



L2VPN TUTORIAL

Path to Convergence

Muhammad Waris Sagheer (waris@cisco.com)

Paresh Shah (pashah@cisco.com)

SANOG2006

Agenda

- 1. Introduction to L2VPNs**
- 2. Signaling Concepts**
- 3. VPWS Transports**
- 4. VPWS Service Interworking**
- 5. Virtual Private LAN Service**
- 6. Pseudo Wire Stitching**
- 7. Quality of Service**
- 8. Demonstration**
- 9. Q&A**

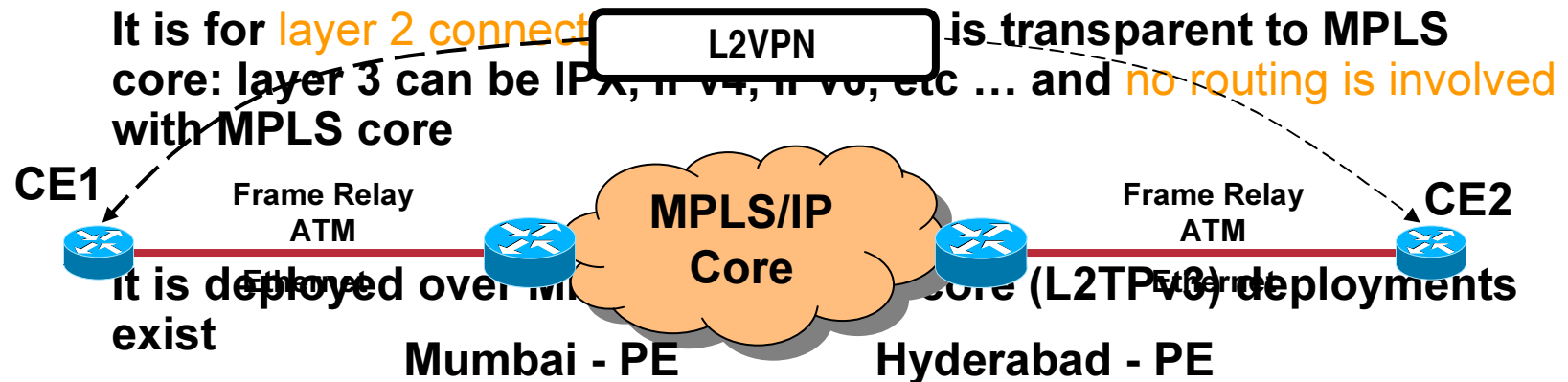
Introduction to L2VPNs



What is L2VPN?

- L2VPN provides an **end-to-end layer 2 connection** to an enterprise office in Mumbai and Hyderabad over a SP's MPLS or IP core

It can be Ethernet, Frame Relay, ATM, HDLC, PPP, etc ...



Why is L2VPN needed?

- Allows SP to have a **single infrastructure** for both IP and legacy services

Migrate legacy ATM and Frame Relay services to MPLS/IP core without interruption to existing services

Provisioning new L2VPN services is **incremental (not from scratch)** in existing MPLS/IP core

Capital and Operational savings of converged IP/MPLS network

- SP provides new **point-2-point** or **point-2-multipoint** services
Customer can have their own routing, qos policies, security mechanisms, etc ...
- Based on IETF drafts that promote open architecture and vendor interoperability

Layer 3 and Layer 2 VPN Characteristics

LAYER 3 VPNS

- SP devices forward customer packets based on **Layer 3 information** (e.g. IP addresses)
- SP is involved in customer IP routing
- Support for **any access** or backbone technology
- **IP** specific
- **Foundation for L4–7 services!**
- Example: RFC 2547bis VPNs (L3 MPLS-VPN)

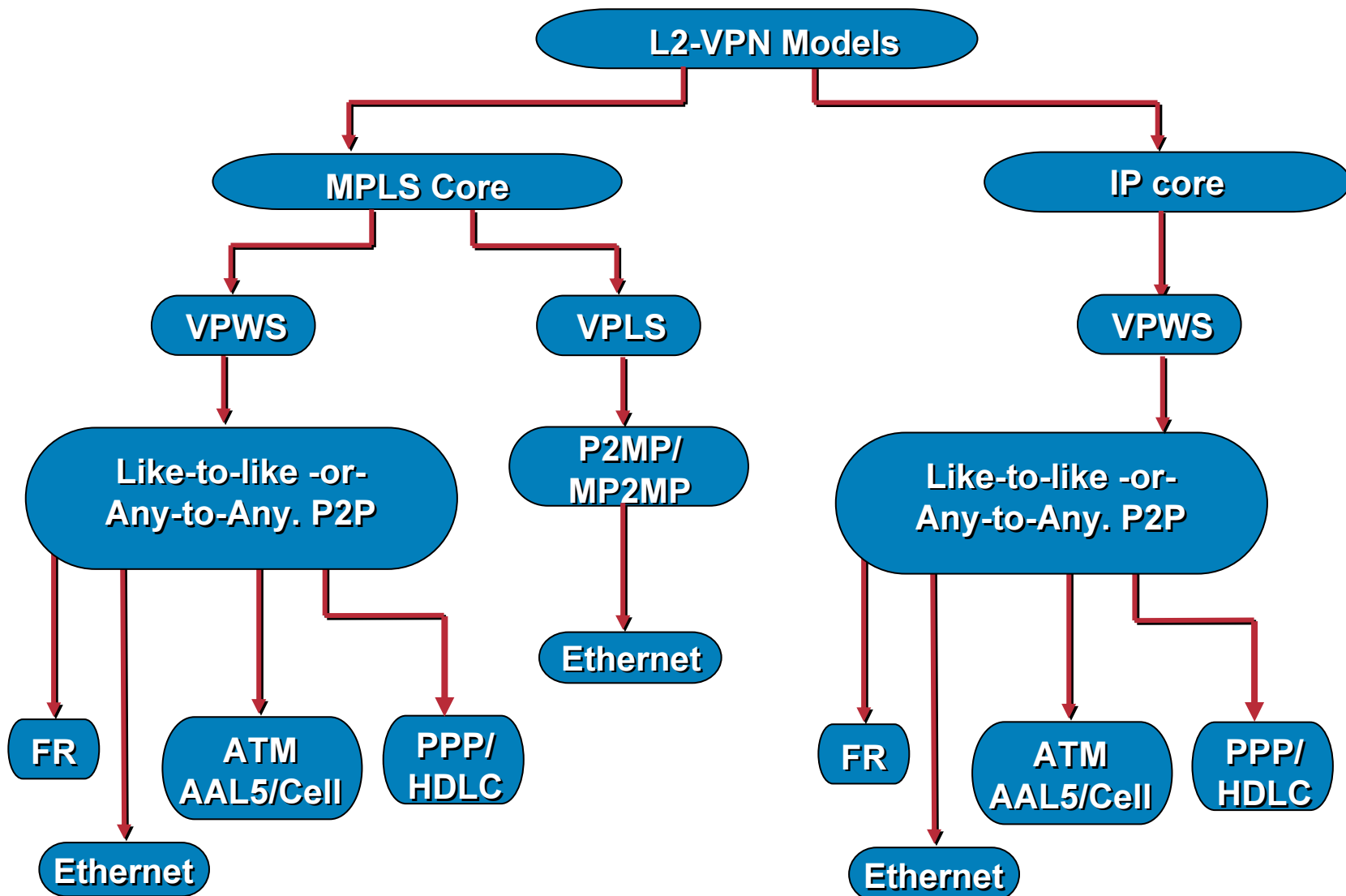
LAYER 2 VPNS

- SP devices forward customer frames based on **Layer 2 information** (e.g. DLCI, VPI/VCI, MAC, VLAN ID)
- No SP involvement in customer IP routing
- Enterprise stays in **control** of L3 policies (Routing, QoS)
- **Multiprotocol** support
- Example: FR—ATM—Ethernet

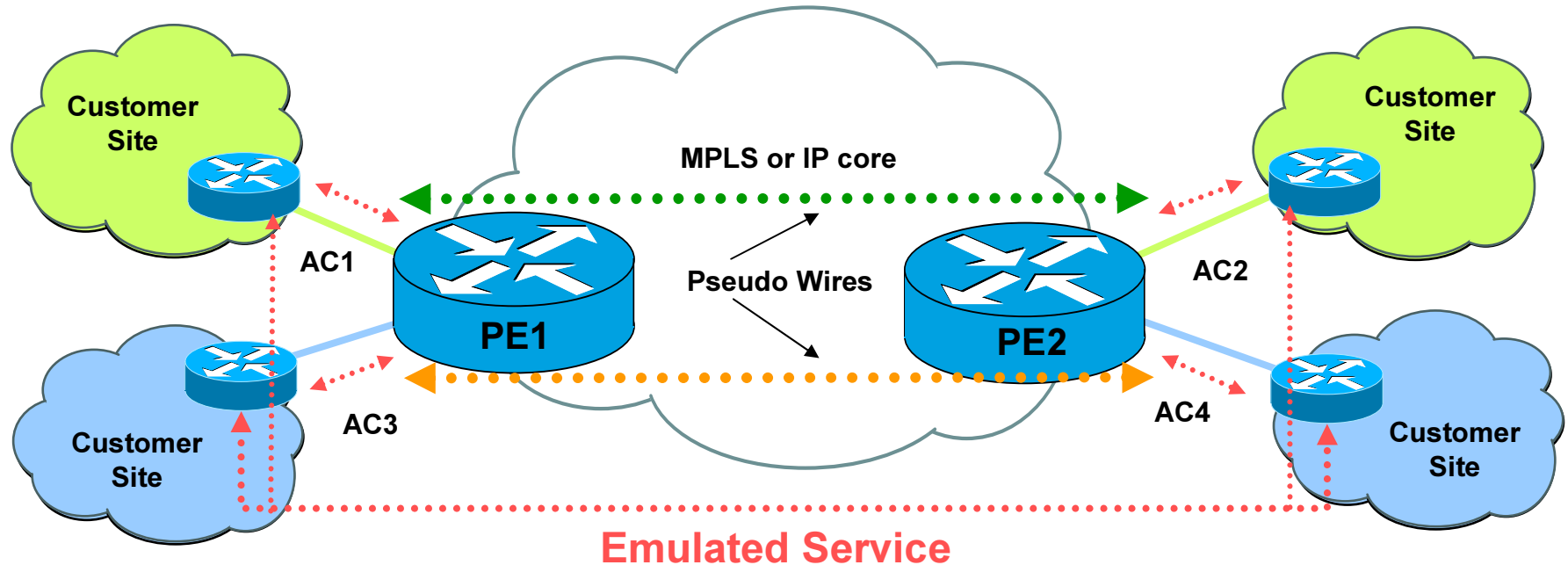
The Choice of L2VPN over L3VPN Will Depend on **How Much Control** the Enterprise Wants to Retain.

L2 VPN Services Are **Complementary** to L3 VPN Services

L2VPN Models



VPWS - Pseudo Wire Reference Model



A pseudo-wire (PW) is a connection between two provider edge (PE) devices which connects two attachment circuits (ACs).

Emulated Services:

- Ethernet
- 802.1Q (VLAN)
- ATM VC or VP
- HDLC
- PPP
- Frame Relay PVC

Pseudo Wire – Basic Building Blocks

**Control
Connection**

Adds scale thru: **Session Management, Error Notification, L2 Access management interworking, etc.**

Required Components

**Transport
Component**

This is the delivery header of the encapsulated packet. This can be a Label (MPLS) or an IP Header. (Typically the IP address of the Loopback interface on Provider Edge (PE) routers.)

**Tunneling
Component**

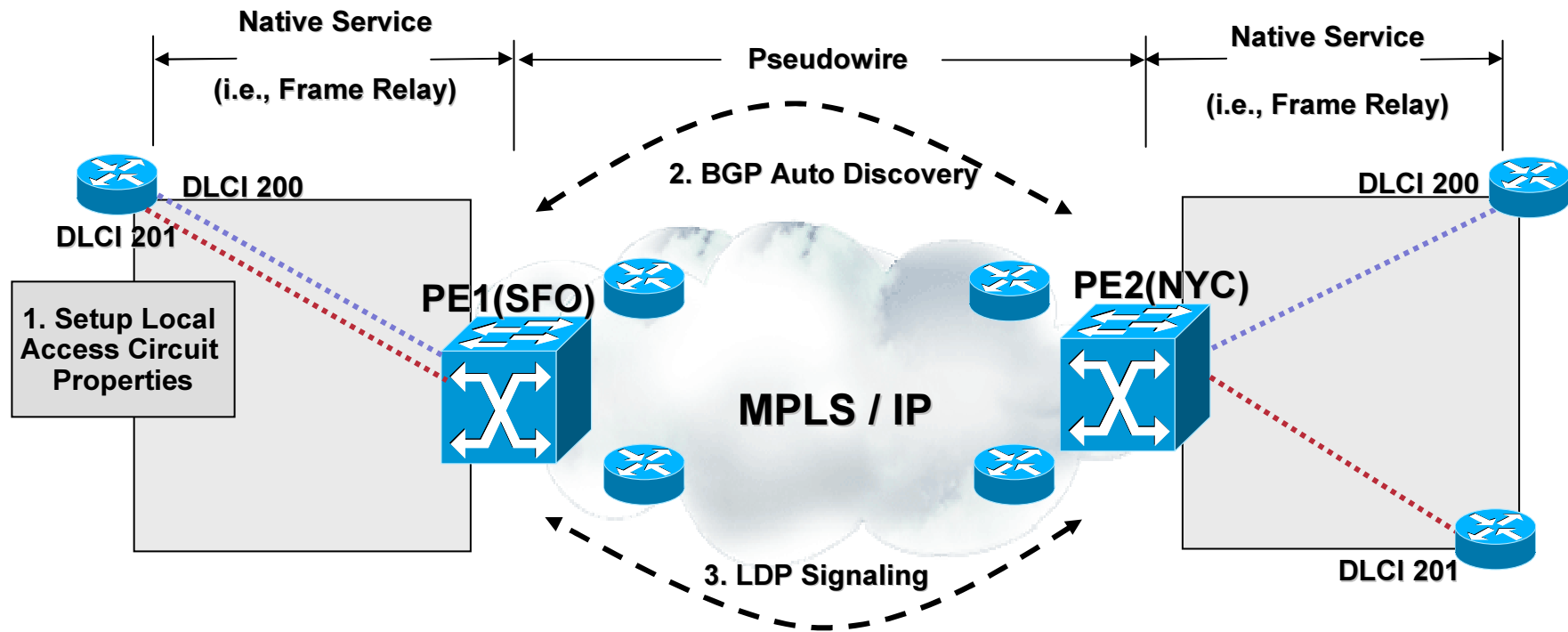
A Unique identifier used to identify a particular circuit / port on a given PE. (VC Label or VC ID)

L2 PDU

The Layer 2 PDU that is the subject of transport (I.e. traffic received from the Customer Edge router, typically Ethernet, Frame Relay, HDLC frames,..etc.)

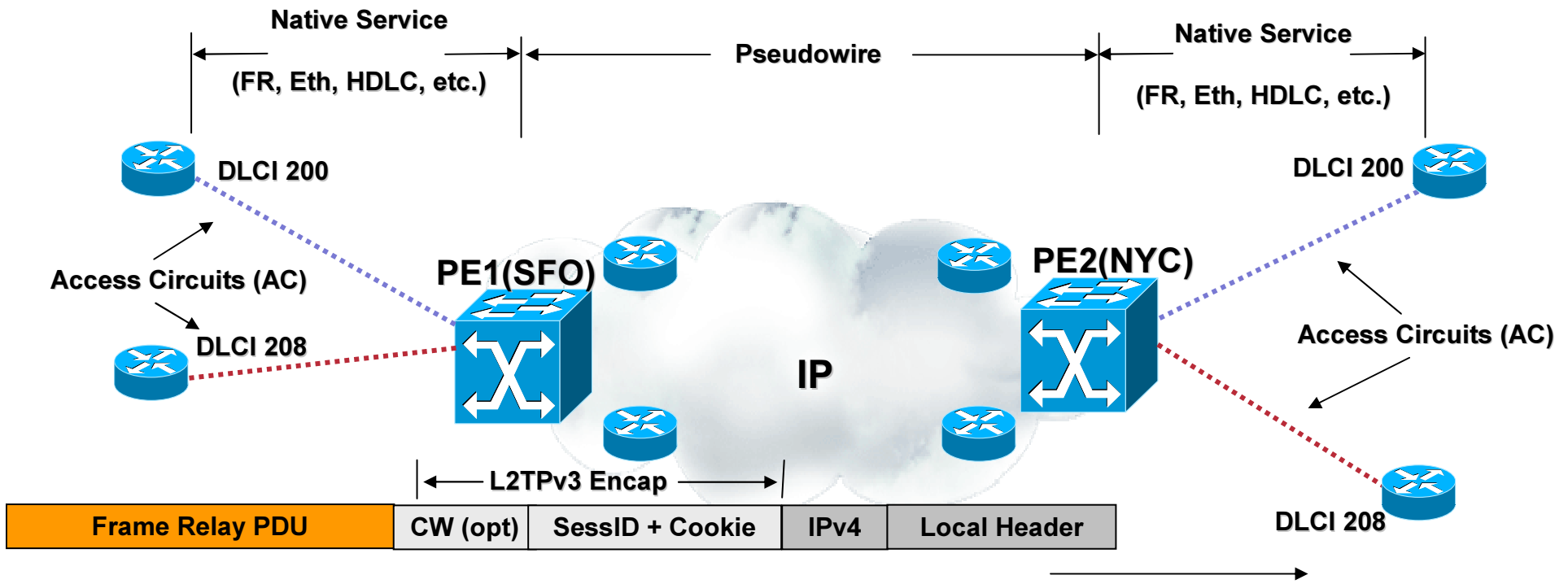
“Connectivity between PEs assumed; verified through ICMP or LSP ping.”

Building Blocks for L2VPNs – Control Plane, Generic Requirements (MPLS Example)



- The appearance of **Native Service** support from CE to CE (SP network is transparent)
- Ability to **Discover** other PE members for a given VPNs (i.e., BGP)
- Requires **Signaling** and **Interwork** with native services (i.e., Frame Relay LMI)
- **Packet Forwarding** – Negotiating the appropriate VC Label / Session ID

Building Blocks for L2VPNs – Data Plan Components – IP Core (IP Example)



Transport Header

Delivery Header (IPv4 Header) = Transport an L2 PDU from ingress to egress PE; comprised of IPv4 loopback addresses (DA, SA)

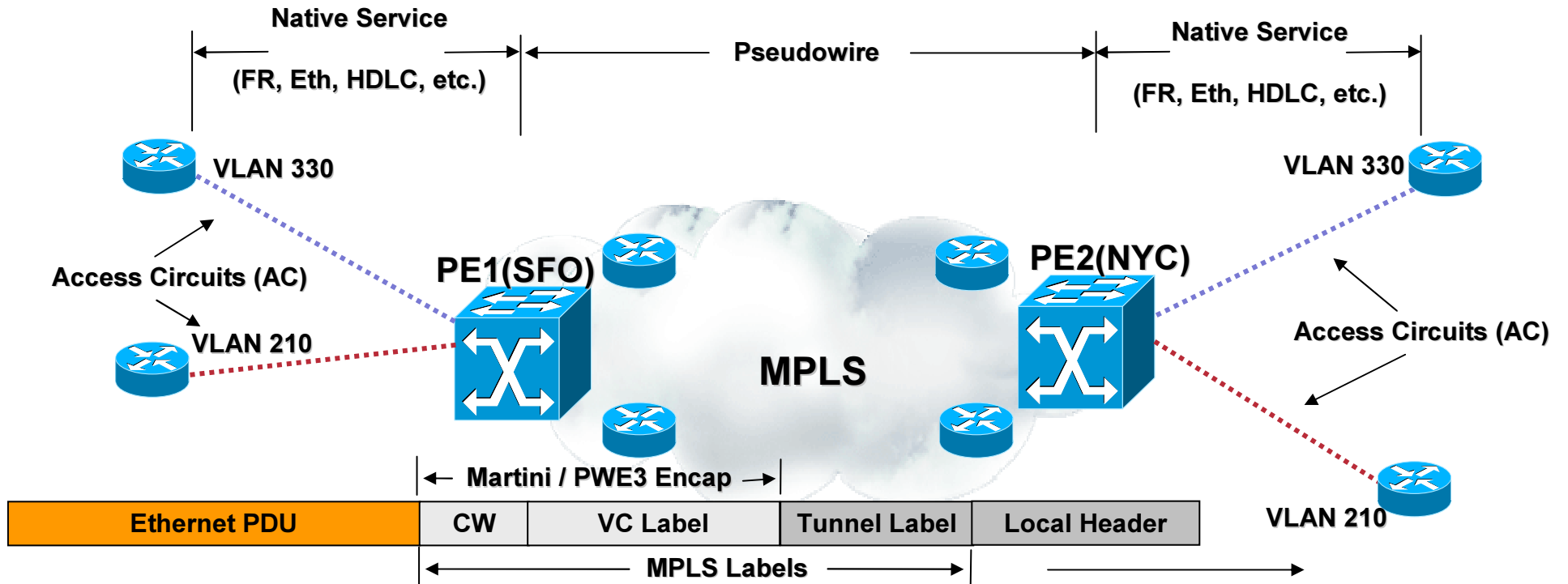
Tunneling Header

4 byte Session ID + Optional 8 byte Cookie = Signalled or Statically configured

L2 PDU

L2 Specific Sublayer + Customer Payload (Layer 2 PDU)

Building Blocks for L2VPNs – Data Plan Components – MPLS Core (Ethernet)



Transport Header

Tunnel Label (MPLS Label) (4 Byte) = Established thru LDP link establishment or RSPV-TE signalling. Forms a uni-directional path to the destination PE

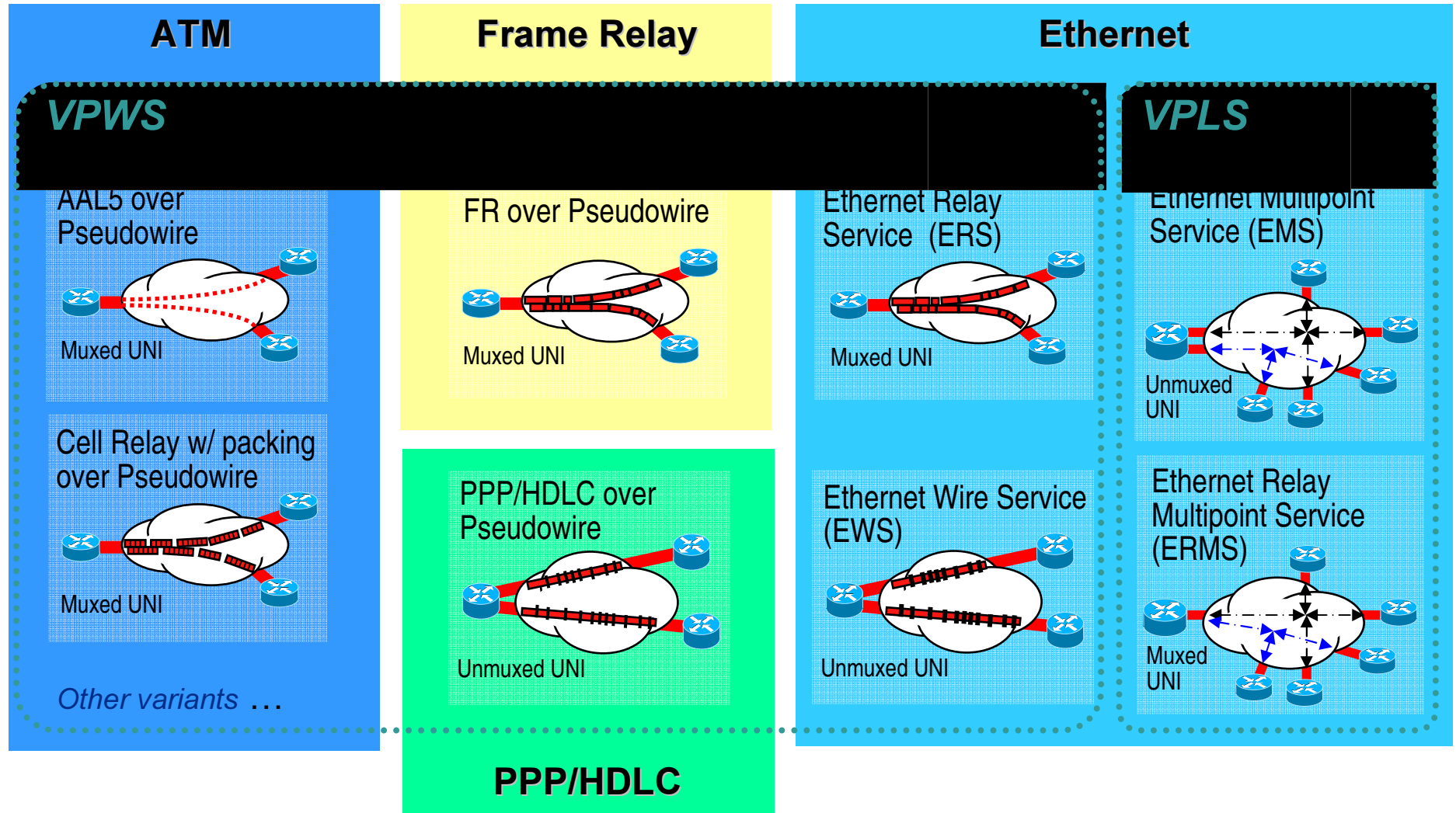
Tunneling Header

VC Label (MPLS Label) (4 Byte) = Signalled thru Extended LDP session established between PE pairs (Martini ; PWE3 based drafts)

L2 PDU

Control Word (opt.) (4 Byte) + Customer Payload (may not transport entire L2 header)

Service Offerings – L2VPN Transport Services



IETF Standardization Activity

- **IETF working group PWE3**

‘Pseudo Wire Emulation Edge to Edge’;

Requirements detailed in

- *draft-ietf-pwe3-requirements*
- *draft-ietf-pwe3-framework*

- **Develop standards for the encapsulation & service emulation of “pseudo wires”**

Across a packet switched backbone

- **Focused on Point-to-Point circuit emulation**

PSN tunnel -> GRE, MPLS, L2TP

Service -> Ethernet, ATM, PPP, FR, HDLC and so on ..

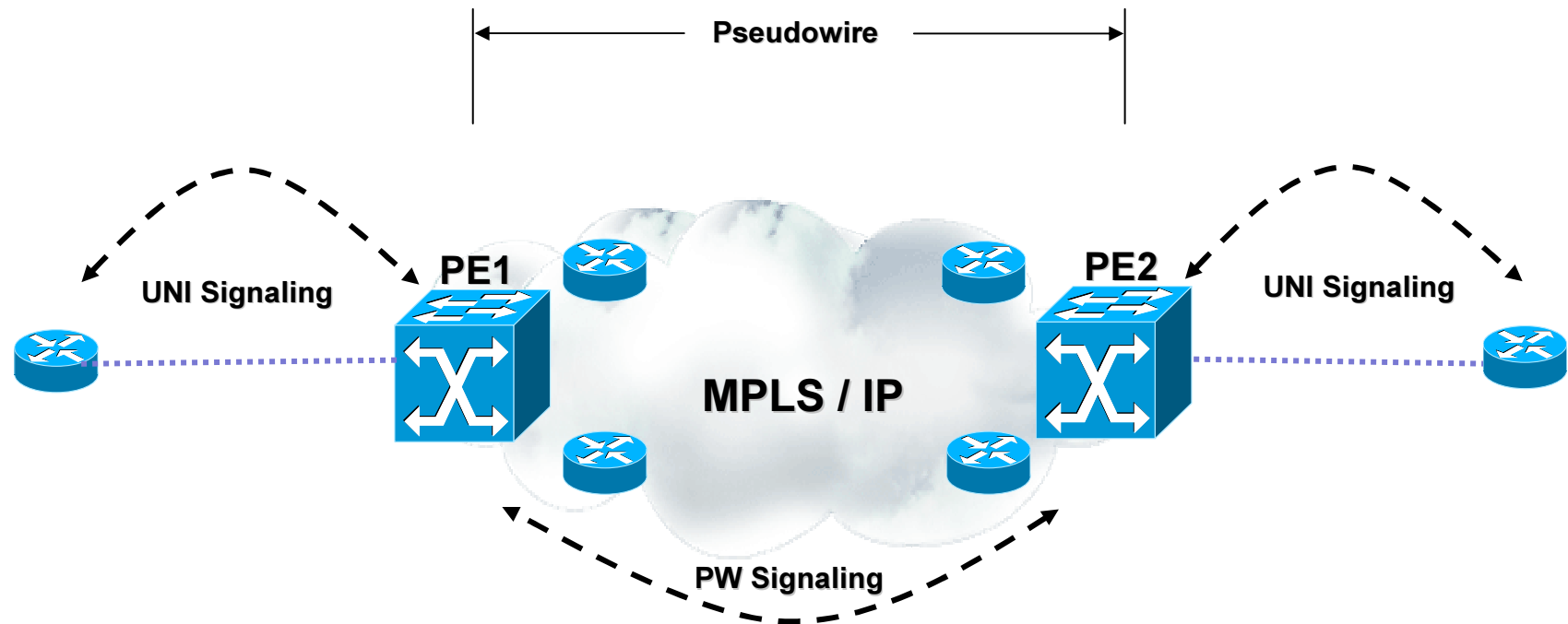
Pseudo Wire – *Cisco IETF Technology Adoption*

- **Layer 2 Transport (VPWS)**
 - **L2TPv3**
 - draft-ietf-l2tpext-l2tp-base-xx
 - draft-ietf-l2tpext-l2tpmib-base-xx
 - **MPLS (P2P, formerly draft-martini)**
 - draft-ietf-pwe3-control-protocol-xx
 - draft-ietf-pwe3-[atm, frame-relay, ethernet, etc.]
- **Layer 2 VPN (VPLS)**
 - draft-ietf-l2vpn-vpls-ldp-xx (LDP Based Signalling)
- **Auto-Provisioning**
 - draft-ietf-ppvpn-bgpvpn-auto-xx (BGP auto-discovery)

Signaling Concepts

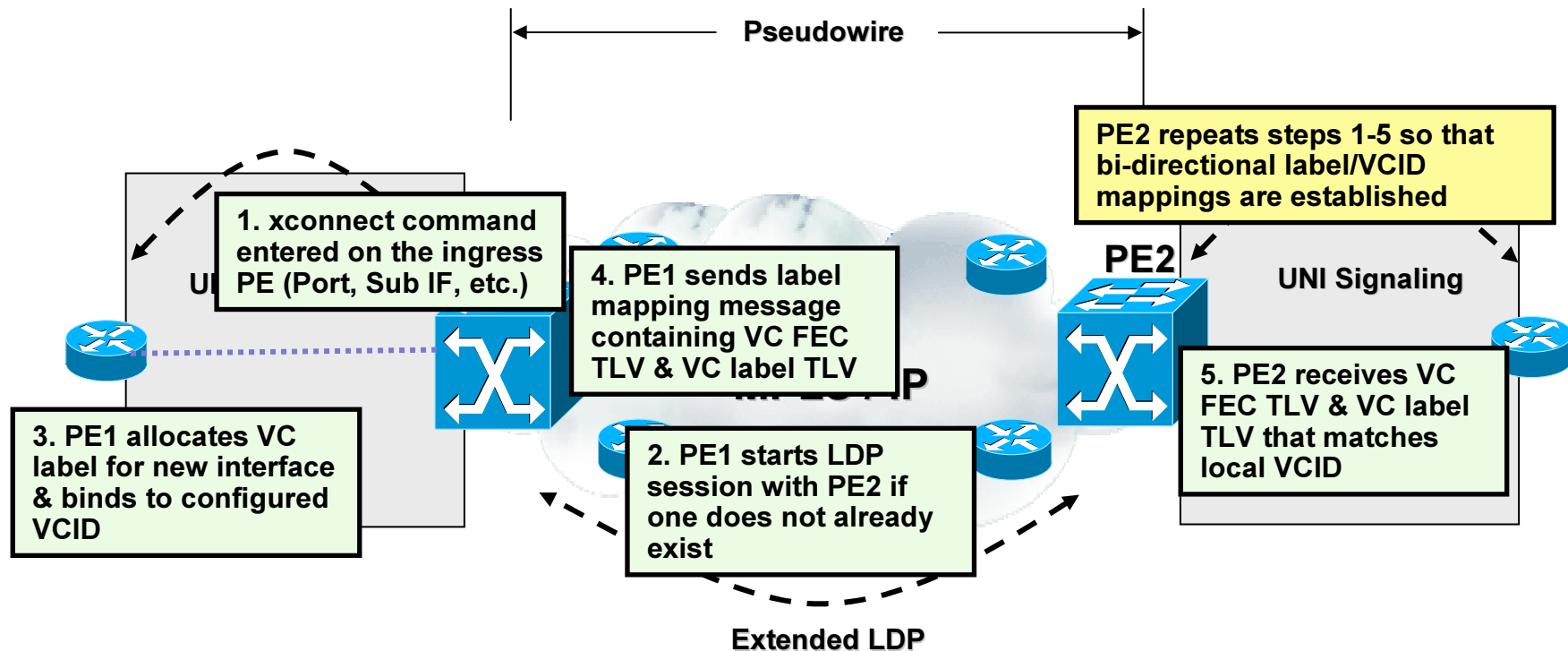


Virtual Private Wire Service – *Provisioning – Overview*



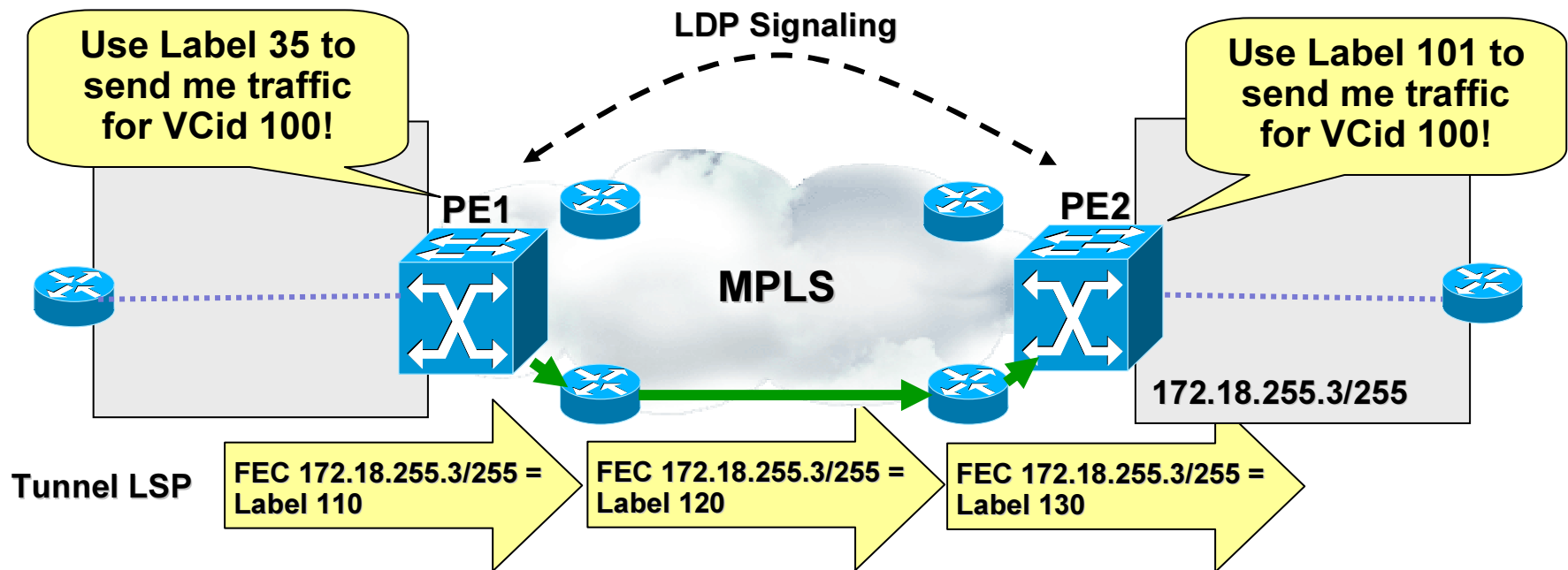
- Establishment of a control connection (LCCE Signaling or LDP)
- Provide an interface to local UNI to (activate, deactivate, delete) attachment circuits (ACs)
- Negotiate Session IDs (L2TPv3) & PW labels (LDP) between PEs.

Any Transport over MPLS – Provisioning



- One LDP session can signal multiple pseudowires
- Provides a dynamic mechanism to interface with UNI signaling
- Requires a common VCID to successfully bind ACs together.
- Pseudowire (VC) labels are assigned by the remote peer and don't require global uniqueness between PEs.

Virtual Private Wire Service – LDP Signaling – Forwarding Equivalence Class



- **FEC = “Forwarding Equivalence Class”** : a set of packets forwarded in the same manner by an MPLS LSR
 - L2VPNs – FEC is used to bind a PW label to VC ID / Service type*
 - L3VPNs – FEC is used to bind a VPN label to set of customer prefixes
- **Used to MUX customer data onto a given Tunnel**
 - Similar concept to DLCI, VPI/VCI in Frame and ATM networks

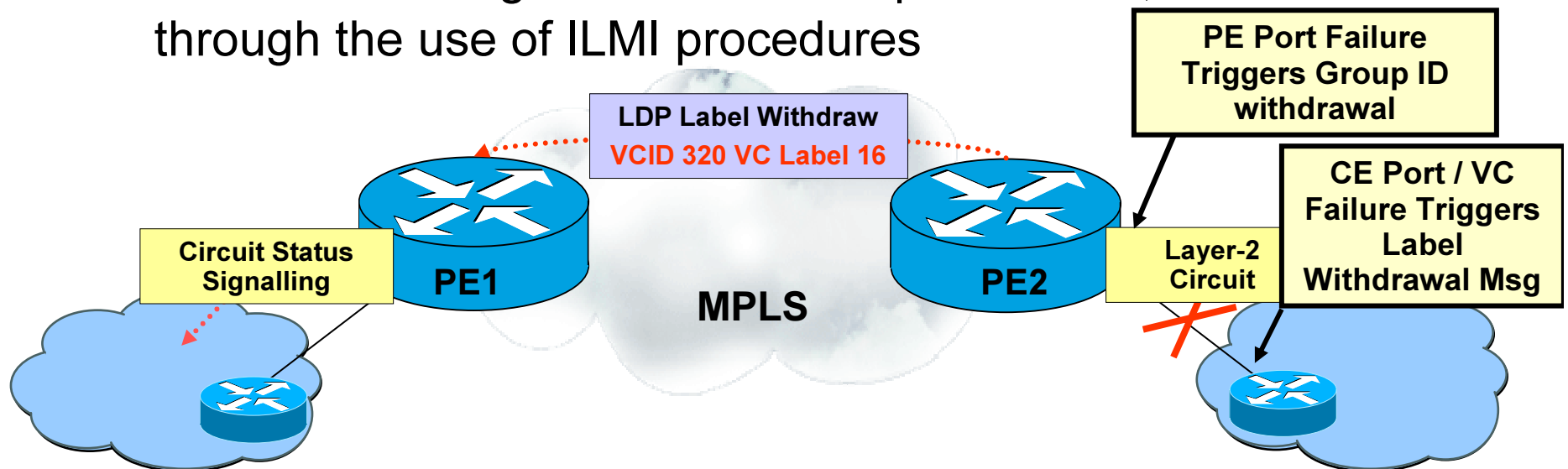
Virtual Private Wire Service– VC Label Withdrawal Procedures - Example

- If a PE router detects a condition that affects normal service it **MUST** withdraw the corresponding VC label

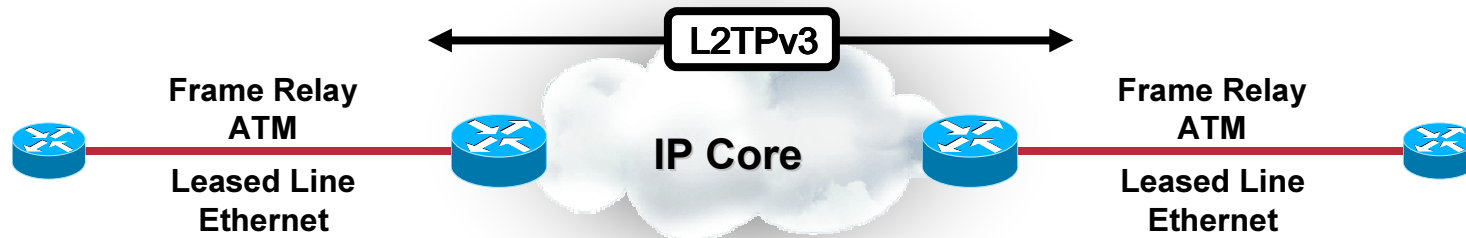
Through the use of LDP signalling

- A PE router may provide circuit status signalling

FR **MUST** through the use of LMI procedures; ATM **SHOULD** through the use of ILMI procedures

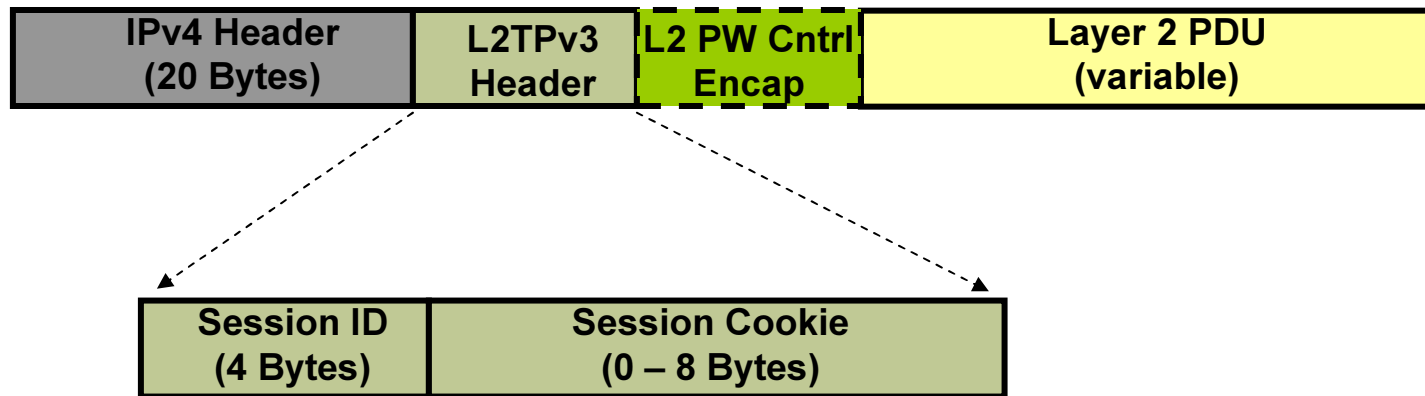


Layer 2 Tunneling Protocol version 3 – *The Native IP Service Provider's Answer to VPWS*



- **Designed for Service Providers with Native IP networks**
- **Based on L2TP (used in Remote Access) & Cisco innovation Universal Transport Interface (UTI)**
- **Open “standards track” architecture allows for extensibility**
- **Fixed header allows for high performance / HW accelerated decapsulation**
- **With IGP reachability, simple edge configuration is the only thing required!**

Layer 2 Tunneling Protocol version 3 – *Data Messages Format*



IPv4 Header - The delivery header for the Tunnel. Always destined for an LCCE.

L2TPv3 header – Consists of two parts; (1) **Session ID** used to uniquely identify the correct Session on the Remote system, and (2) the **Cookie** used as an added measure of session integrity or validation between peers.

L2-Specific Sublayer - Sequence numbers, priority bits, and any additional flags needed to support the L2 emulation for the given PW type. There is a default defined in the L2TPv3 base specification, though this may vary among PW types if necessary.

Payload - Payload to be transported by L2TPv3. Typically the entire link-level frame.

Virtual Private Wire Service – Summary

- **MPLS PW & L2TPv3** are point-to point technologies for the transport of Layer 2 PDUs across and **native** or **MPLS** enabled **IP** cores.
- **MPLS PW** uses **Directed LDP** sessions to exchange **PW Labels** between participating peers while **L2TPv3** uses **Control messages** to negotiate **Session IDs**
- **MPLS PW & L2TPv3** can use an optional **Control Word** to preserve information in transported PDUs
- **MPLS PW & L2TPv3** provides **interworking** with access circuit management protocols to maintain **VC** status consistency (i.e. label withdrawal or call disconnect notification in the event of edge service loss, etc.)

VPWS Transports

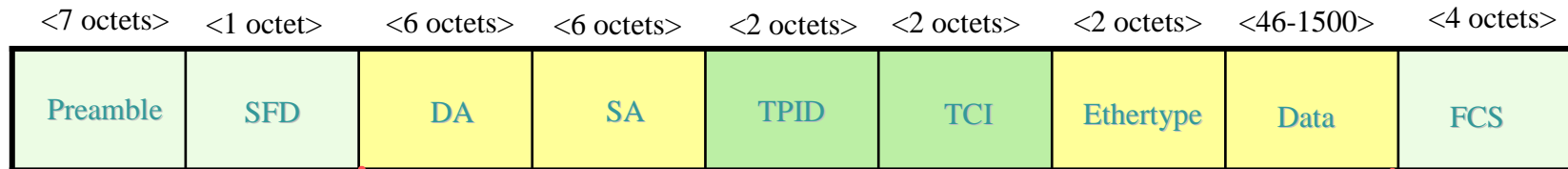


VPWS Transports

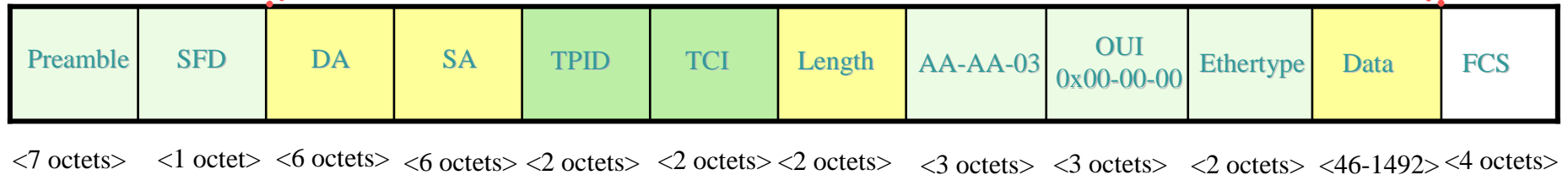
- **EoMPLS**
- **FRoMPLS**
- **ATMoMPLS**
- **PPPoMPLS**
- **HDLCoMPLS**

EoMPLS Transport Formats

Ethernet II Encapsulation



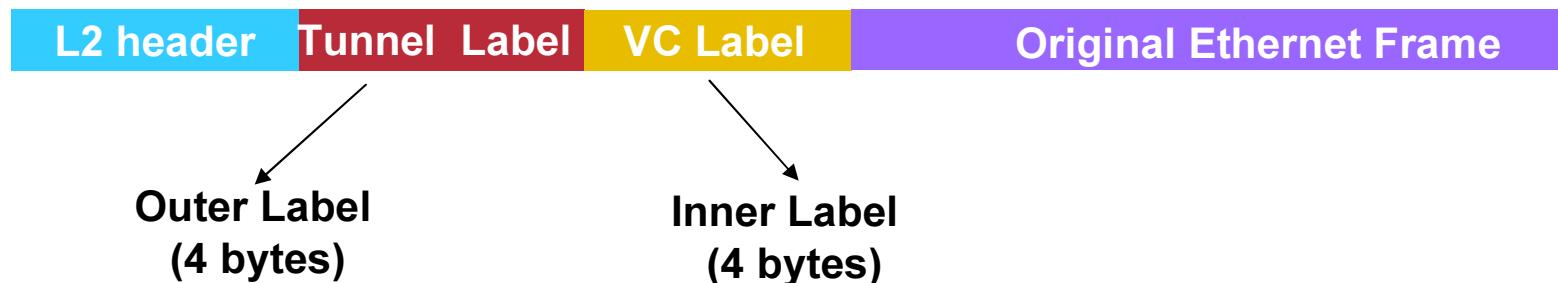
Transported using AToM



802.3/802.2/SNAP Encapsulation

EoMPLS Encapsulation

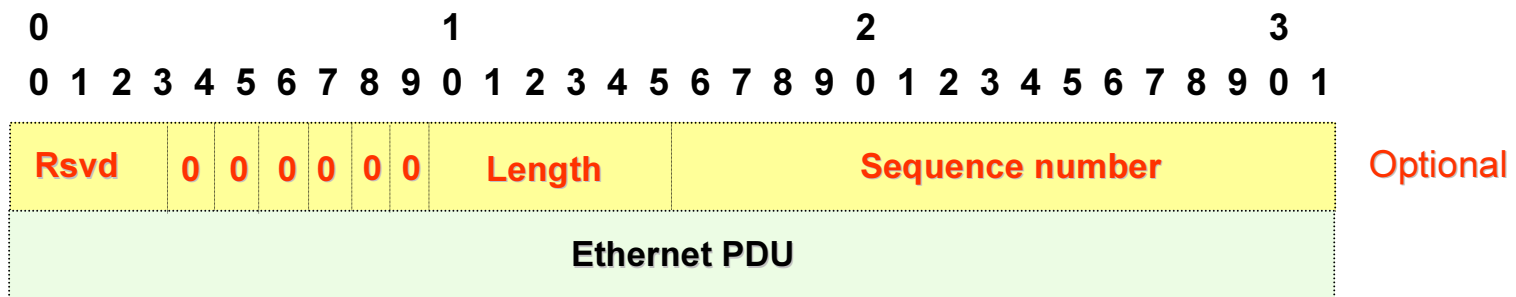
- PE router encapsulates VLAN packet and route it across MPLS backbone
- Two levels of labels (8 bytes)
 - Tunnel label, outer label, to forward the packet across the network
 - VC-based label, to bind L2 interface where packets must be forwarded
- VC (virtual circuit), 32 bit identifier used uniquely to identify the VC per tunnel
- VC type-0x0004 is used for VLAN over MPLS application
- VC type-0x0005 is used for Ethernet port tunneling application (port transparency)
- VC is an label switch path (LSP) tunnel



EoMPLS Encapsulation Details

- **Ethernet PDUs are transported without the preamble, SFD and FCS**
 - but including all VLAN information such as VCID
- **The control word is optional**
 - C bit is set by default in Cisco implementation (except 7600)
- **If the control word is used then the flags must be set to zero**

The VLAN tag is transmitted unchanged but may be overwritten by the egress PE router

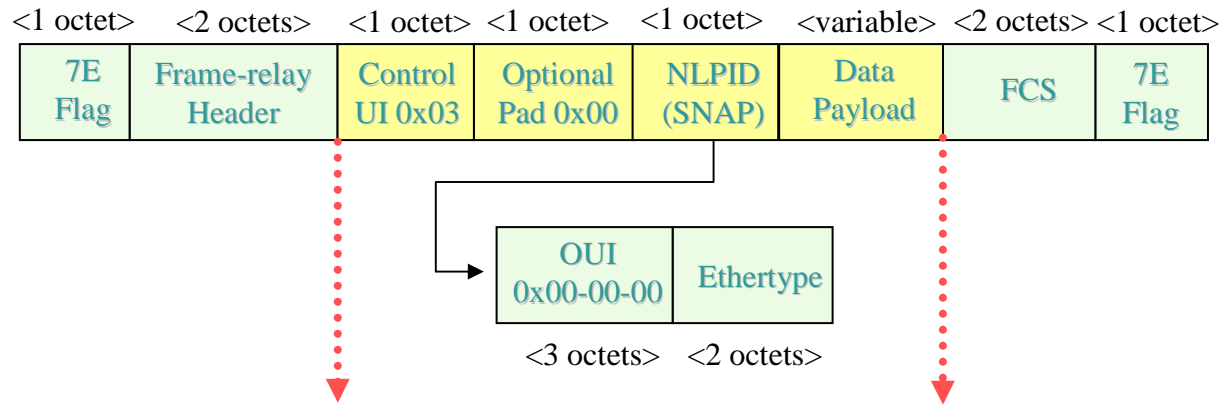


VPWS Transports

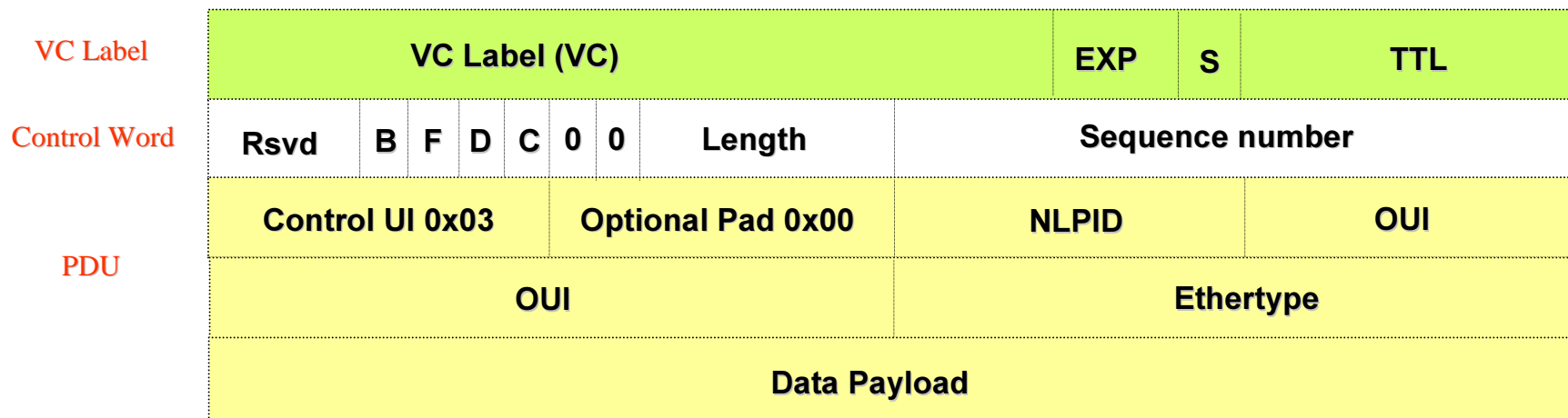
- **EoMPLS**
- **FRoMPLS**
- **ATMoMPLS**
- **PPPoMPLS**
- **HDLCoMPLS**

RFC 1490 Encapsulation

RFC 1490 Frame Relay Encapsulation

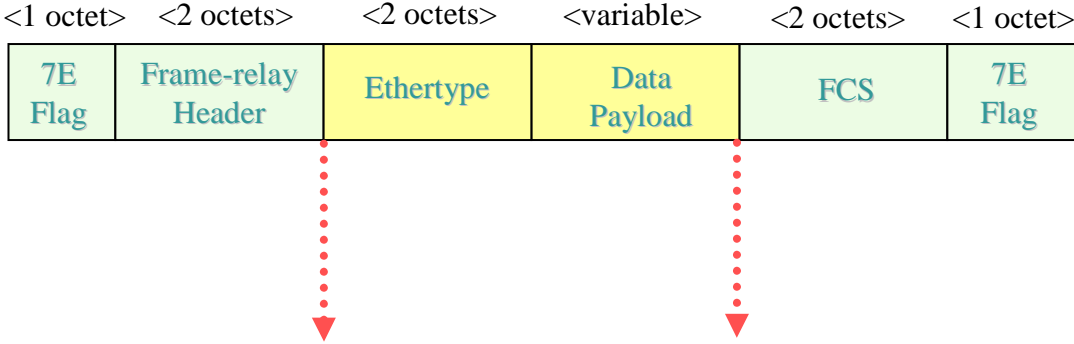


Transported using AToM

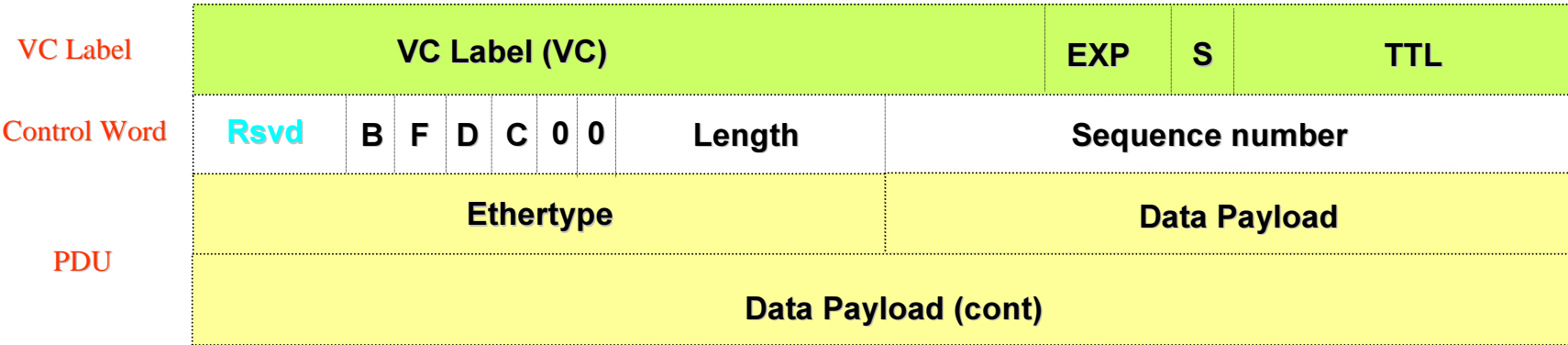


Cisco Proprietary Encapsulation

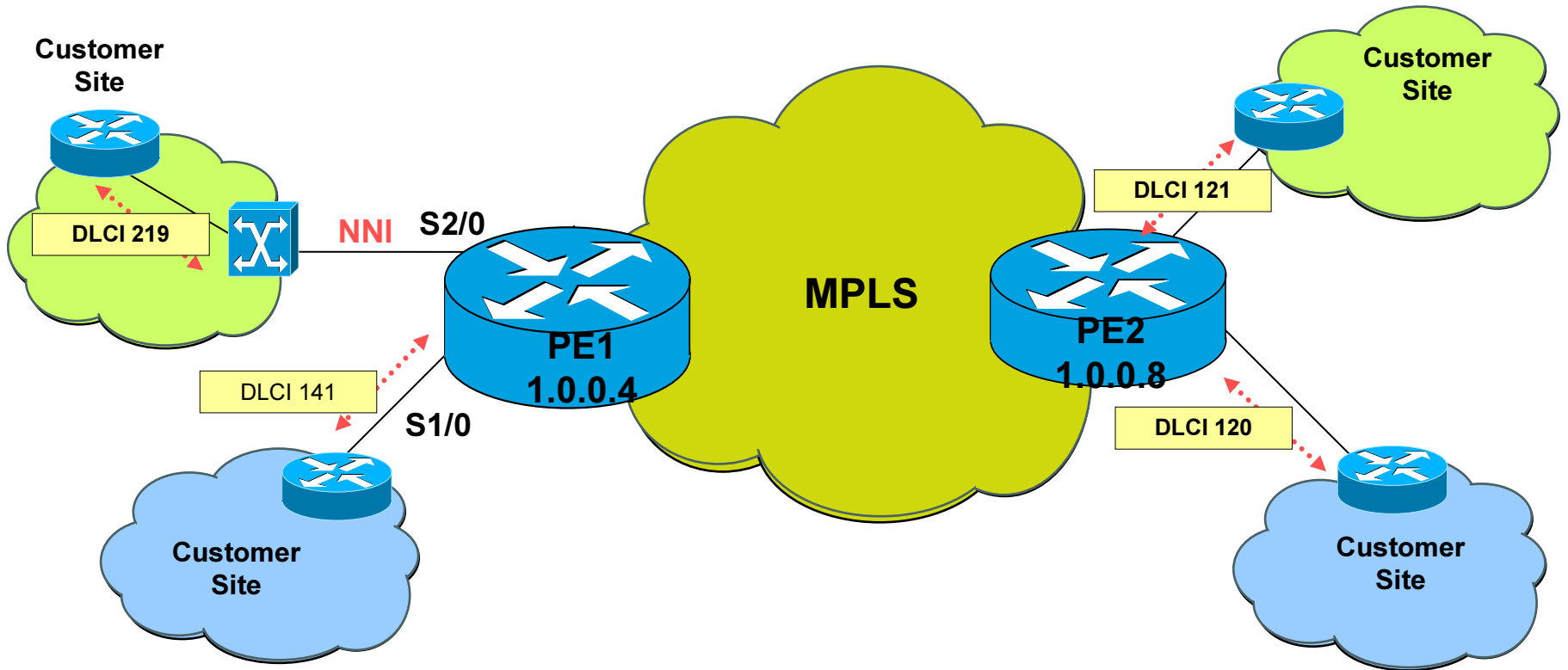
Cisco Frame Relay Encapsulation



Transported using AToM



Distributed DLCI Switching



Frame-relay DLCI to DLCI Transport

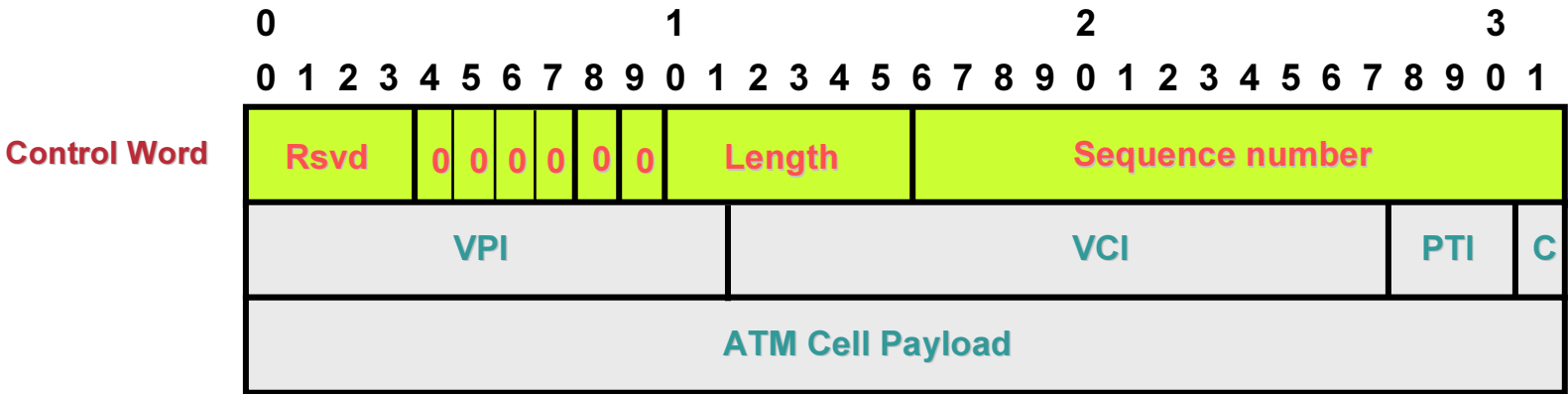
VPWS Transports

- EoMPLS
- FRoMPLS
- **ATMoMPLS**
- PPPoMPLS
- HDLCoMPLS

Cell Relay Over MPLS

- **Single Cell per MPLS packet**
 - Applying labels per cell
 - Do not distinguish between payload cells and signaling
 - OAM and RM also transported
- **Control Word is optional**
- **HEC is not carried inside MPLS network**
- **Idle cells are not carried over MPLS network**

Cell Relay Encapsulation Details



- Single cell is encapsulated
- Control word is optional
- Control word flags should be set to zero and ignored

ATM Cell Packing – Why Important?

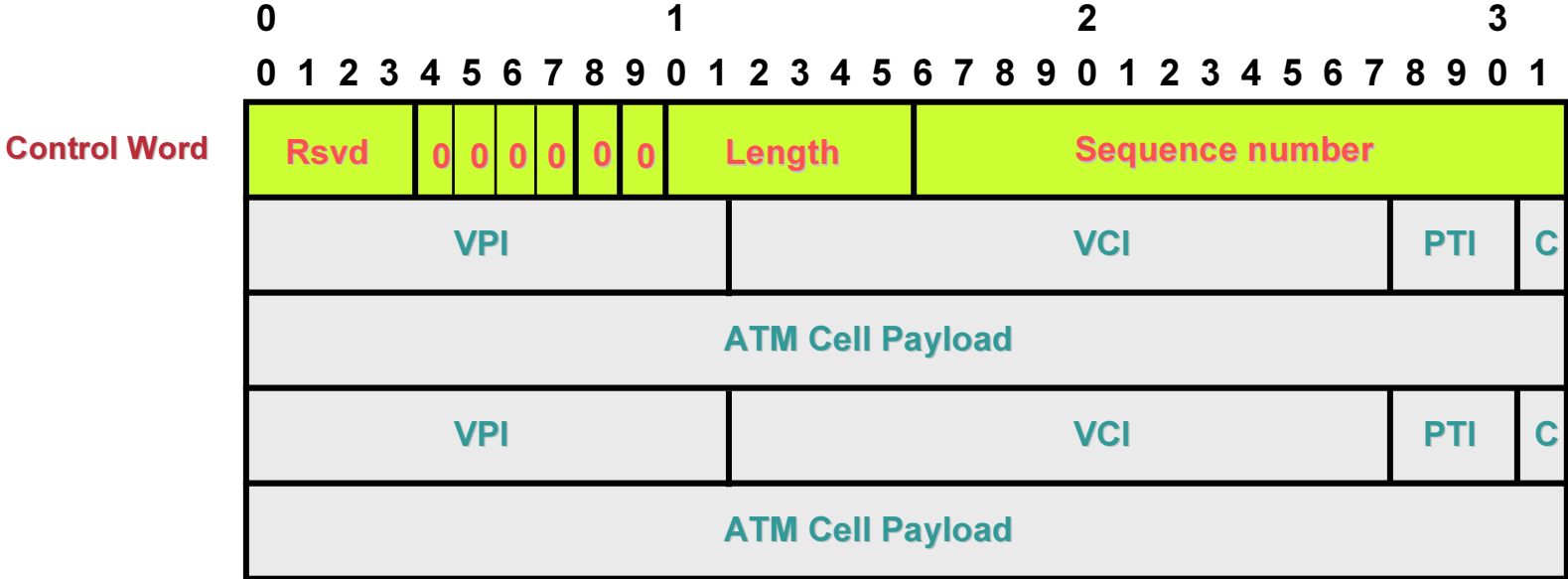
Overview:

- Used to mitigate Cell to MPLS Packet MTU inefficiencies
- Concatenated ATM Cell (52 Bytes); No HEC
- Maximum 28 Cells per MPLS Frame (<1500 byte MTU)

Components:

- Maximum cells to pack
- Maximum Cell Packing Timeout (MCPT)
- PE will send packed cells based on the minimum configured packing characteristics between the two PEs. Each PE honors the MNCP configured on the other.

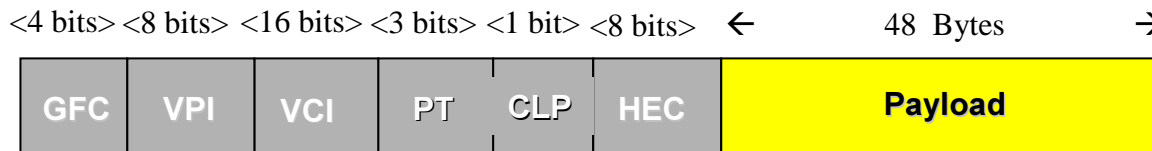
Cell Packing Encapsulation Details



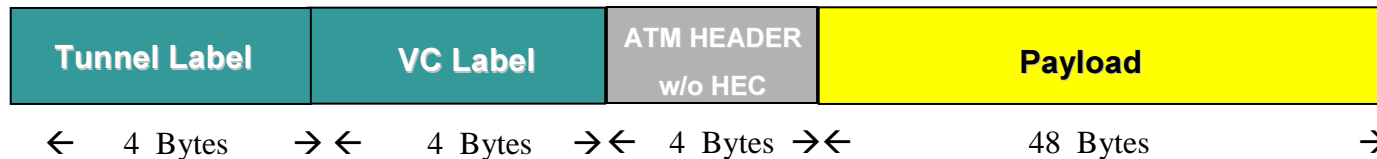
- Multiple cells are packed per MPLS packet
- All cells must belong to the same VC / VP
- Packing controlled by max number of cells and timer
- Control word is optional

Cell Packing Encapsulation

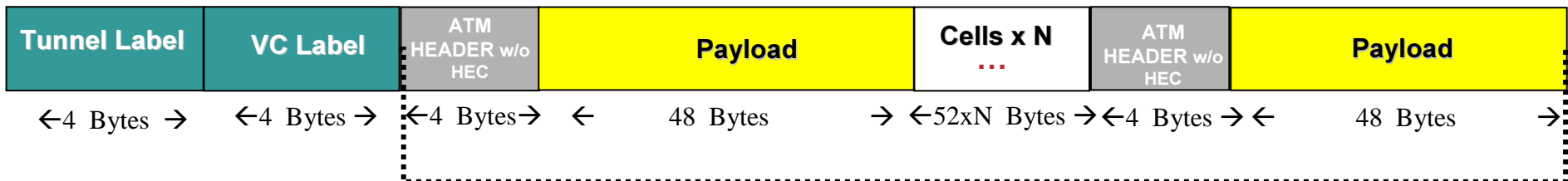
- **ATM Cell**



- **Single Cell Relay**



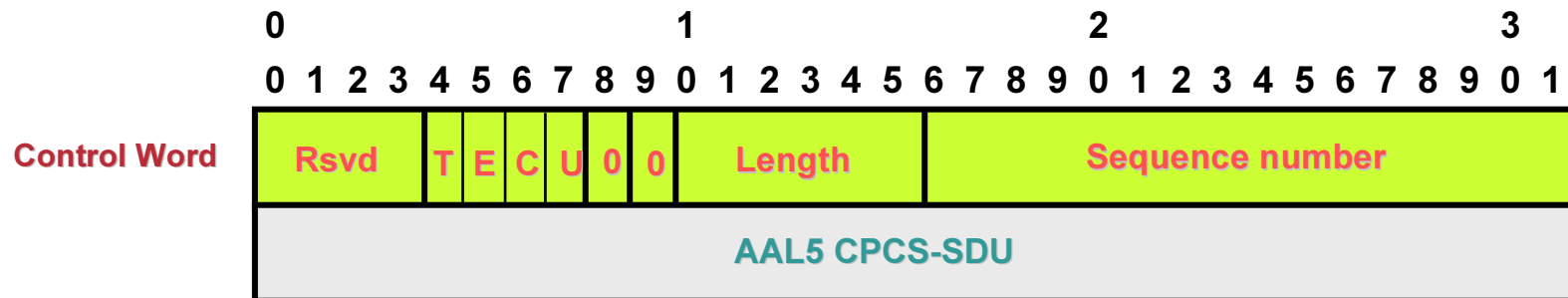
- **Packed Cell Relay**



PACKED CELLS MAX 28

28*52=1456 bytes

AAL5 Encapsulation Details



- AAL5 SDUs are encapsulated
- Control word is required
- Control word flags encapsulate transport type, EFCI, CLP, C/R bit
- Service allows transport of OAM and RM cells

VPWS Transports

- **EoMPLS**
- **FRoMPLS**
- **ATMoMPLS**
- **PPPoMPLS & HDLCoMPLS**

Cisco HDLC & PPP Encapsulation Details

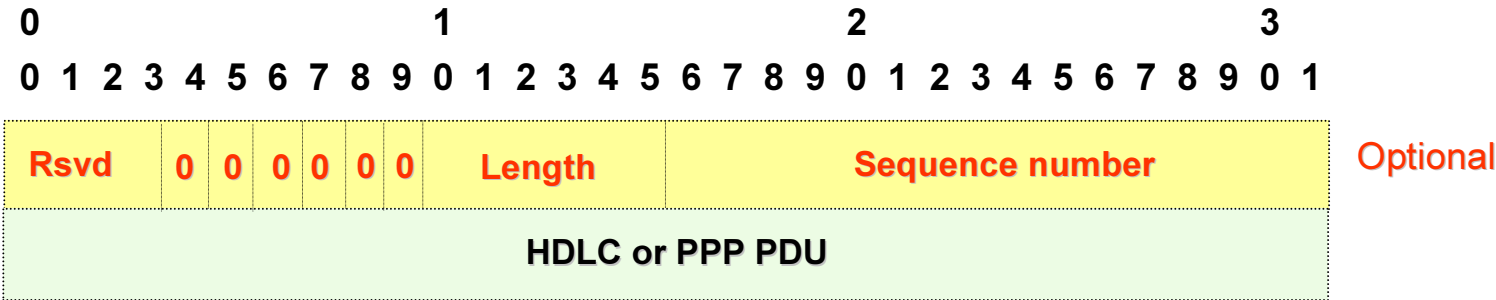
- **Cisco HDLC & PPP PDUs are transported without flags or FCS**

PPP frames also do not carry HDLC address & control information

- **The control word is optional**

C bit is set by default in Cisco Implementation

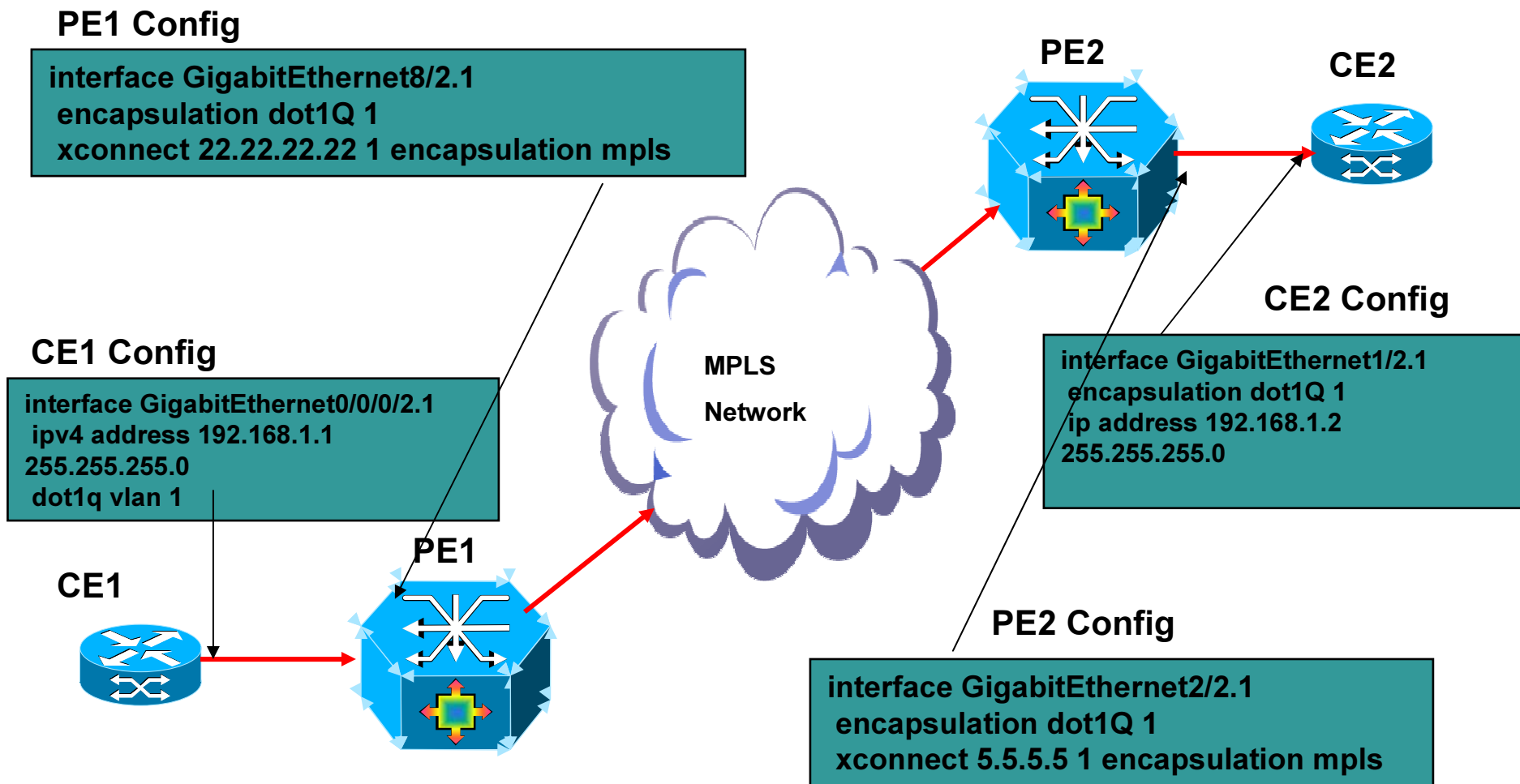
- **If the control word is used then the flags must be set to zero**



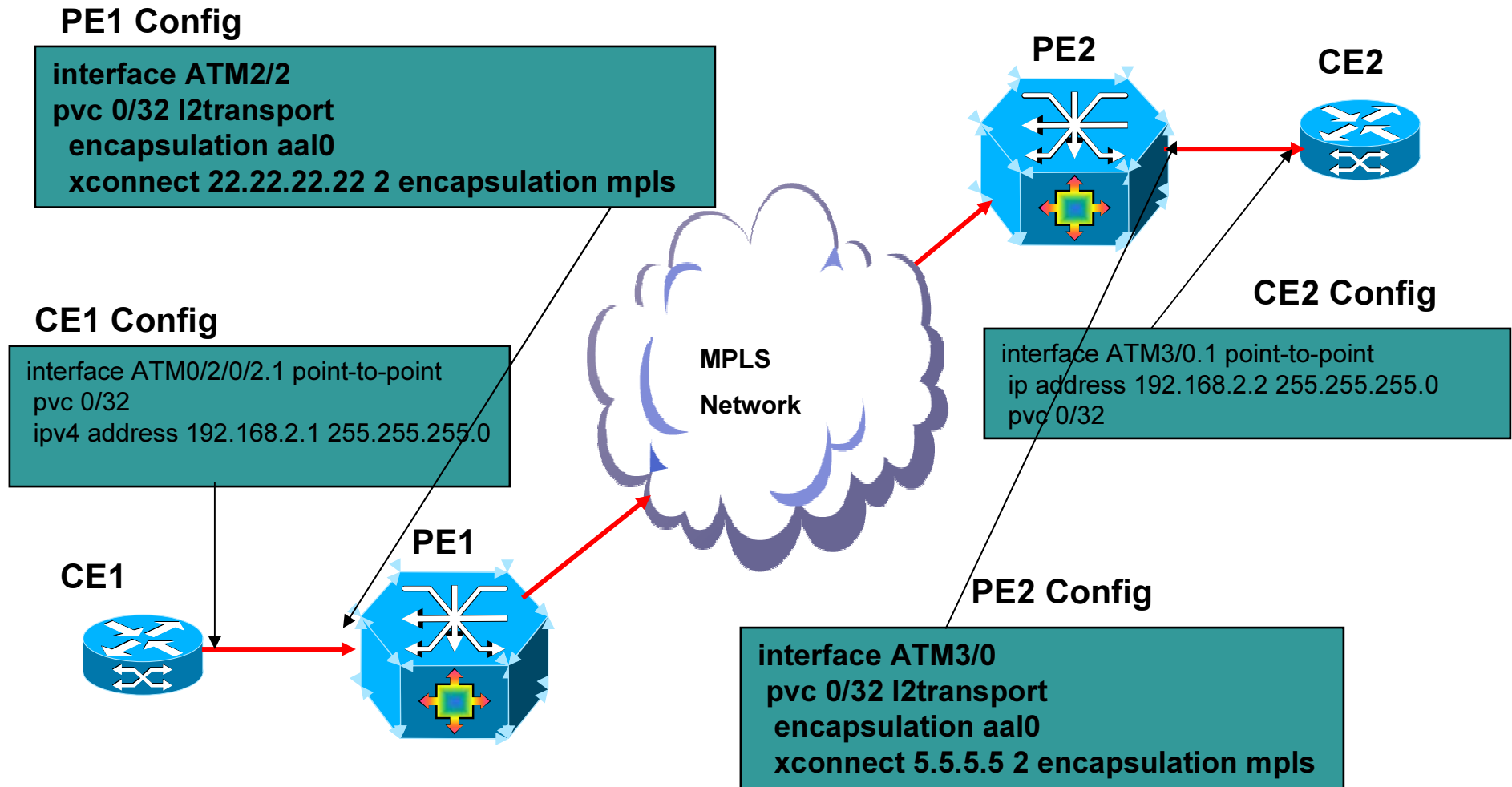
DEMO



Point to Point VLAN over MPLS



Point to Point Cell Relay over MPLS



|

VPWS Service Interworking

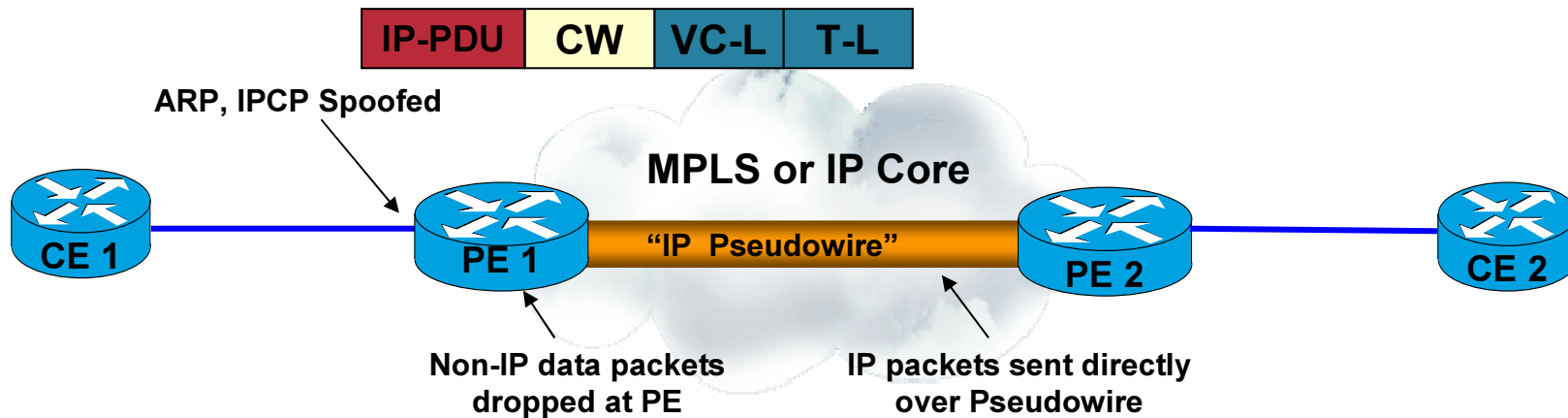


L2VPN Service Interworking – *Overview – What is it's purpose?*

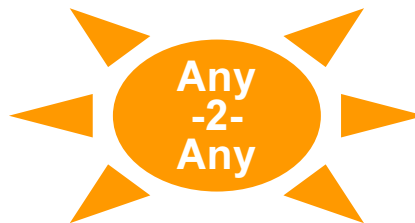
L2VPN Service Interworking enables:

- Flexibility for the Service Provider in offering access circuits that best meet the customer's needs (i.e. multiple services on a common port)
- Expanded transport options from **Like-to-Like** → **Any-to-Any**
- Allows different Layer 2 encapsulations to connect at opposite ends of the network. (i.e. ATM PVC → Ethernet 802.1Q VLAN)
- Provide circuit based services in addition to packet based services
- Transparent trunking of customer IGP independent of access media

IP Interworking

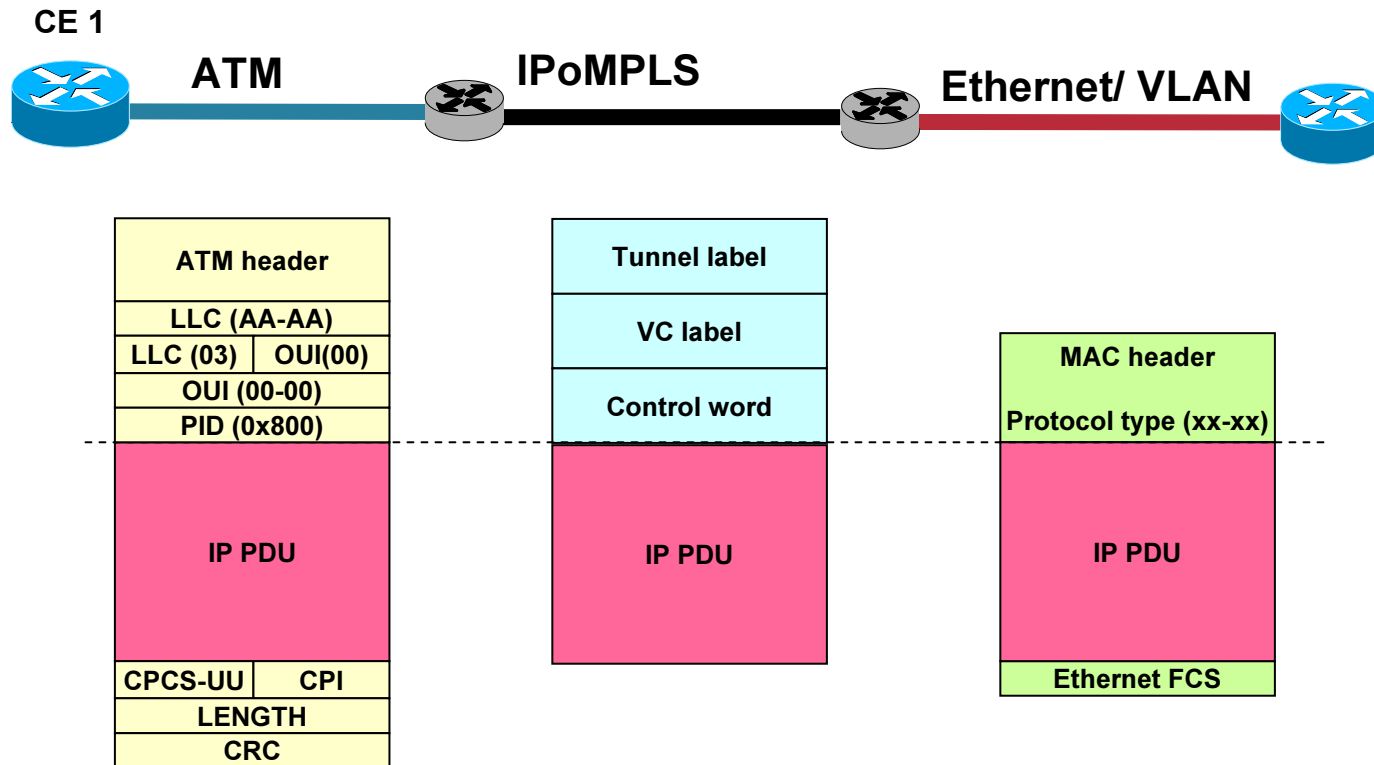


Ethernet
Frame Relay
PPP/HDLC
ATM



Ethernet
Frame Relay
PPP/HDLC
ATM

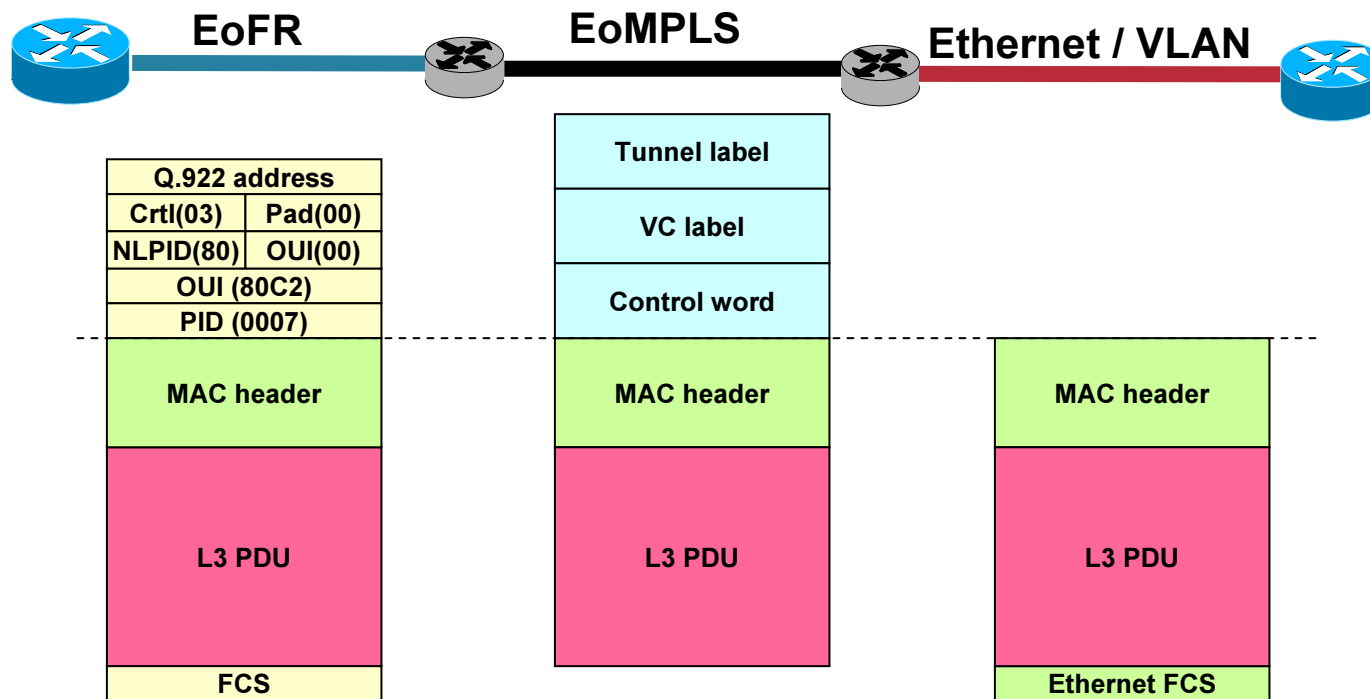
Ethernet to AAL5: IP interworking



Ethernet to FR bridged interworking

CE 1

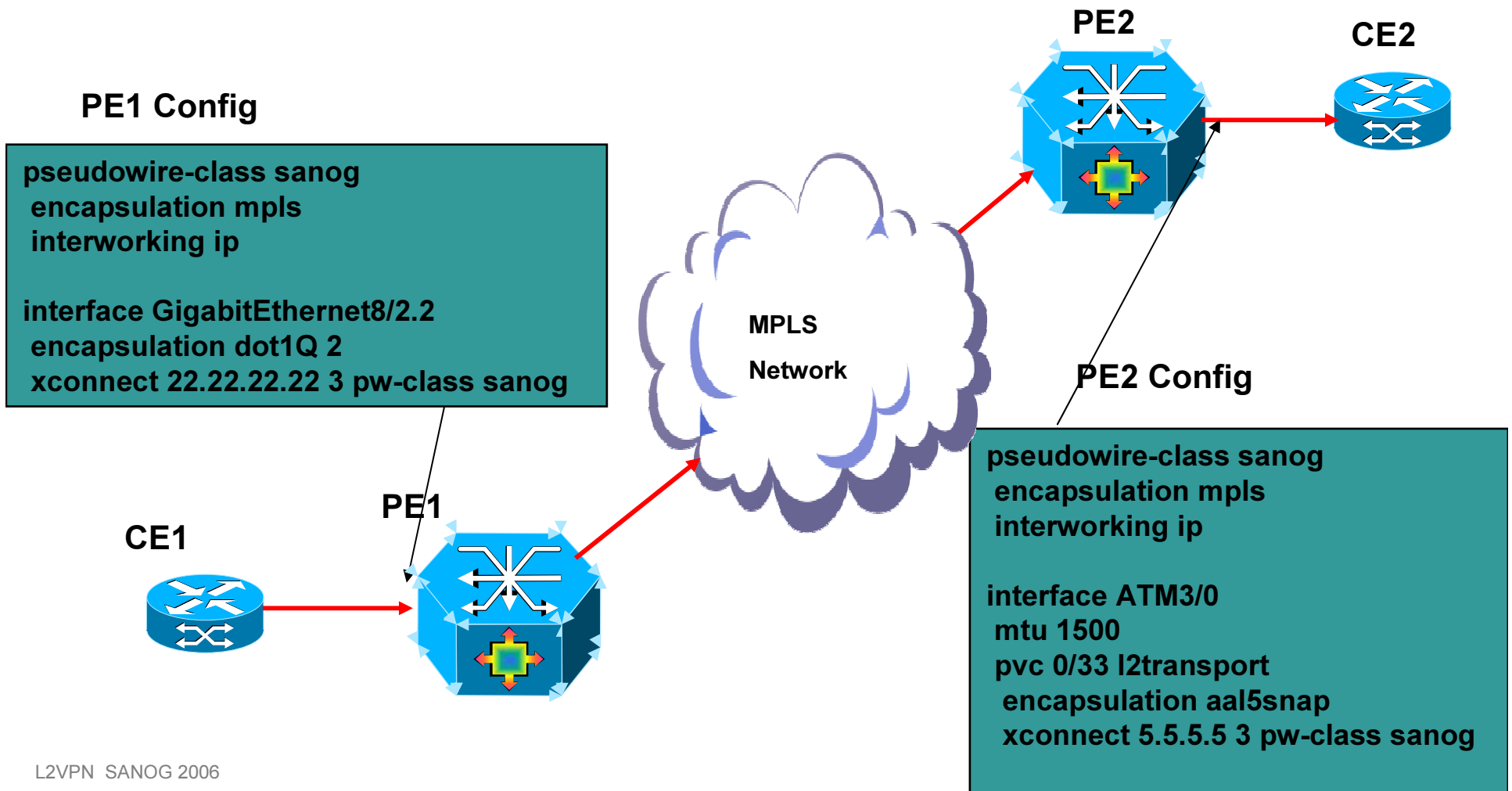
Running RBE/IRB (RFC2427)



DEMO



Point to Point VLAN to ATM Interworking



Virtual Private LAN Service (VPLS)



Virtual Private LAN Service

- **VPLS Overview**
- **VPLS Architectures**

VPLS – Overview

- **Architecture**

It is an end-to-end architecture that allows IP/MPLS networks to provide Layer 2 multipoint Ethernet services while using LDP as signaling protocol

- **Bridge emulation**

Emulates an Ethernet bridge

- **Bridge functions**

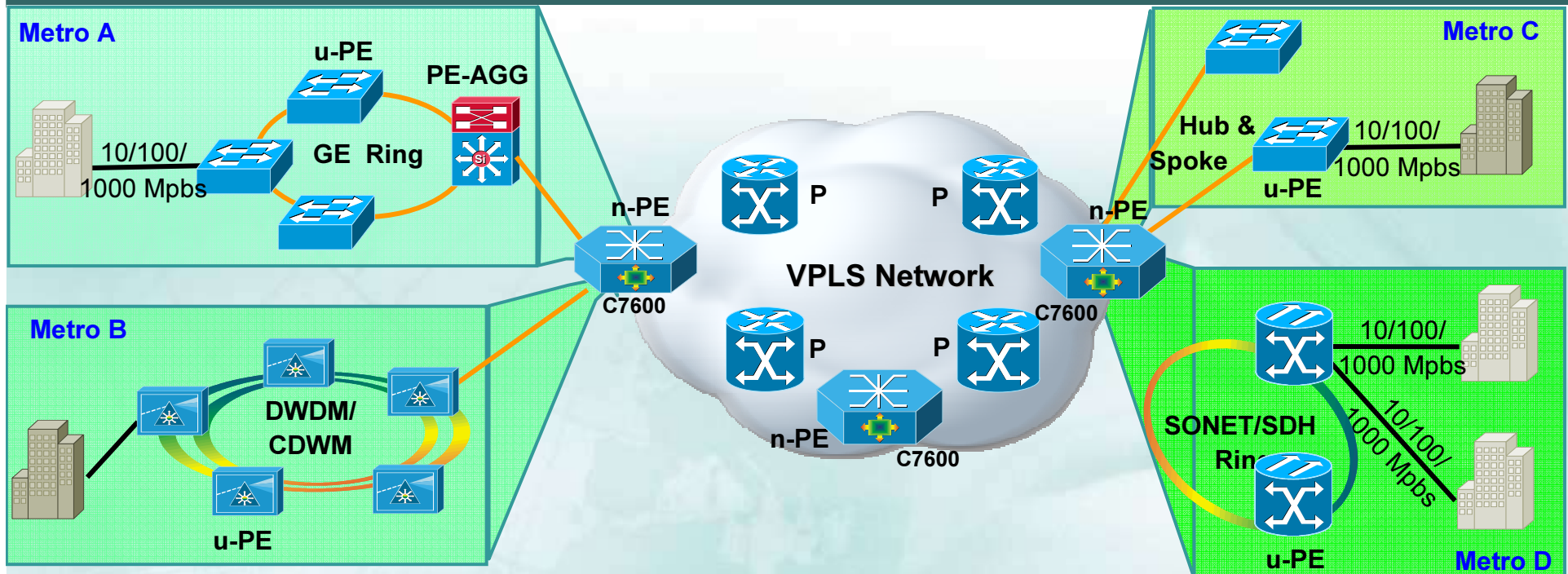
Operation is the same as for an Ethernet bridge, ie forwards using the destination MAC address, learns source addresses and floods broad-/multicast and unknown frames

- **Several drafts in existence**

draft-ietf-l2vpn-vpls-ldp-xx.txt

draft-ietf-l2vpn-vpls-bgp-xx.txt

VPLS Overview for Metro Ethernet



- Delivers Ethernet-based multipoint L2 VPN service
- Enhances L2 VPN scalability (geographic sites & no. of customers)
- Leverages existing SP MPLS Core
- Supports operational speeds of GB to 10 GB
- On track for IETF standardization: Draft Lasserre-Kompella
- Uses familiar Ethernet user network interface

VPLS: Requirements

A Virtual Switch MUST operate like a conventional L2 switch!

Flooding / Forwarding:

- MAC table instances per customer and per customer VLAN (L2-VRF idea) for each PE
- VSI will participate in learning, forwarding process

Address Learning / Aging:

- Self Learn Source MAC to port associations
- Refresh MAC timers with incoming frames
- New additional MAC TLV to LDP

Loop Prevention:

- Create partial or full-mesh of EoMPLS VCs per VPLS
- Use “split horizon” concepts to prevent loops
- Announce EoMPLS VPLS VC tunnels

VPLS Characteristics

- **Auto-discovery of VPN membership**
 - Reduces VPN configuration and errors associated with configuration
- **Signaling of connections between PE devices associated with a VPN**
- **Forwarding of frames**
 - AToM uses Interface based forwarding
 - VPLS uses IEEE 802.1q Ethernet Bridging techniques
- **Loop prevention**
 - MPLS Core will use a full mesh of PWs and “split-horizon” forwarding
 - H-VPLS edge domain may use IEEE 802.1s Spanning Tree, RPR, or SONET Protection

SP Ethernet

VPLS Overview: VSI Functions

- **MAC Address Management**
 1. **Dynamic Learning of MAC Address on Physical Ports and VC**
 2. **Aging of MAC address**
 3. **Withdraw of MAC address**
 4. **Flooding of Multicast, Unicast, Unknown**
- **Data Forwarding**
- **Customer STP BPDU tunneled across SP Cloud.**

SP Ethernet

VPLS Overview: VPLS Learning

- **Unqualified**

- Single port assigned for all customer VLANs

- Single broadcast domain for all customer VLANs

- Single MAC address space (no overlap!)

- **Qualified**

- Each VLAN has its own VPLS instance

- A VLAN has its own broadcast space and MAC address space

- Customer MAC addresses MAY overlap

- One FIB per customer VLAN

- Broadcast domain limited to VLAN scope

MTM – Mac Table Management

SP Ethernet

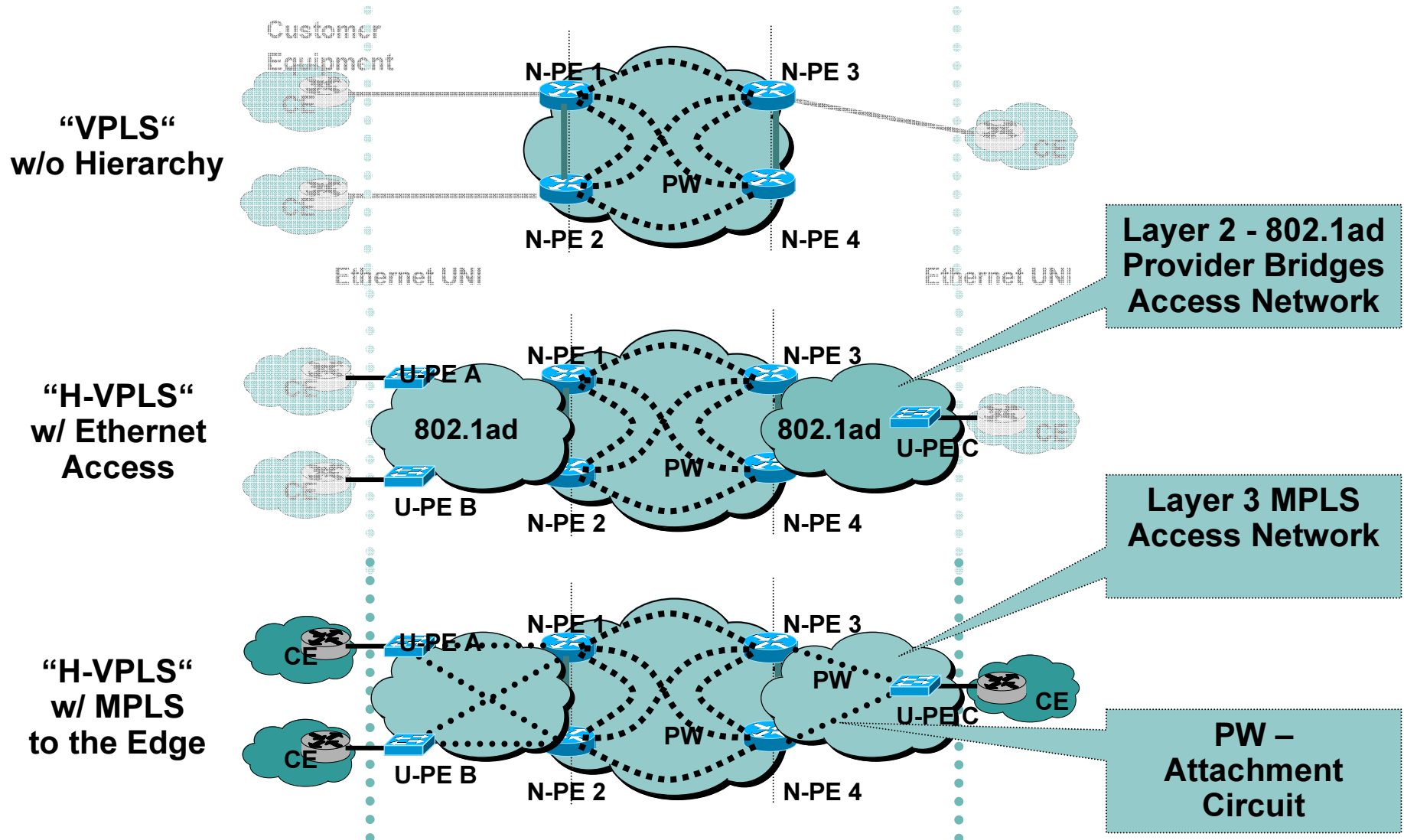
Virtual Private LAN Service

- **VPLS Overview**
- **VPLS Architectures**

VPLS Architectures

- **VPLS defines two Architectures**
 - Non-Hierarchical (Single PE)**
 - Hierarchical (Distributed PE)**
 - 802.1ad (aka QinQ) Access**
 - MPLS Access**
- **Each Architecture has different scaling characteristics**

Way to Build a L2 Core: VPLS—Virtual Private LAN Services



VPLS Architecture: Characteristics - Direct Attachment (Flat)

Overview:

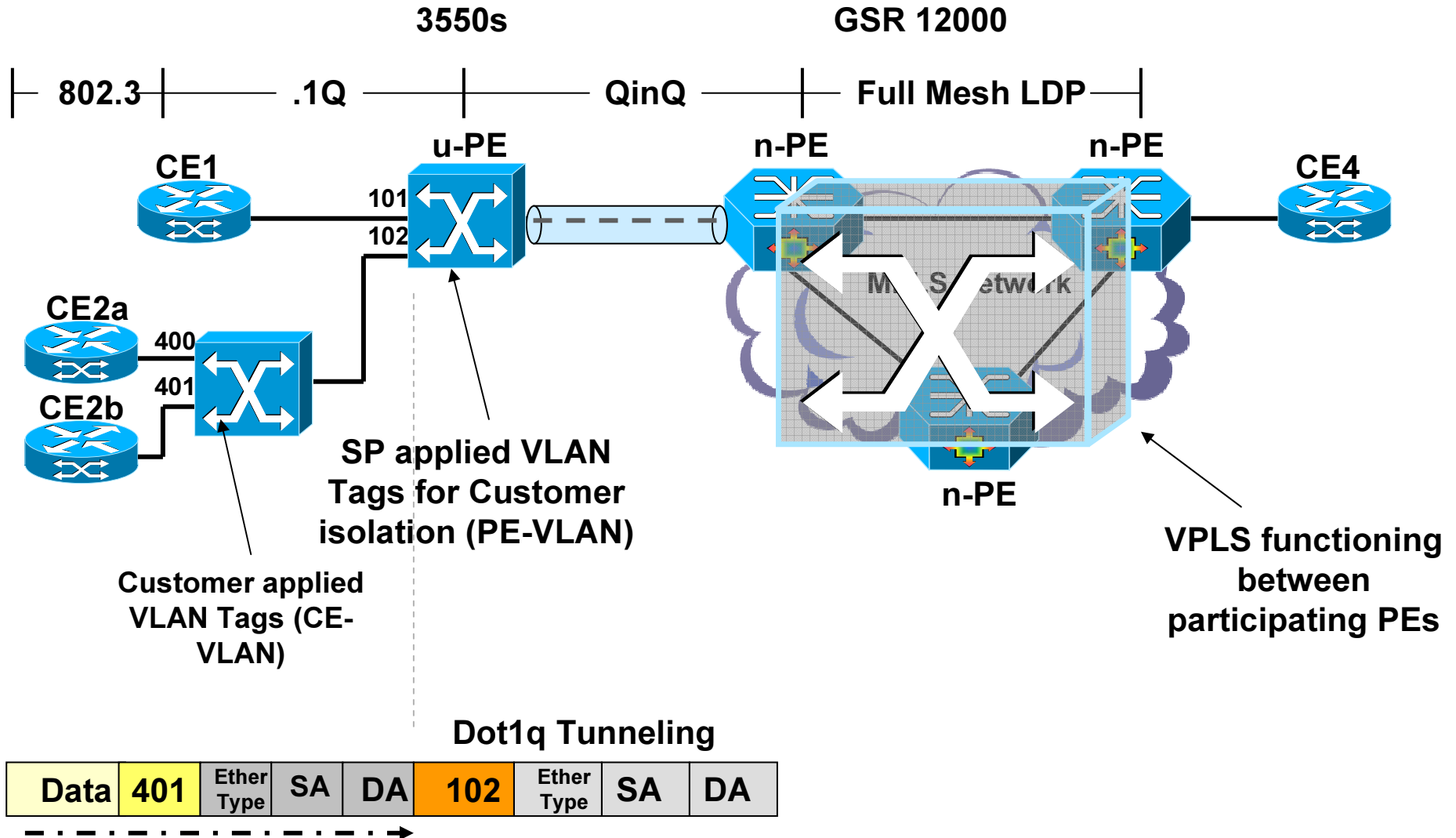
- Okay for small customer implementations
- Simple provisioning
- Full mesh of directed LDP sessions required between participating PEs
- VLAN and Port level support (no QinQ)

Drawbacks:

- No hierarchical scalability
- Scaling issues:
 - PE packet replication
 - Full mesh causes classic - $N*(N-1) / 2$ concerns

SP Ethernet

VPLS Architecture: Architecture – Ethernet Edge H-VPLS



VPLS Architecture: Characteristics – H-VPLS

Benefits:

- **Best for larger scale deployment**
- **Reduction in packet replication and signaling overhead on PEs**
- **Full mesh for core tier (Hub) only**
- **Attachment VCs “virtual switch ports” effected through Layer 2 tunneling mechanisms (MPLS PW, L2TPv3, QinQ)**
- **Expansion affects new nodes only (no re-configuring existing PEs)**

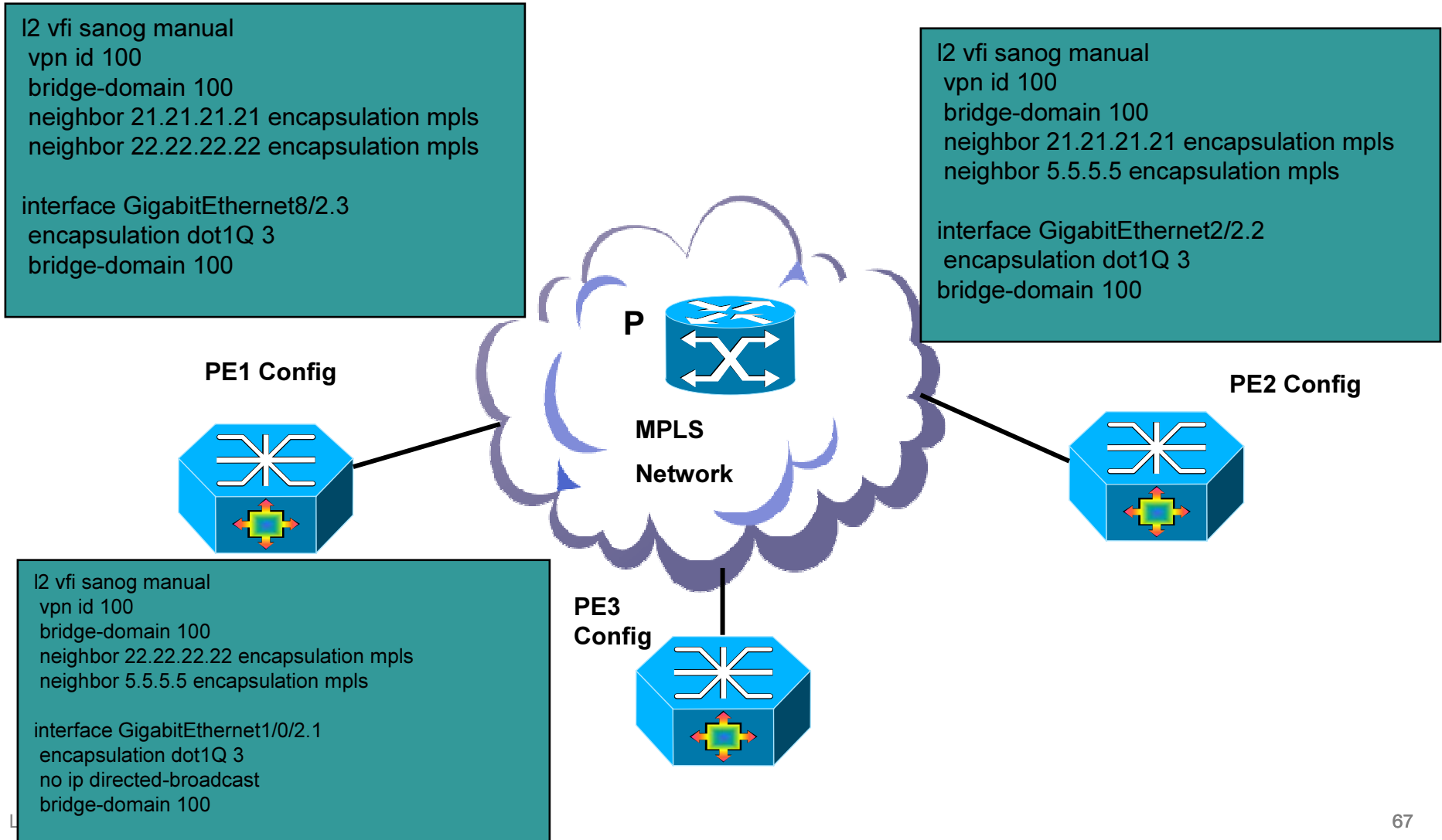
Drawbacks:

- **More complicated provisioning**
- **MPLS Edge H-VPLS requires MPLS to u-PE**
 - Complex operational support**
 - Complex network design**
 - Expensive Hardware support**

DEMO



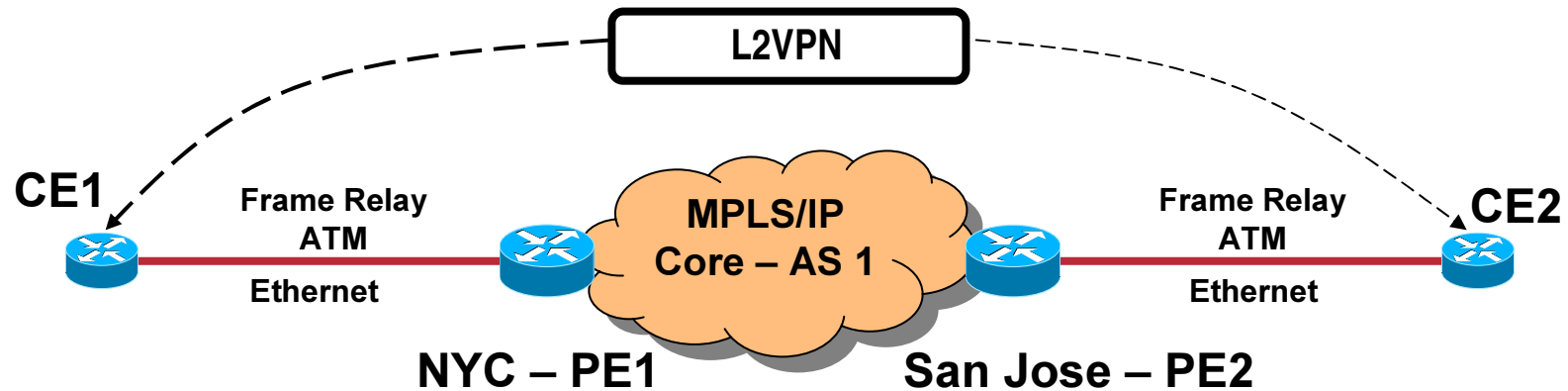
VPLS with Dot1q ACs



Pseudo Wire Stitching



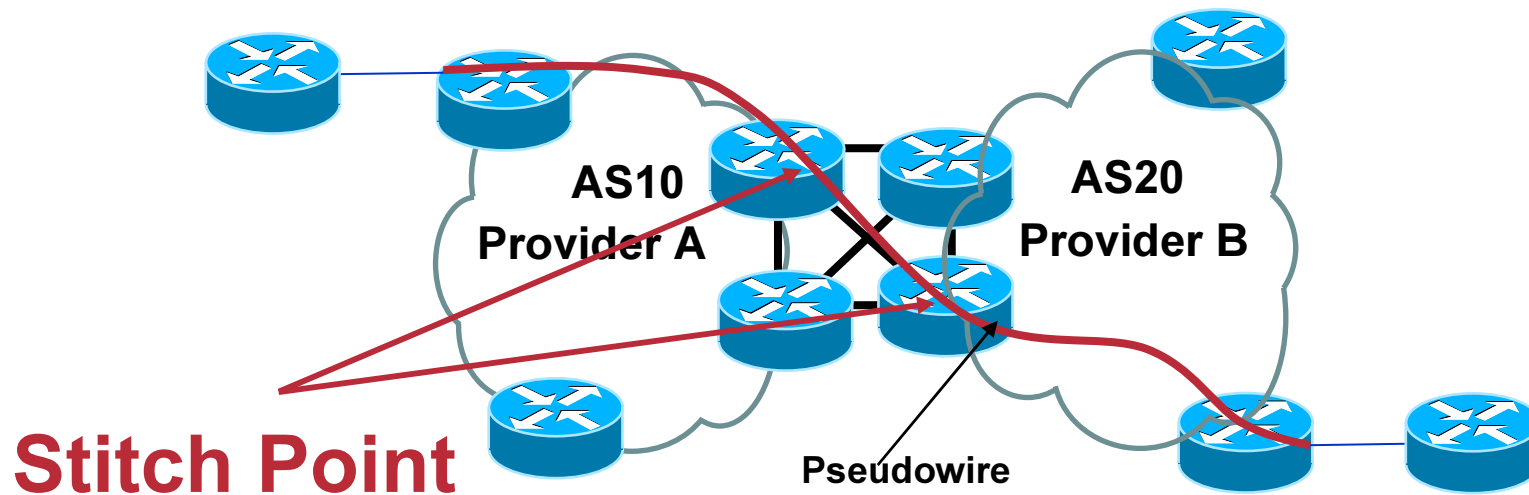
Pseudo Wire Extensibility Limitation



- **What is the Problem?**

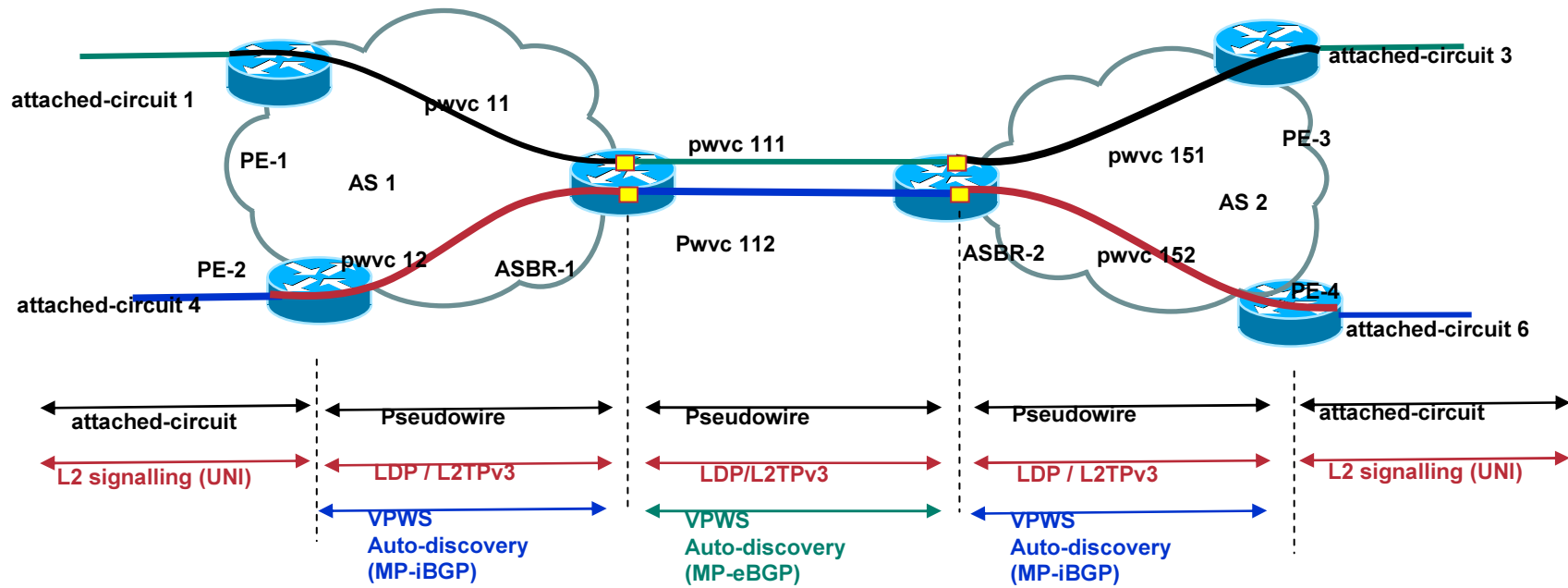
- L2VPNs are created by inter-connecting 2 attachment circuits using a Pseudo Wire (AToM and L2TPv3) to provide an end to end L2 connection but in **1 BGP AS**
- End to end L2VPN deployment not possible across multiple ASes
- Changes in the control and data plane code are required for inter-working them across multiple ASes

Tunnel Stitching at ASBR – Stitch Point



- Tunnel stitching solves this problem by inter-connecting pseudo wires belonging to different autonomous systems and thus providing an end-2-end path
- Tunnel stitch point refers to the ASBR where tunnel stitching is performed
- Achieved through inter-working of data and control planes at the stitch point

Pseudo Wire Stitching Reference Model



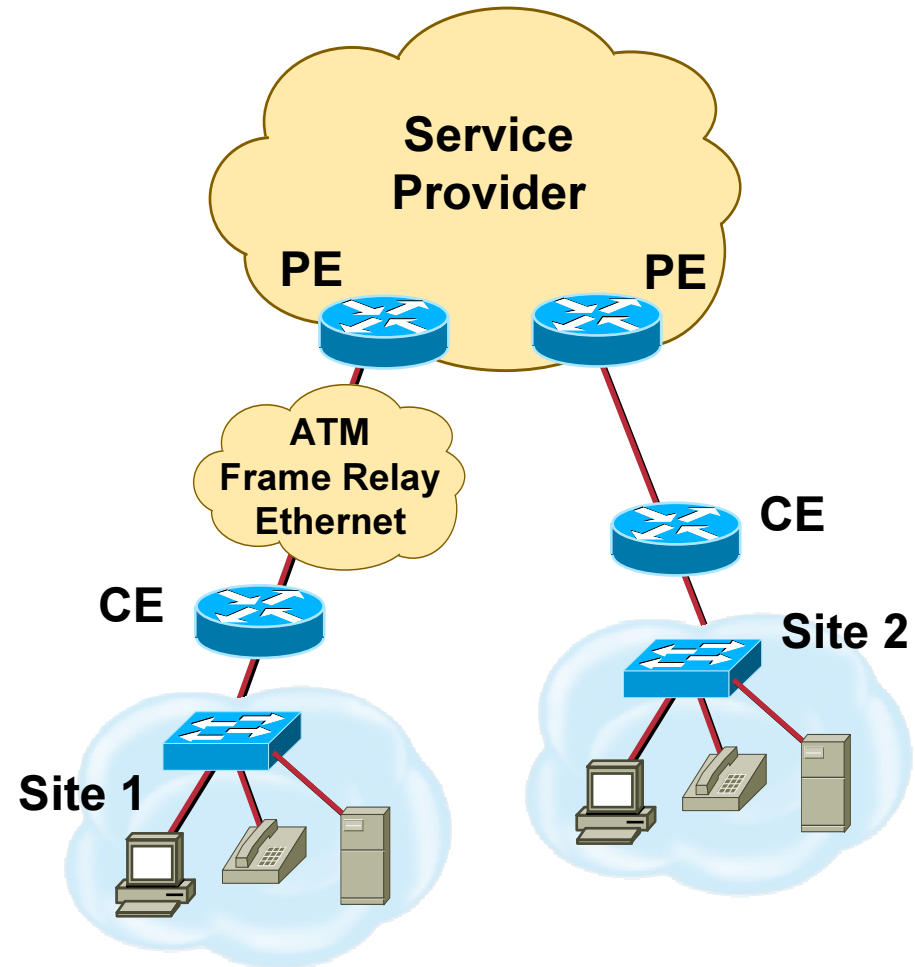
- Each pseudo wire segment can independently employ draft-martini or L2TPv3 signaling and encapsulations
- The ASBRs are responsible for "cross-connecting" the pseudowire control channels and pseudowire data planes
- Easy provisioning and Scalable
- In the end-2-end L2VPN path, you can have 1 or multiple stitch point

Quality of Service

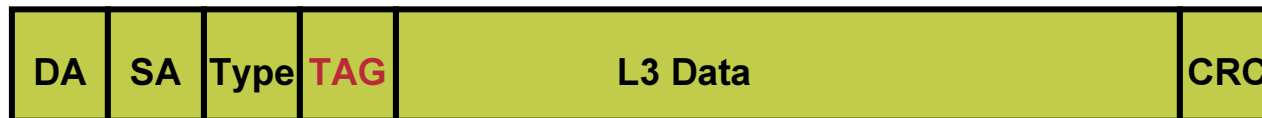
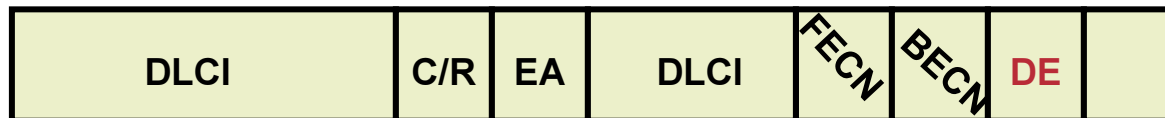
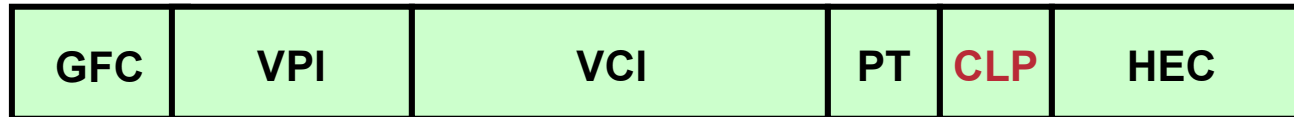


L2VPN Deployment – *The Layer-2 Service Level Agreement (SLA)*

- Point to point guarantees are the basis for the SLA
 - Delay
 - Delay Variation
 - Loss
- In a port trunking environment the FR / ATM / Ethernet access network may enforce SLA
 - Marking, Policing
 - Queuing, Shaping, etc.
- Pay-as-you-grow services can be implemented for many service encaps
- Service Provider Equipment is transparent to the Customer



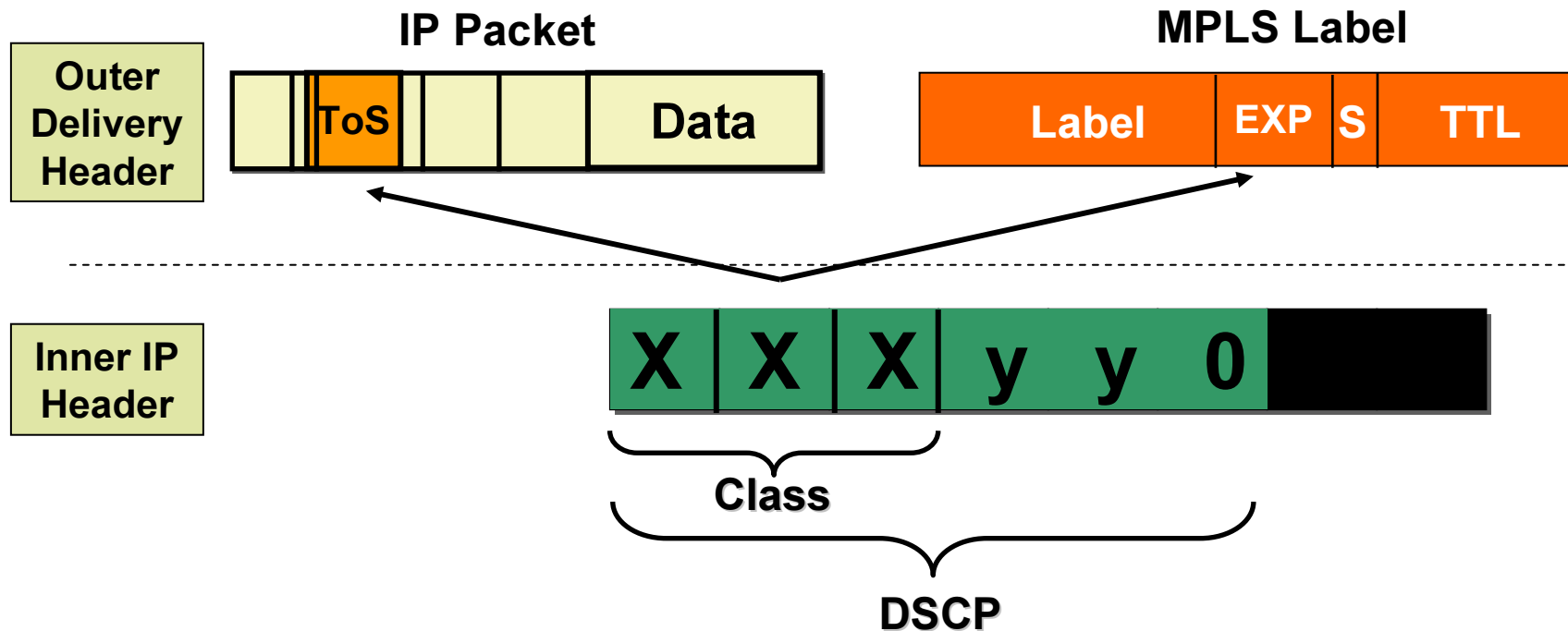
L2VPN Deployment – *How does an SP prioritize traffic?*



Combination of VC IDs & / or Discard indication can be used to classify traffic correctly.

- Provide Service on Port, Layer 2 or Layer 3 Info
- Enforcing sub-rate leased line access
- L2 PDUs provide options for setting frame priority (ex: CLP, DE, 802.1p)
- L2 PVC or Inner L3 Precedence (e.g. ATM VP 10 = CBR VCs)

L2VPN Deployment – *Precedence Equivalence: MPLS / IP*



- IP Precedence is the most often used in determining different traffic priorities (0 – 7) (Customer set, SP optionally enforces)
- Most SPs implement 3 – 5 traffic classes (Best Effort → Mission Critical)
- Enforce policies through shaping, marking, policing

L2VPN QoS Deployment

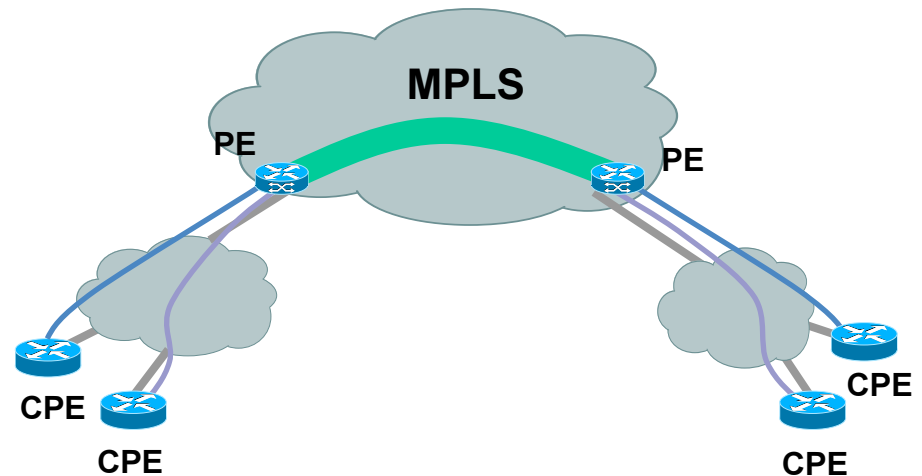
L2 QoS Functions:

- **Shaping** – When transmit rates are higher than expected buffering or queuing is used delay excess traffic, opposite of policing
- **Marking** – The ability to differentiated packets by setting properties within the Layer 2 or Layer 3 header like the IP precedence, or L2 Class of Service or drop priority.
- **Policing** – Used to drop or remark with a lower priority IP Precedence or MPLS EXP bits in traffic that is in excess of contract.
- **Queuing** – Congestion management by giving correct priority to traffic classes one can manage time-sensitive applications without penalizing lower priority traffic. (CBWFQ)

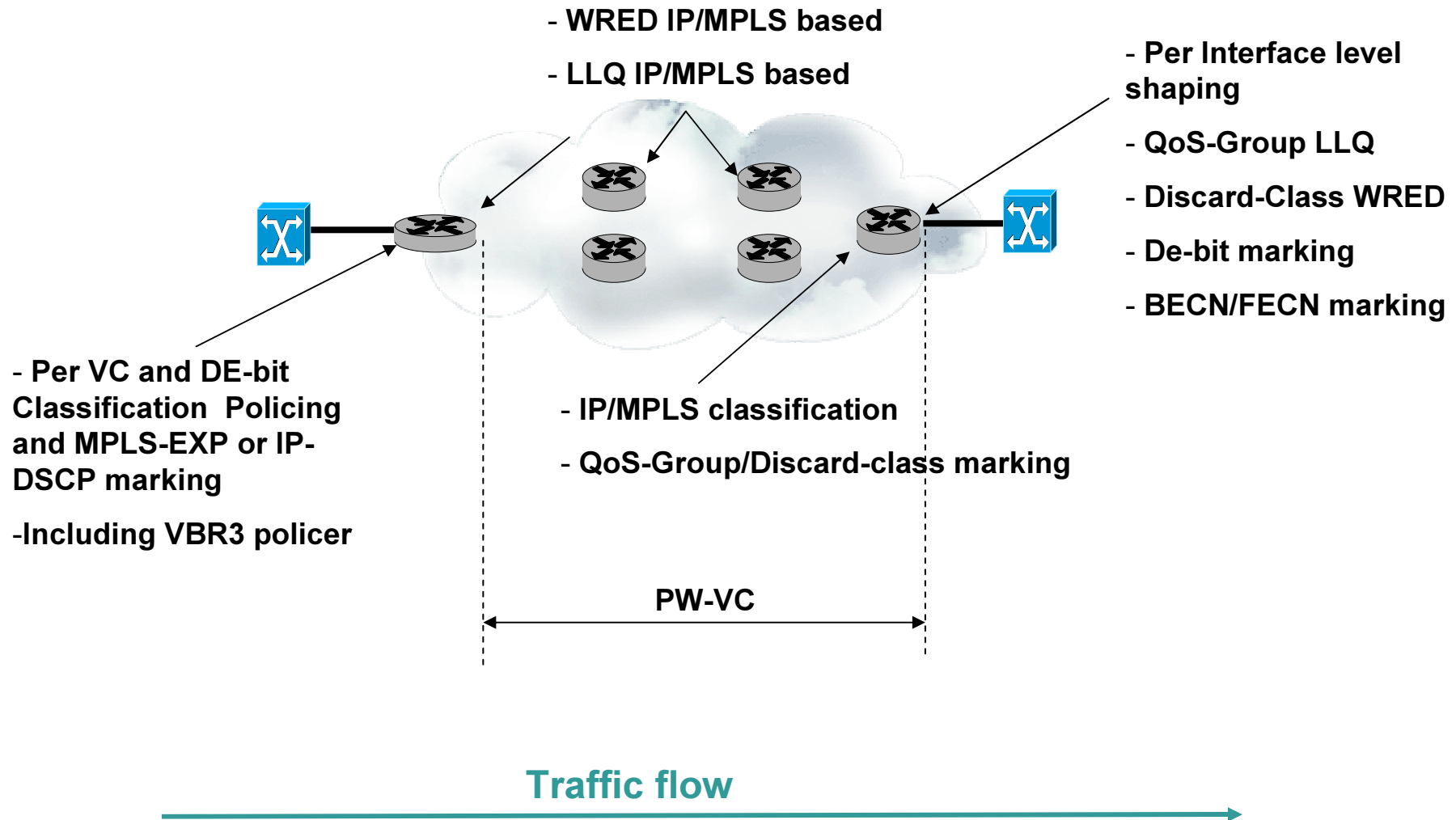
L2 Tunnel Selection

- Tunnel defined as preferred path in pseudo-wire class
- Pseudo-wire class applied to attachment circuit xconnect
- Fallback can be disabled if TE tunnel unreachable

L2 Service Transported over a TE Tunnel
(Point-to-Point Tight SLA)



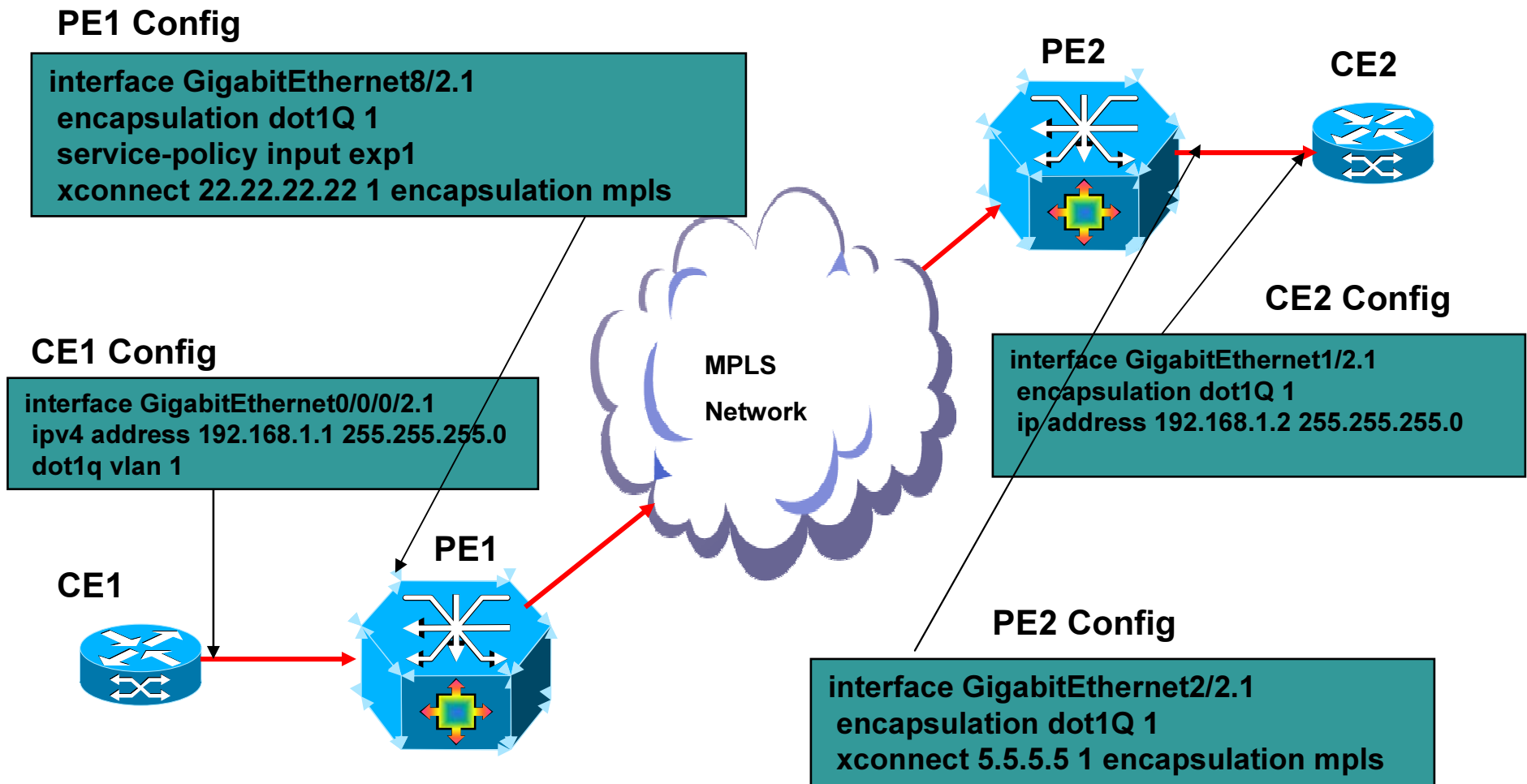
The Complete Picture



DEMO



QoS on Point to Point VLAN over MPLS



Questions & Answers



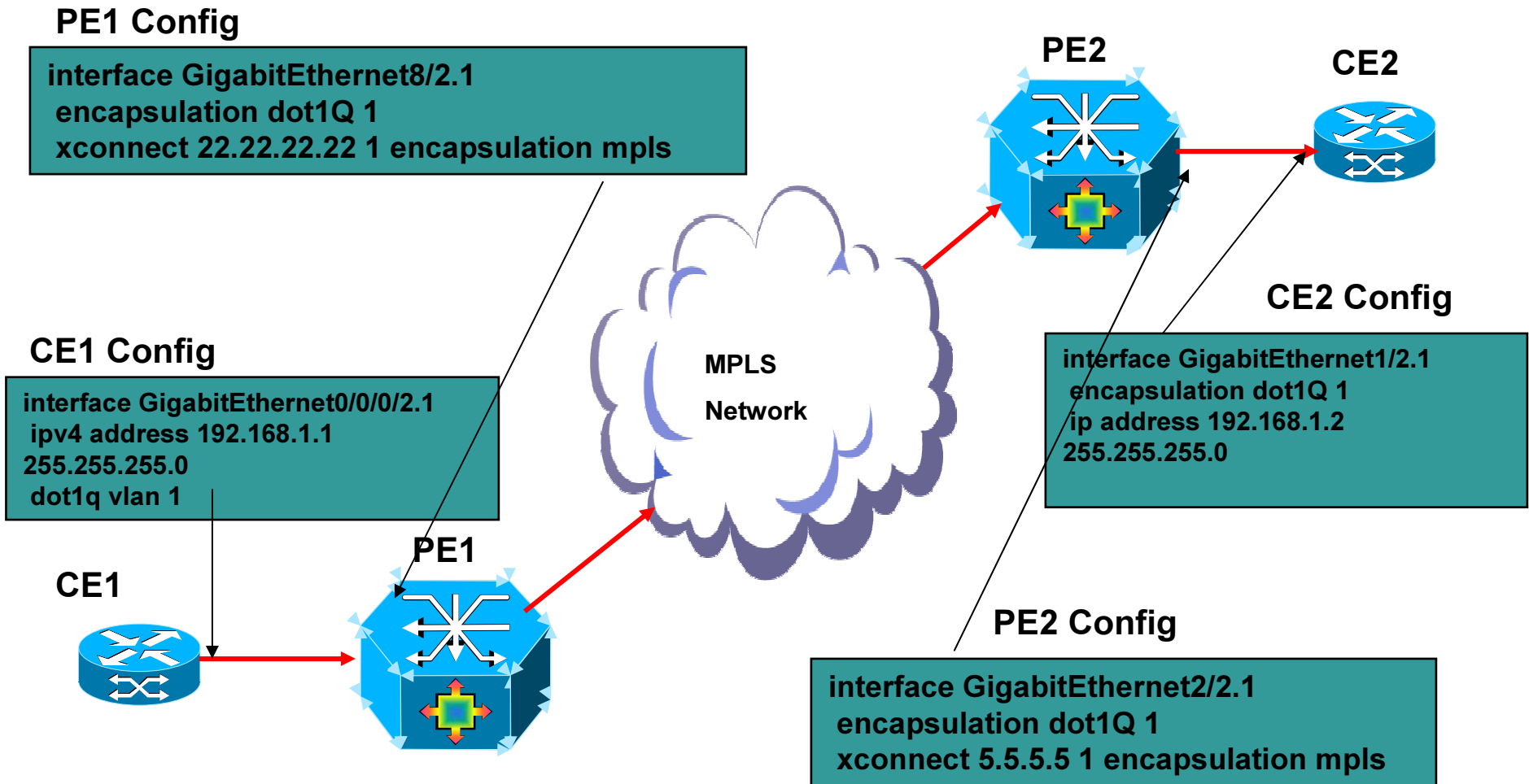
Backup



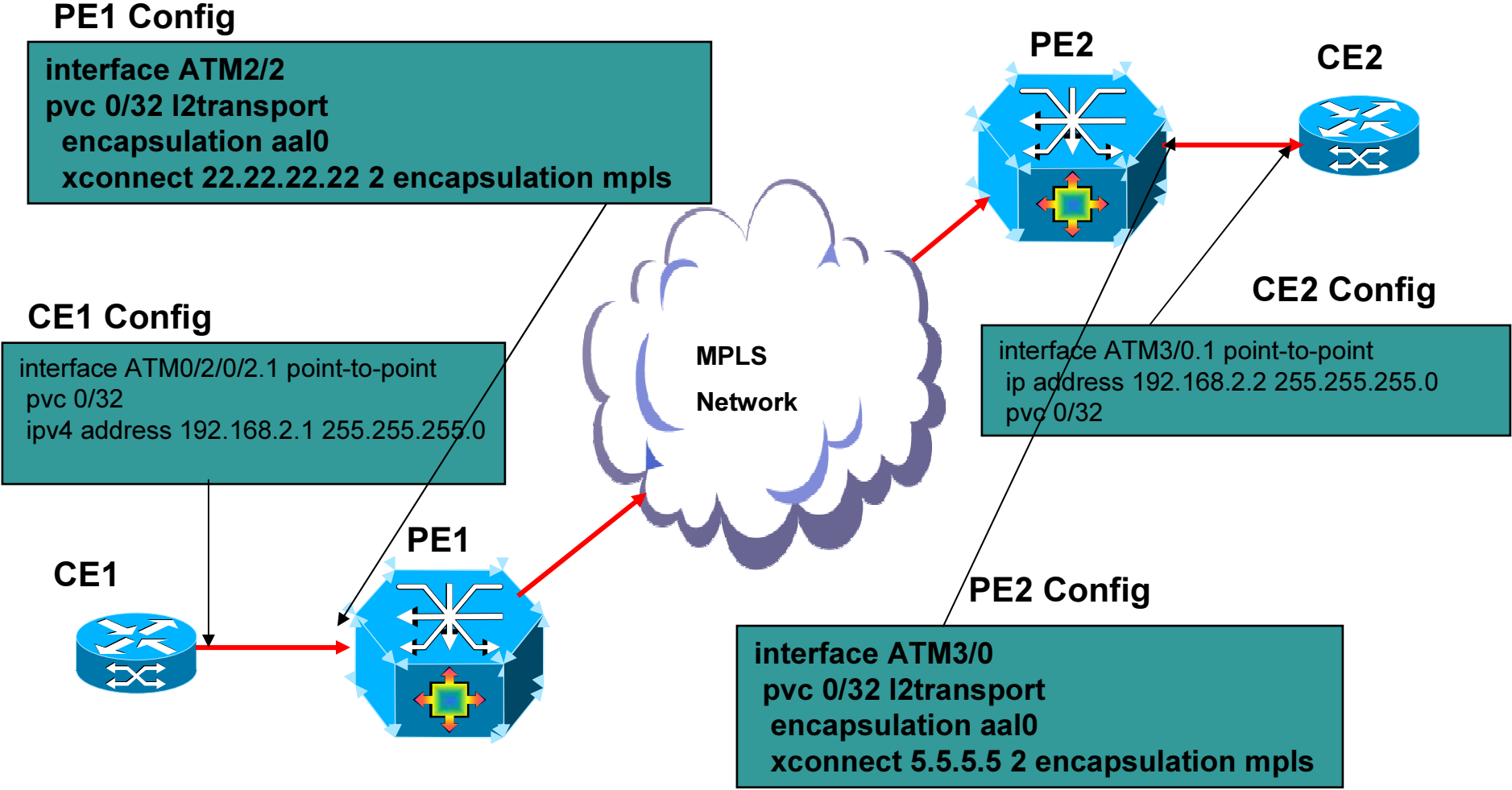
DEMO



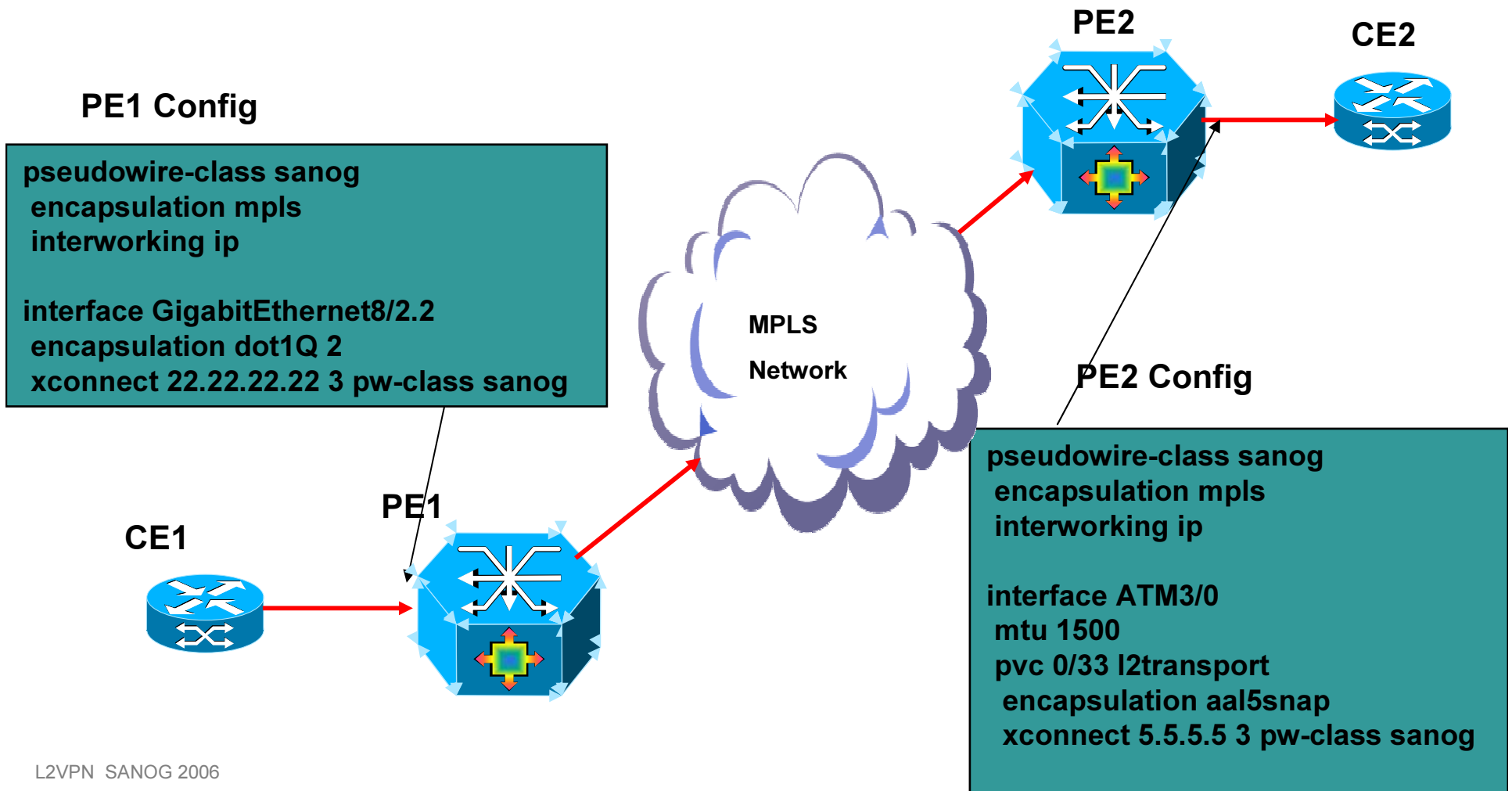
Point to Point VLAN over MPLS



Point to Point Cell Relay over MPLS



Point to Point VLAN to ATM Interworking

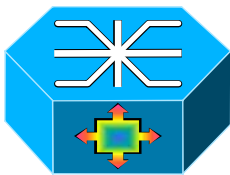


VPLS with Dot1q ACs

```
I2 vfi sanog manual
vpn id 100
bridge-domain 100
neighbor 21.21.21.21 encapsulation mpls
neighbor 22.22.22.22 encapsulation mpls

interface GigabitEthernet8/2.3
encapsulation dot1Q 3
bridge-domain 100
```

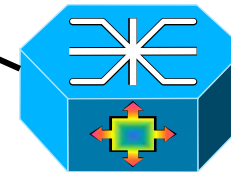
PE1 Config



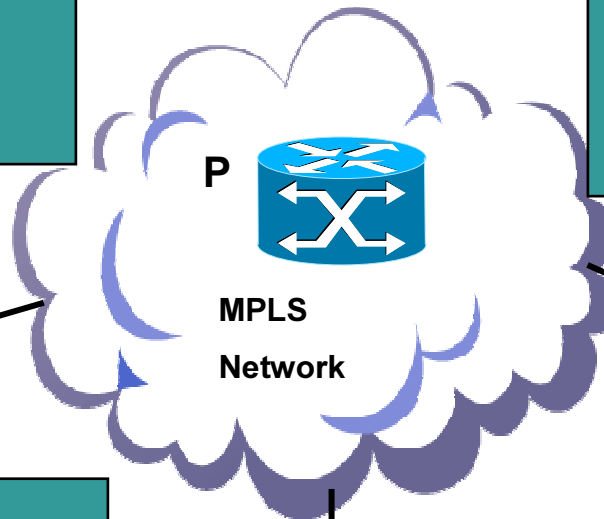
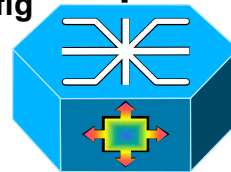
```
I2 vfi sanog manual
vpn id 100
bridge-domain 100
neighbor 21.21.21.21 encapsulation mpls
neighbor 5.5.5.5 encapsulation mpls

interface GigabitEthernet2/2.2
encapsulation dot1Q 3
bridge-domain 100
```

PE2 Config



PE3 Config



P

MPLS
Network

QoS on Point to Point VLAN over MPLS

