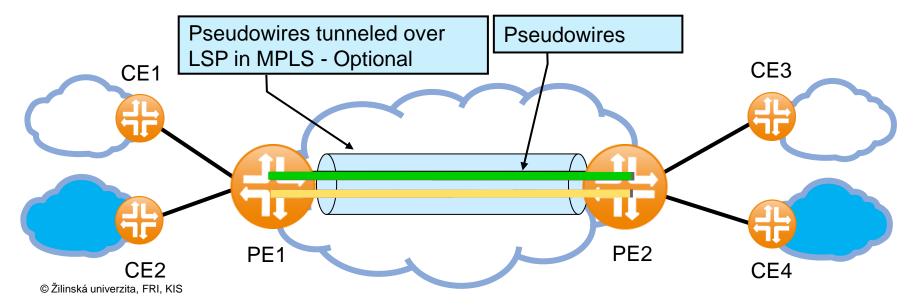
## **VPLS**

## L2 VPN types for Ethernet emulation over MPLS

- E-Line or Point2Point
  - a service connecting two customer Ethernet ports over a network
  - The IETF uses PWs (pseudowires) encapsulation for an emulated bi-directional point-to-point circuit or VPWS (Virtual Private Wire Service) for any point-to-point L2 service (Ethernet, FR, ATM, PPP, etc)
  - E-MPLS VPWS which emulates Ethernet over MPLS
- E-LAN or Any2Any
  - a multipoint service connecting a set of customer endpoints, giving the appearance to the customer of a bridged Ethernet network connecting the sites
  - The IETF uses VPLS (Virtual Private LAN Service) for a multipoint-to-multipoint service
  - The same PW encapsulation is used as for E-Line
- E-Tree or P2MP Service
  - a multipoint service connecting one or more multicast roots and a set of multicast leaves
  - VPLS + Multicast LSPs

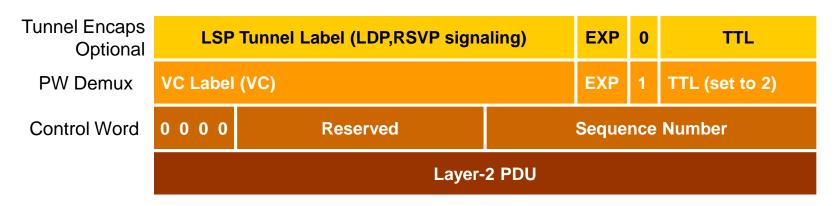
#### **Pseudo Wire Reference Model**

- A Pseudowire (PW) is a bidir logical connection between two provider edge devices connecting two attachment circuits (ACs)
- In MPLS core a Pseudowire may use one or two MPLS labels in a stack
  - VC (Virtual Circuit) Label identifying unidirectional pseudowire circuit (tunneled or not)
  - Unidirectional Tunnel LSP between PE routers to transport PW PDU from PE to PE using tunnel label(s)
    - Transport tunnel offers better scalability



#### **Ethernet PW Tunnel Encapsulation**

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1



- VC Label (inner label) used by receiving PE
  - to determine the Virtual Routing Instance (VRF)
  - Egress interface for L2PDU forwarding (Port based)
  - Egress VLAN used on the CE facing interface (VLAN Based)
- EXP can be set to the values received in the L2 frame
- Control Word is Optional (as per RFC)
  - First 4 bits set to 0 to indicates PW data
  - sequencing capability to detect out of order packets optional

## PW Frame (no transport tunnel)

```
■ Ethernet II, Src: JuniperN_04:8b:ac (5c:5e:ab:04:8b:ac), Dst: JuniperN_04:8b:ae (5c:5e:ab:04:8b:ae)

■ 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 100
   000. .... = Priority: Best Effort (default) (0)
   ...0 .... = CFI: Canonical (0)
   \dots 0000 0110 0100 = ID: 100
   Type: MPLS label switched packet (0x8847)
■ MultiProtocol Label Switching Header, Label: 262148, Exp: 0, S: 1, TTL: 255
   MPLS Label: 262148
   MPLS Experimental Bits: 0
   MPLS Bottom Of Label Stack: 1
   MPLS TTL: 255
Ethernet II, Src: JuniperN_04:8b:95 (5c:5e:ab:04:8b:95), Dst: JuniperN_04:8b:93 (5c:5e:ab:04:8b:93)

■ Internet Protocol Version 4, Src: 172.16.1.1 (172.16.1.1), Dst: 172.16.1.3 (172.16.1.3)

■ Internet Control Message Protocol
   Type: 0 (Echo (ping) reply)
   Code: 0
   Checksum: 0x0abe [correct]
   Identifier (BE): 23058 (0x5a12)
   Identifier (LE): 4698 (0x125a)
   Sequence number (BE): 2 (0x0002)
   Sequence number (LE): 512 (0x0200)
   [Response To: 255]
   [Response Time: 0.298 ms]
```

#### Virtual Private LAN Service (VPLS)

- VPLS defines an architecture allows MPLS networks offer Layer 2 multipoint Ethernet Services
- SP emulates an Ethernet bridge network (broadcast domain)
- Virtual Bridges linked with MPLS Pseudo Wires
- Full Mesh required
- There are two standards for VPLS control plane:
  - One uses BGP for auto-discovery and signaling of PWs (RFC 4761)
  - The other one uses LDP for signaling of PWs (RFC 4762)
- The data plane scheme used by both these mechanisms is same:
  - Flooding of unknown unicast, learning / aging MAC addresses
  - A key notion in the data plane is of split-horizon forwarding

#### **Control Plane - MPLS PW Types**

```
0x0001 Frame Relay DLCI (Martini Mode)
0x0002 ATM AAL5 SDU VCC transport
0x0003 ATM transparent cell transport
0x0004 Ethernet Tagged Mode (VLAN)
0x0005 Ethernet (Port)
0x0006 HDLC
0x0007 PPP
0x0008 SONET/SDH Circuit Emulation
0x0009 ATM n-to-one VCC cell transport
0x000A ATM n-to-one VPC cell transport
0x000B IP Layer2 Transport
0x000C ATM one-to-one VCC Cell Mode
0x000D ATM one-to-one VPC Cell Mode
```

```
0x000E ATM AAL5 PDU VCC transport
0x000F Frame-Relay Port mode
0x0010 SONET/SDH Circ. Emu. over Packet
0x0011 Structure-agnostic E1 over Packet
0x0012 Structure-agnostic T1 over Packet
0x0013 Structure-agnostic E3 over Packet
0x0014 Structure-agnostic T3 over Packet
0x0015 CESoPSN basic mode
0x0016 TDMoIP AAL1 Mode
0x0017 CESoPSN TDM with CAS
0x0018 TDMoIP AAL2 Mode
0x0019 Frame Relay DLCI
```

# Virtual Connection Distribution mechanism via LDP

- VC labels are exchanged across a targeted LDP session between PE routers
  - Generic Label TLV within LDP Label Mapping Message
- LDP FEC element defined to carry VC information
  - Such PW Type (RFC 4446) and VCID
- VC information exchanged using Downstream Unsolicited label distribution procedures
- Separate "MAC List" TLV for VPLS
- 1. Xconnect command entered on ingress interface, PE1 starts targeted LDP session with PE2
- PE1 allocates VC label for interface & binds to configured VC ID
- PE1 sends VC label mapping message containing VC FEC TLV, VC label TLV
- 4. PE2 receives it and matches local VC ID
- PE2 repeats it as so that bidir VC mapping is established

#### LDP label distribution

```
Label Distribution Protocol
   Version: 1
   PDU Length: 46
   LSR ID: 10.255.255.3 (10.255.255.3)
   Label Space ID: 0
 Label Mapping Message
     0... = U bit: Unknown bit not set
     Message Type: Label Mapping Message (0x400)
     Message Length: 36
     Message ID: 0x00000008
   ■ Forwarding Equivalence Classes TLV
       00.. .... = TLV Unknown bits: Known TLV, do not Forward (0x00)
       TLV Type: Forwarding Equivalence Classes TLV (0x100)
       TLV Length: 20

□ FFC Flements

       ■ FEC Element 1 VCID: 10
          FEC Element Type: Virtual Circuit FEC (128)
          0... = C-bit: Control Word NOT Present
           .000 0000 0000 0101 = VC Type: Ethernet (0x0005)
          VC Info Length: 12
          Group ID: 0
          VC ID: 10

    ■ Interface Parameter: MTU 1500

        ■ Interface Parameter: VCCV
            ID: VCCV (0x0c)
            Length: 4

    □ CC Type

              .... 0 = PWE3 Control Word: False
              .... ..1. = MPLS Router Alert: True
              .... .1.. = MPLS Inner Label TTL = 1: True
          ■ Generic Label TLV
       00.. .... = TLV Unknown bits: Known TLV, do not Forward (0x00)
       TLV Type: Generic Label TLV (0x200)
       TLV Length: 4
       Generic Label 262148
```

#### **BGP Label distribution**

```
■ Ethernet II, Src: JuniperN_04:8b:ae (5c:5e:ab:04:8b:ae), Dst: JuniperN_04:8b:ac (5c:5e:ab:04:8b:ac)

■ 802.1Q Virtual LAN, PRI: 6, CFI: 0, ID: 100

■ Internet Protocol Version 4, Src: 10.255.255.3 (10.255.255.3), Dst: 10.255.255.1 (10.255.255.1)

∃ Transmission Control Protocol, Src Port: 60640 (60640), Dst Port: bgp (179), Seq: 163, Ack: 163, Len: 88

∃ Border Gateway Protocol
 ■ UPDATE Message
     Marker: 16 bytes
     Length: 88 bytes
     Type: UPDATE Message (2)
     Unfeasible routes length: 0 bytes
     Total path attribute length: 65 bytes
   ■ Path attributes

    ⊕ ORIGIN: IGP (4 bytes)

■ AS_PATH: empty (3 bytes)

     ■ EXTENDED_COMMUNITIES: (19 bytes)

∃ Flags: 0xc0 (Optional, Transitive, Complete)

        Type code: EXTENDED_COMMUNITIES (16)
        Length: 16 bytes
      □ Carried Extended communities
          UnknownRoute Target: 100:201

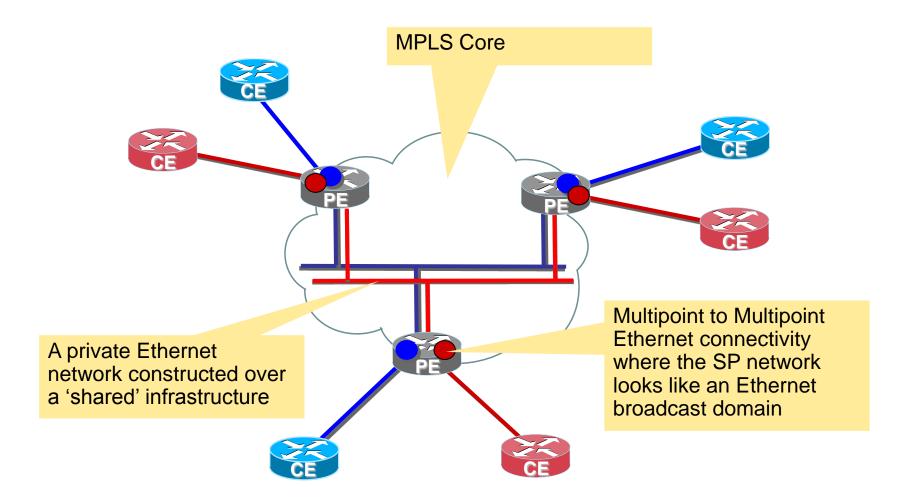
■ UnknownLayer 2 Information: Upknown, Control Flags: none, MTU: 0 bytes

     ■ MP_REACH_NLRI (32 bytes)

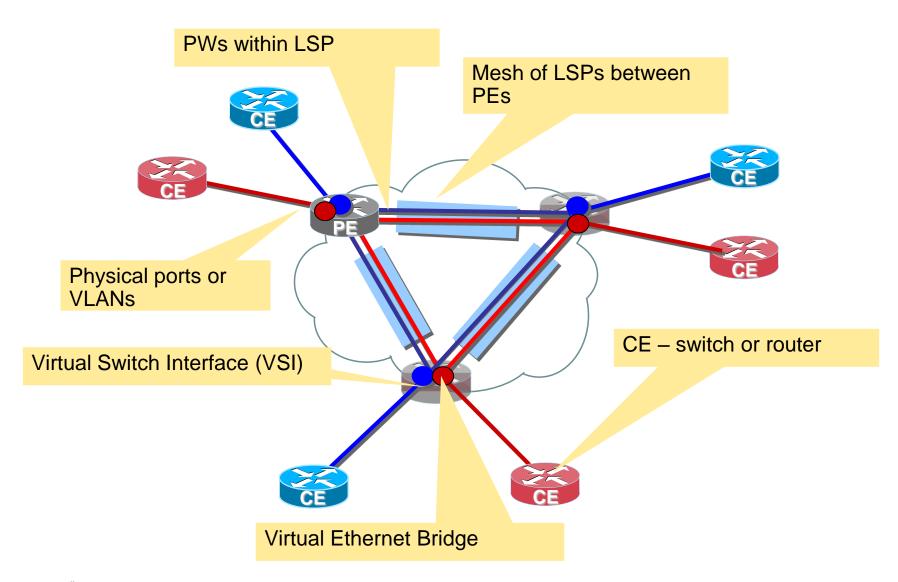
■ Flags: 0x90 (Optional, Non-transitive, Complete, Extended Length)

        Type code: MP_REACH_NLRI (14)
        Length: 28 bytes
        Address family: Layer-2 VPN (25)
        Subsequent address family identifier: VPLS (65)
       ■ Next hop network address (4 bytes)
          Next hop: IPv4=10.255.255.3 (4)
        Subnetwork points of attachment: 0
       ■ Network layer reachability Information (19 bytes)
          RD: 100:0.0.0.203, CE-ID: 3, Label-Block Offset: 1, Label-Block Size: 8 Label Base 262401 (bottom)
```

#### **VPLS Customer's View**



## **VPLS Components**



#### **VPLS Virtual Switch Instance Functions**

 Virtual Switch Instance/Interface (VSI) is an Ethernet bridge function entity of a VPLS instance on a PE. It emulates legacy bridging by forwarding Layer 2 frames based on MAC address and VLAN tag

#### Flooding / Forwarding

- MAC table instances per customer (port/vlan) for each PE
- Forwarding Instance (VFI) will participate in learning and forwarding process
- Associate ports to MAC, flood unknowns to all other ports

#### Address Learning / Aging

- LDP enhanced with additional MAC List TLV (label withdrawal)
- MAC timers refreshed with incoming frames

#### Loop Prevention

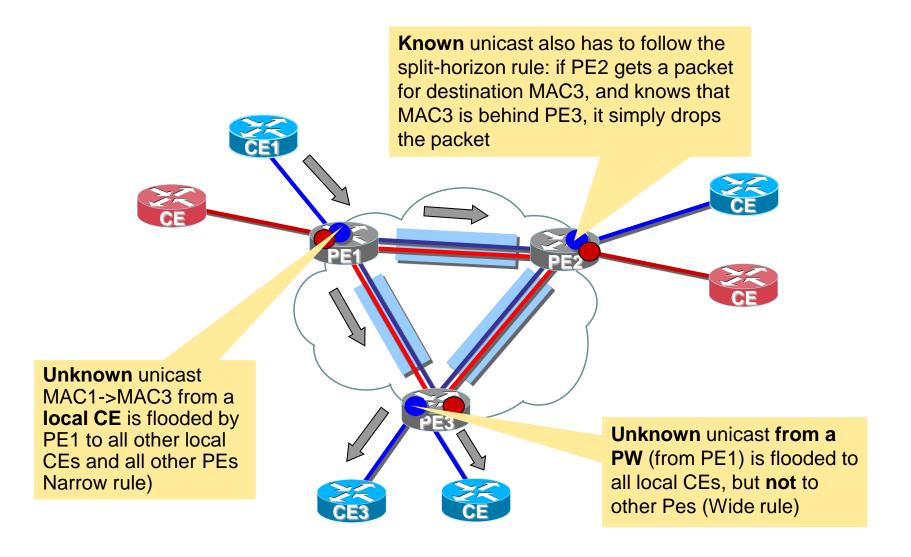
- Create full-mesh of Pseudo Wire VCs (EoMPLS)
- Unidirectional LSP carries VCs between pair of PEs
- A VPLS use "split horizon" concepts to prevent loops

## **Split Horizon Semantics**

#### Two types of split horizon

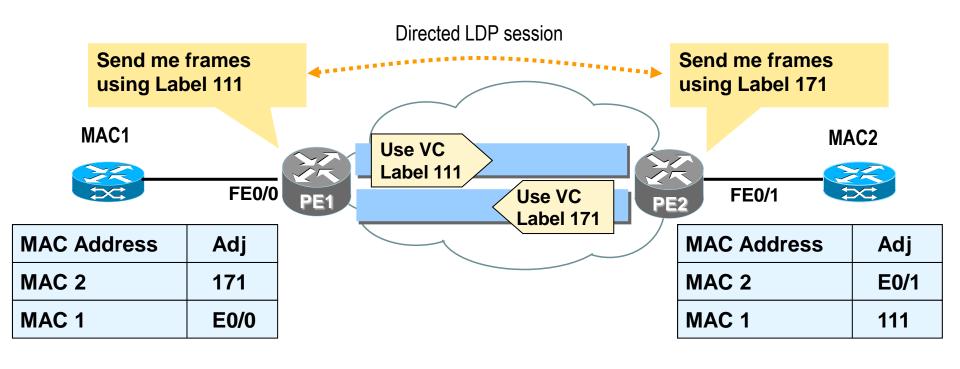
- "Narrow"
  - receive from one CE, send to all other CEs
  - don't send to CE from which the packet came
  - Forwards to all PWs within VPLS instance
- "Wide"
  - receive from a device in a VPLS instance, don't send to any device in that VPLS instance, just to local CEs

## **Split Horizon Example**



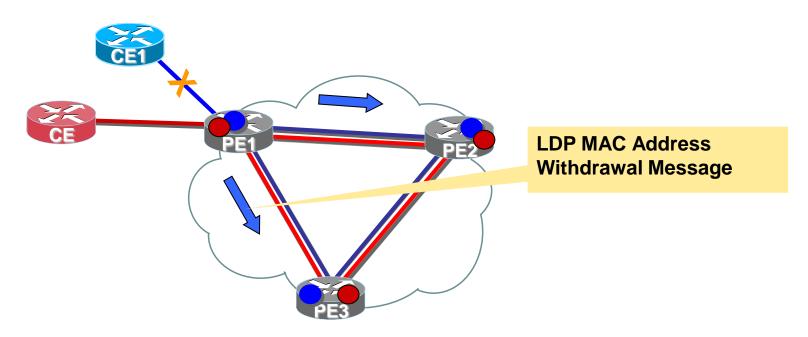
#### **MAC Address Learning and Forwarding**

- Broadcast, Multicast, and Unknown Unicast are learned via the received label associations
- Two unidir LSPs associated with a VC (Tx & Rx)
- If inbound or outbound LSP is down then the entire Pseudowire is considered down



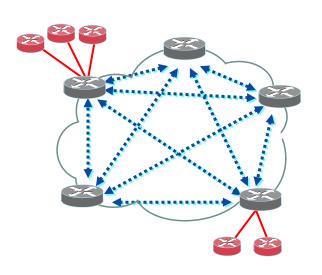
#### LDP MAC Address Withdrawal Message

- Message speeds up convergence process
- otherwise PE relies on MAC Address Aging Timer
- Upon failure PE removes locally learned MAC addresses
- Send LDP Address Withdrawal Message to remote PEs in VPLS (using LDP session)
  - New MAC List TLV (type-length-value) is used to withdraw addresses



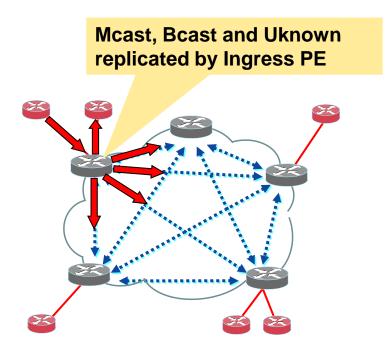
## Single Domain VPLS Scalability

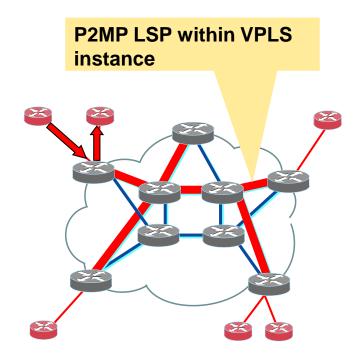
- Suitable for simple/small implementations
- Full mesh of PWs required (for both LDP and BGP based)
  - N\*(N-1)/2 Pseudo Wires
  - For LDP based full mesh of directed LDP sessions required
  - For LDP based manual configuration & provisioning issues
- No hierarchical scalability
- Stability / Recovery issues
- Potential packet replication overhead
  - PE-to-PE flooding used to use ingress replication of unknowns, broadcast and multicast
  - CPU overhead for replication



#### **Ingress Replication versus P2MP LSPs**

 Some vendors integrate into VPLS point-to-multipoint LSPs in such a way that broadcast, multicast, and unknown unicast will be forwarded using point-to-multipoint LSP instead of ingress replicating





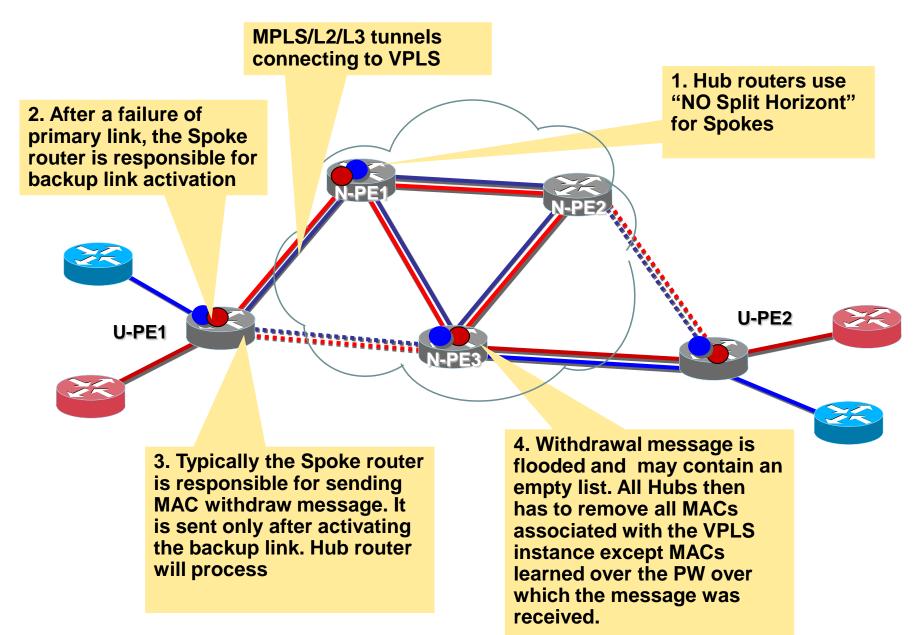
## **Control Plane**

Characteristic	LDP VPLS	BGP VPLS	
Full-mesh requirement	Alleviated only somewhat by H-VPLS, though at the expense of introducing changes and additional overhead in the dataplane	' I	
Provisioning task of adding or removing PE router	Only somewhat simplified by H-VPLS	Highly simplified by use of Route Reflector	
Provisioning task of adding or changing VPLS cutomer sites	Manual or through provisioning	Automated by BGP'd autodiscovery	
VPLS with P2MP LSP integration to scale forwarding and data planes	Currently still not fully standardized, based on mLDP draft-ietf-l2vpn-vpls-mcast-13.txt	Already implemented by some vendors using RSVP-TE signalling	
Signaling overhead	Increases in proportion to the total number of PWs in the network	Minimal because each signaling update can be used to establish multiple PWs	

#### **Multi-Domain VPLS**

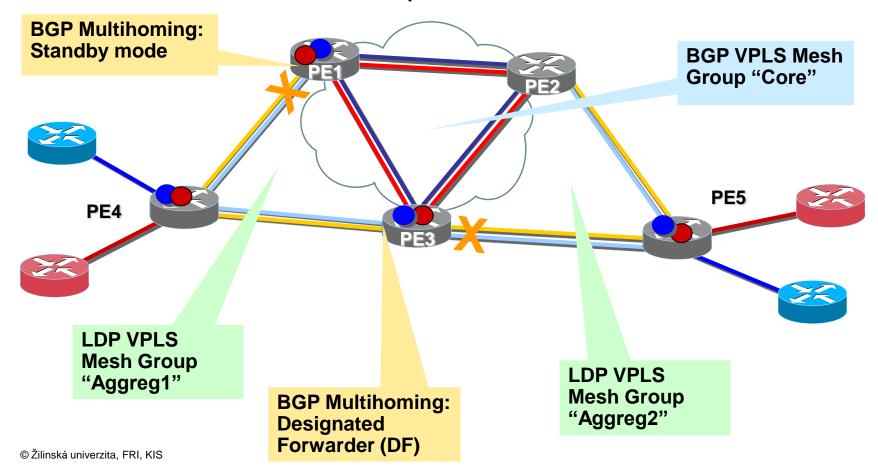
- Hierarchical LDP based VPLS (H-VPLS)
  - Consists of two levels in a Hub and Spoke topology
  - Hub or N-PE (Network facing PE router) consists of full mesh VPLS Pseudo Wires in MPLS core
  - Spokes or U-PE (User facing PE router) consist of a single or pair of PWs connecting to VPLS (Hub) PEs
- Multi domain concept with BGP VPLS in the core
  - BGP VPLS domain in the core
  - Typically LDP VPLS domains in the aggregation part of the network
    - The reason for LDP in aggregation might be lack of BGP on edge PE routers - LDP is considered as a less complex technology
  - From the perspective of BGP VPLS, the LDP VPLS domain is equivalent to a CE
  - The LDP VPLS domain can be multihomed to redundant interworking BGP VPLS PE routers
  - LDP VPLS domain has no idea about core BGP VPLS domain

#### H-VPLS



#### Core BGP VPLS interconnected to Access LDP ones

- In JunOS for instance (Juniper), domain or "mesh-group" construct can be used to create any of both mentioned 2level hierarchy
- The mesh-groups is notion in regards to split-horizon rules in VPLS which introduce "no split horizon" rule

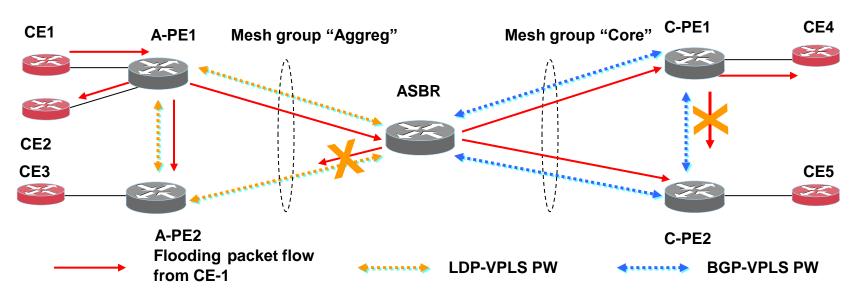


## Multi Domain Split Horizon

- ASBR (autonomous-system-boundary) PE router hosts more than one mesh groups (any combination of LDP and BGP based)
- Split-horizon is much more complex now:
  - PE border router must not forward packets received from mesh gr 1 back to mesh gr 1, or from mesh gr 2 to mesh gr 2, or from core to core
  - But forwards in case of different mesh groups
- For a multi-homed site, a PE can be a designated forwarder only if the site hosted on it is operationally UP

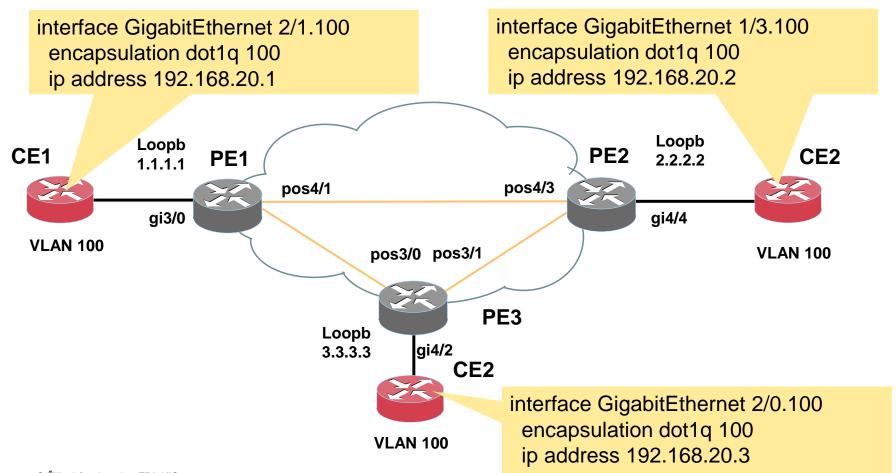
#### Mesh Groups: Flooding example

- Assume CE-1 sends a broadcast ARP request packet. It will be flooded by APE1, to all PEs in "Aggreg" domain (including ASBR) as it's fully-meshed.
- ASBR receives this packet from APE1 and forwards it to all the mesh-groups (PWs) except the one in which packet is received, as a result packet is forwarded on all PWs part of mesh-group "Core".
- Upon receiving this packet from APE1, ASBR learns the CE1 MAC address via APE1 PW.

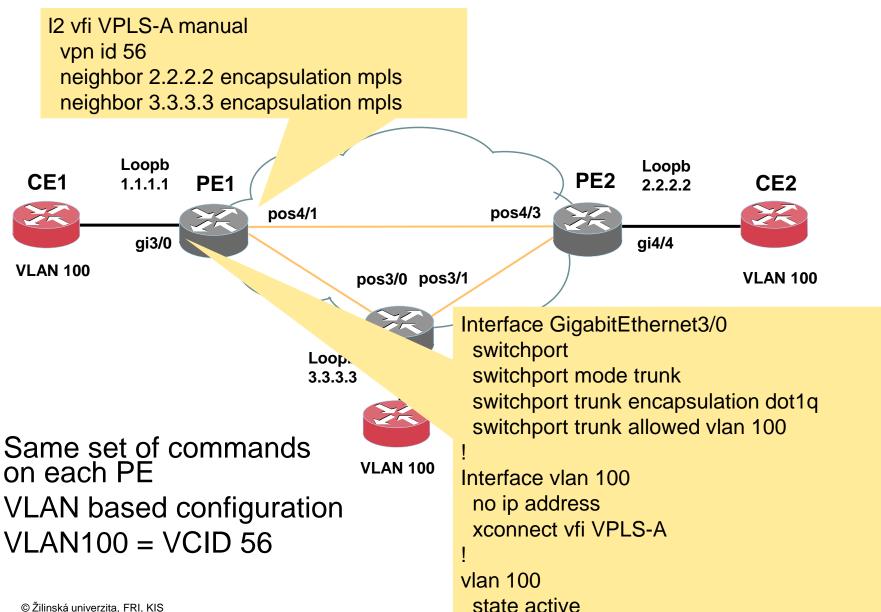


## Simple VPLS Configuration Example (c7600)

- CEs are all part of same VPLS instance (VCID = 56)
- CE router connects using VLAN 100 over sub-interface
- Can also be just port based (no VLAN)



## Simple VPLS Configuration Example



## show mpls 12

PE2#show mpls I2 vc detail

Local interface: VFI VPLS-A up

Destination address: 1.1.1.1, VC ID: 56, VC status: up

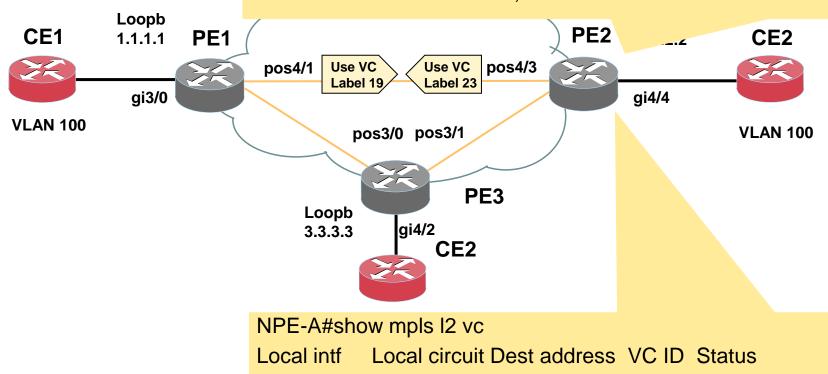
Tunnel label: imp-null, next hop 156.50.20.1

Output interface: POS4/3, imposed label stack {19}

Create time: 1d07h, last status change time: 00:40:14

Signaling protocol: LDP, peer 1.1.1.1:0 up

MPLS VC labels: local 23, remote 19

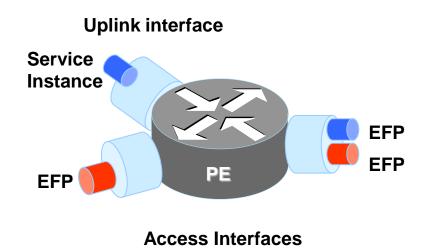


------

VFI VPLS-A VFI 1.1.1.1 56 UP VFI VPLS-A VFI 3.3.3.3 56 UP

## Providing a Service on an Access Port

- EFP (Ethernet Flow Point) Overview on VLAN or port
  - interface fe-slot/port.<sub-intf no.> l2transport
  - <match criteria commands> (VLAN tags, MAC, Ethtype)
  - <rewrite commands> (VLAN tags pop/push/translation)
  - <feature commands> (QoS, ACL, etc.)



#### **VLAN Manipulation / Rewriting**

- Provides port isolation allowing multiple customers single switch to have same internal VLANs
- Flexibility for Service Provider to choose VLANs to pass over core
- Does NOT require mixing of Customer and Provider VLANs

PE (config-subif)#rewrite ingress tag?

pop Remove one or more tags

push Push one or more tags

translate Replace tags with other tags

PE (config-subif)#rewrite ingress tag push dot1q 100 ? second-dot1q Push another Dot1Q tag symmetric All rewrites must be symmetric

PE (config-subif)#rewrite ingress tag pop?

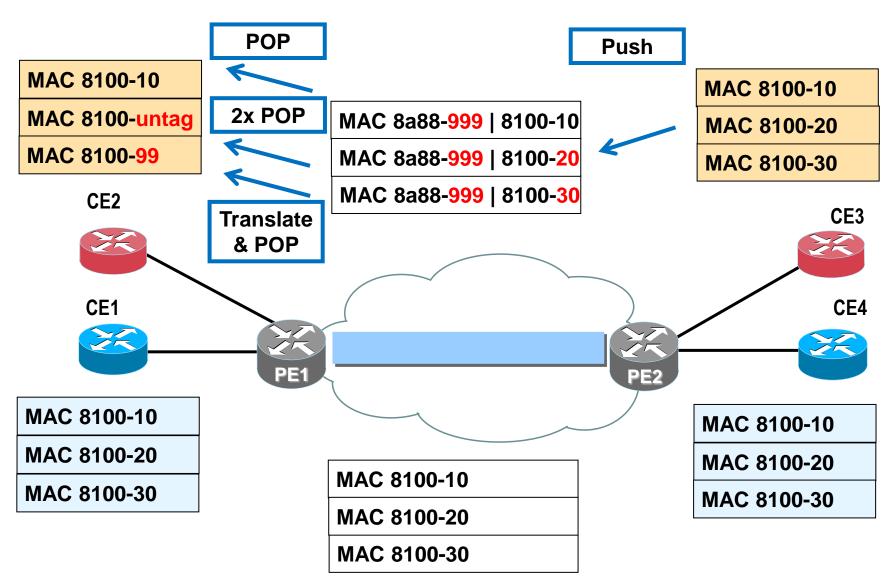
1 Remove outer tag only
2 Remove two outermost tags

PE(config-subif)#rewrite ingress tag translate?
1-to-1 Replace the outermost tag with another tag
1-to-2 Replace the outermost tag with two tags
2-to-1 Replace the outermost two tags with one

PE (config-subif)#rewrite ingress tag push ? tag
dot1ad Push a Dot1ad tag
dot1q Push a Dot1Q tag
other

**2-to-2** Replace the outermost two tags with two other tags

#### QinQ Mapping over SP's Network Example



## **VPLS LDP based example - JunOS**

Local interface: lsi.17827079, Status: Up, Encapsulation: ETHERNET

Description: Intf - vpls VPLS-LDP-1 neighbor 10.255.255.3 vpls-id 10

```
lab@MX80# run show vpls mac-table logical-system R1
                                Logical system : R1
                                Routing instance : VPLS-LDP-1
                                Bridging domain : __VPLS-LDP-1__, VLAN : none
                                                               Logical
                                  MAC
                                                      MAC
                                                      flags.
                                  address
                                                               interface
                                   5c:5e:ab:04:8b:91
                                                               lsi.17827078
lab@MX80# run show configuration
                                  5c:5e:ab:04:8b:93 D
                                                               lsi.17827079
instance-type vpls;
                                   5c:5e:ab:04:8b:95 D
                                                               ge-1/2/0.101
vlan-id none;
interface ge-1/2/0.101;
protocols {
   vpls {
                                lab@MX80# run show arp logical-system R101
       no-tunnel-services;
                               MAC Address
                                                                                           Interface
                                                 Address
                                                                 Name
                               5c:5e:ab:04:8b:91 172.16.1.2
       vpls-id 10;
                                                                 172.16.1.2
                                                                                           ge-1/2/5.101
       neighbor 10.255.255.2; 5c:5e:ab:04:8b:93 172.16.1.3
                                                                 172.16.1.3
                                                                                           ge-1/2/5.101
       neighbor 10.255.255.3; Total entries: 2
lab@MX80# run show vpls connections logical-system R1
Layer-2 VPN connections:
Instance: VPLS-LDP-1
 VPLS-id: 10
                                          Time last up
   Neighbor
                             Type St
                                                                # Up trans
   10.255.255.2(vpls-id 10) rmt
                                          Apr 16 10:15:45 2013
                                   Up
     Remote PE: 10.255.255.2, Negotiated control-word: No
     Incoming label: 262145, Outgoing label: 262145
     Negotiated PW status TLV: No
     Local interface: lsi.17827078, Status: Up, Encapsulation: ETHERNET
       Description: Intf - vpls VPLS-LDP-1 neighbor 10.255.255.2 vpls-id 10
   10.255.255.3(vpls-id 10) rmt Up
                                          Apr 16 10:15:50 2013
     Remote PE: 10.255.255.3, Negotiated control-word: No
     Incoming label: 262146, Outgoing label: 262145
     Negotiated PW status TLV: No
```

Flags

none

none

## **VPLS BGP based example - JunOS**

```
lab@MX80# run show vpls mac-table logical-system R1
                               Logical system : R1
                               Routing instance : VPLS-BGP-2
                                Bridging domain : __VPLS-BGP-2__, VLAN : none
                                                               Logical
                                  MAC
                                                      MAC
                                  address
                                                      flags.
                                                                interface
lab@MX80# run show configuratio
                                  5c:5e:ab:04:8b:91
                                                                lsi.17827081
instance-type vpls;
                                  5c:5e:ab:04:8b:93
                                                                lsi.17827080
vlan-id none:
                                  5c:5e:ab:04:8b:95
                                                               ge-1/2/0.201
interface ge-1/2/0.201;
route-distinguisher 100:201;
                               [edit]
vrf-target target:100:201;
                               lab@MX80# run show arp logical-system R201
protocols {
                               MAC Address
                                                 Address
                                                                                            Interface
                                                                                                                Flags
                                                                  Name
   vpls {
                               5c:5e:ab:04:8b:91 172.16.2.2
                                                                 172.16.2.2
                                                                                            ge-1/2/5.201
                                                                                                                none
       site-range 20:
                               5c:5e:ab:04:8b:93 172.16.2.3
                                                                 172.16.2.3
                                                                                            ge-1/2/5.201
                                                                                                                none
       no-tunnel-services:
                               Total entries: 2
        site CE1 {
           site-identifier 1:
            interface ge-1/2/0.201;
lab@MX80# run show vpls connections logical-system R1
Layer-2 VPN connections:
Instance: VPLS-BGP-2
 Local site: CE1 (1)
    connection-site
                              Type St
                                           Time last up
                                                                  # Up trans
                                           Apr 16 11:27:28 2013
                              rmt
                                    Up
     Remote PE: 10.255.255.2, Negotiated control-word: No
     Incoming label: 262402, Outgoing label: 262401
     Local interface: lsi.17827081, Status: Up, Encapsulation: VPLS
       Description: Intf - vpls VPLS-BGP-2 local site 1 remote site 2
                                    Up
                                           Apr 16 11:27:27 2013
                                                                           1
                              rmt
      Remote PE: 10.255.255.3, Negotiated control-word: No
     Incoming label: 262403, Outgoing label: 262401
     Local interface: lsi.17827080, Status: Up, Encapsulation: VPLS
        Description: Intf - vpls VPLS-BGP-2 local site 1 remote site 3
```

#### **Deployment Issues**

- MTU Size
  - Core MTU ≥ Edge MTU + Transport Header + AToM Header + (MPLS Label Stack \* MPLS Header Size)

	Edge	Transport	AToM	MPLS Stack	MPLS Header	Total
EoMPLS Port Mode	1500	14	4 [0]	2	4	1526 [1522]
EoMPLS VLAN Mode	1500	18	4 [0]	2	4	1530 [1526]
EoMPLS Port w/ TE FRR	1500	14	4 [0]	3	4	1530 [1526]

- Broadcast Handling
  - Ethernet switches replicate broadcast/multicast flows once per output interface, VPLS may duplicate packets over the same physical egress interface – for each PW

#### **Deployment Issues**

- Router or a Switch as a CPE?
  - Router
    - Single MAC, SPT iteractions, no broadcast storms
    - BUT does not provide direct L2 connectivity but
  - Switch
    - Too many MAC addresses can exist within VPLS
    - High exposure to broadcast storms
    - Non-predictable and excessive number of Unknown, Broadcast and Multicast frames could behave as a series of "packet bombs"
    - High MAC address learning rate -> CPU issues
- Solution: Ingress Threshold Filters on U-PE or N-PE
  - How to selectively choose which Ethernet Frames to discard?
  - How to avoid dropping Routing and Keepalives (control)
  - How many MAC addresses allowed?
  - Does SP really want to take this responsibility?

# Ďakujem za pozornosť

roman dot kaloc at gmail dot com