/2016 20:55:51
                     ldp
-------------------------------------------------------------------------------
Number of entries : 2
```



All-Active Multihoming DF Selection: Route Type 4

- The DF (Designated Forwarder) election avoids duplicate BUM flooding to all-active CEs
  - EVPN elects a DF per ESI per service
  - DF is responsible for BUM flooding the the Ethernet Segment
- When service-carving mode auto is configured the DF election algorithm will run the function [V(evi) mod N(peers) = i(ordinal)] to identify the DF for a specified service and ESI

```
A:SR-2# tools dump service system bgp-evpn ethernet-segment
"esi-11" evi 1001 df
[09/06/2016 21:55:11] Computed DF: 1.1.1.16 (Remote) (Boot
Timer Expired: Yes)
```



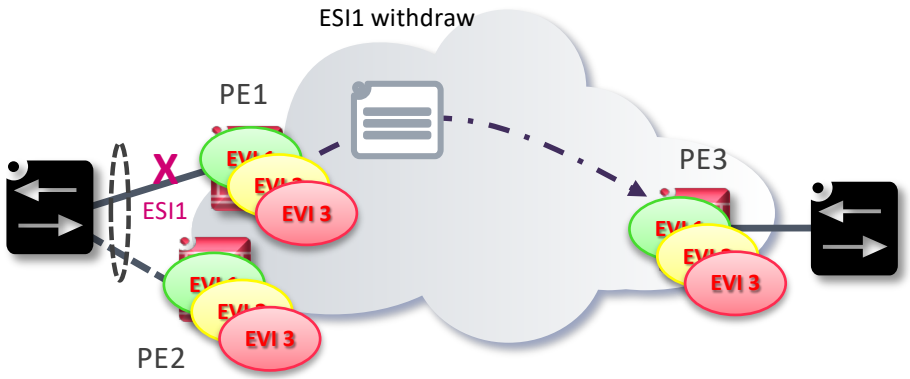

```
A:SR-2# show service system bgp-evpn ethernet-segment name "esi-11" all
===============================================
Service Ethernet Segment
===============================================
Name                    : esi-11
Admin State             : Enabled           Oper State         : Up
ESI                     : 11:11:11:11:11:11:11:11:11:11
Multi-homing            : allActive         Oper Multi-homing  : allActive
ES SHG Label            : 262142
Source BMAC LSB         : <none>
Lag Id                  : 1
ES Activation Timer     : 3 secs (default)
Exp/Imp Route-Target    : target:11:11:11:11:11:11
Svc Carving             : auto
===============================================
EVI Information
=======================================================
EVI                 SvcId               Actv Timer Rem      DF
-------------------------------------------------------------------------
30                  30                  0                   yes
1001                1001                0                   no
-------------------------------------------------------------------------
Number of entries: 2
-------------------------------------------------------------
DF Candidate list
-------------------------------------------------------------
EVI                                 DF Address
-------------------------------------------------------------
1001                                1.1.1.12
1001                                1.1.1.16
-------------------------------------------------------------
Number of entries: 2
```

```
*A:SR-2>config>service>system# info
----------------------------------------------
            bgp-evpn
               ethernet-segment "esi-11" create
                  esi 11:11:11:11:11:11:11:11:11:11
                  service-carving
                     mode auto
                  exit
                  multi-homing all-active
                  lag 1
                  no shutdown
               exit
            exit
```

```
A:SR-6# show router bgp routes evpn eth-seg
===============================================
 BGP Router ID:1.1.1.16      AS:65100      Local AS:65100
===============================================
BGP EVPN Eth-Seg Routes
===============================================
Flag  Route Dist.       ESI                     NextHop
      OrigAddr
-------------------------------------------------------------------------------
u*>i  1.1.1.12:0        11:11:11:11:11:11:11:11:11:11 1.1.1.12
      1.1.1.12
i     1.1.1.16:0        11:11:11:11:11:11:11:11:11:11 1.1.1.16
      1.1.1.16
-------------------------------------------------------------------------------
Routes : 2
```

EVPN Single-Active Multihoming and Mass-Withdraw

In single-active multihoming VPLS, individual MAC flush messages must be sent per service in order to flush the MACs
- Total convergence time grows with the number of services
- MAC-flush creates subsequent flooding



In a single-active multihoming EVPN, a mass-withdraw message is sent for all the services in the ESI.

PEs advertise:
→ MAC/IP address and its ESI (only PE1)
→ AD route per ESI (PE1 and PE2)

If a failure affects the ESI, PE1 simply withdraws the route for the ESI and the remote PE moves all the MACs to the backup PE (PE2)

Total convergence time is uniform for all the services
No need to wait for individual MACs to be withdrawn, no flooding

EVPN NLRI Route Types and Extended Communities

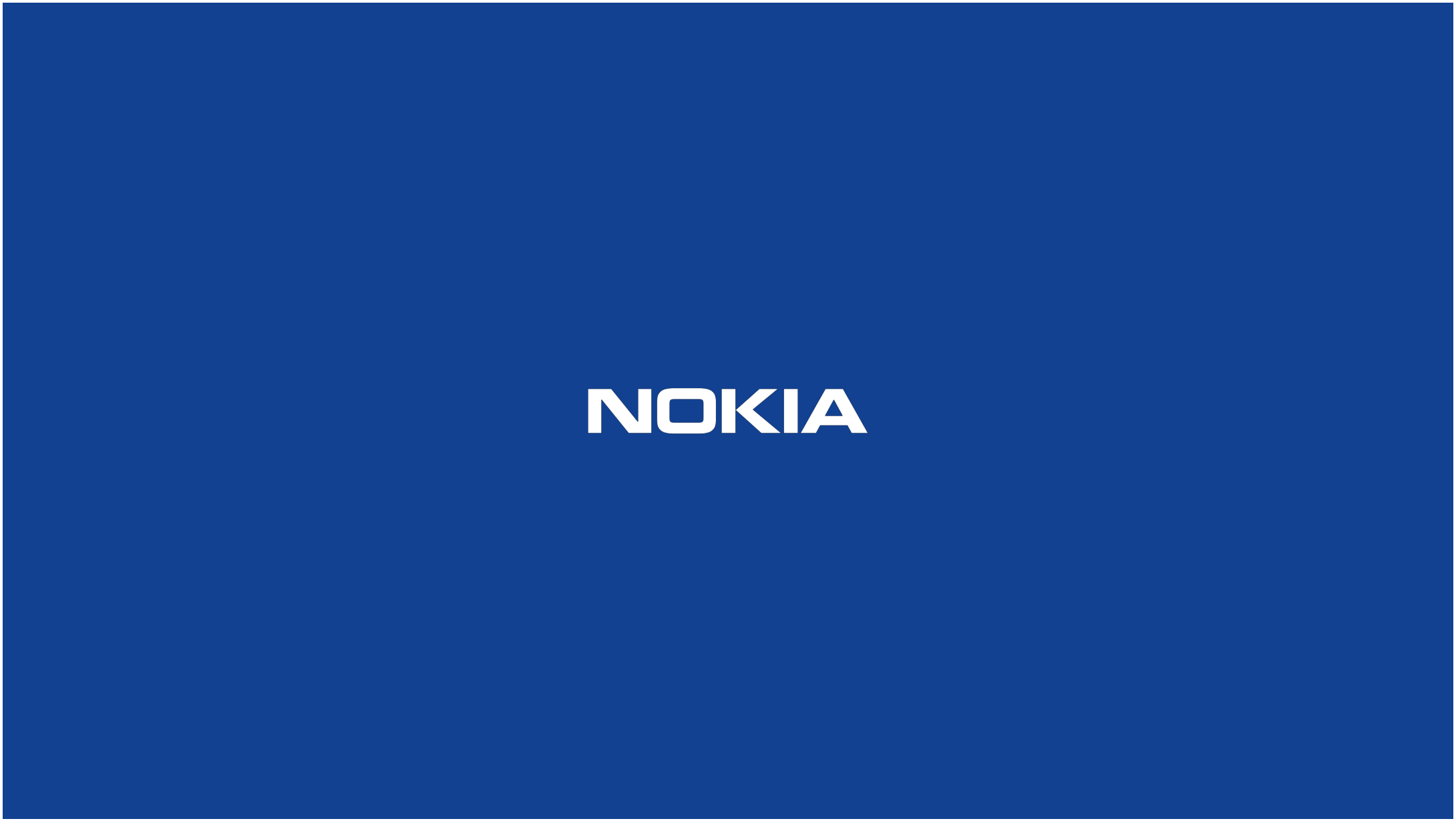
Route Type	Route Description	Route Usage	Reference
1	Ethernet Auto-Discovery (A-D) Route	Endpoint Discovery, Aliasing, Mass-Withdraw	RFC7432
2	MAC Advertisement Route	MAC/IP Advertisement	RFC7432
3	Inclusive Multicast Route	BUM Flooding Tree	RFC7432
4	Ethernet Segment Route	Ethernet Segment Discovery, DF Election	RFC7432
5	IP Prefix Route	IP Route Advertisement	draft-ietf-bess-evpn-prefix-advertisement-11

Extended Community Type	Extended Community Description	Extended Community Usage	Reference
0x06/0x01	ESI Label Extended Community	Split Horizon Label	RFC7432
0x06/0x02	ES-Import Route Target	Redundancy Group Discovery	RFC7432
0x06/0x00	MAC Mobility Extended Community	MAC Mobility	RFC7432
0x03/0x030d	Default Gateway Extended Community	Default Gateway	RFC7432

Summary

Summary

- EVPN provides next-generation VPN solutions for Layer 2 and Layer 3 services over Ethernet
  - Consistent signaled FDB in control plane using MP-BGP vs. flood-and-learn FDB in data plane
  - L3VPN-like operation for scalability and control
  - Flow-based load balancing and all-active multipathing
  - Delivering Layer 2 and Layer 3 services over the same interface, VLAN and VPN
  - Simpler provisioning and management with a single VPN technology
  - ARP/ND security and MAC provisioning
  - MPLS or IP data plane encapsulation choices
- More information
  - IETF BGP Enabled ServiceS (bess) Working Group
    http://datatracker.ietf.org/wg/bess/
  - Requirements: RFC7209
    http://tools.ietf.org/html/rfc7209
  - Base specification: RFC7432
  - http://tools.ietf.org/html/rfc7432


NOKIA