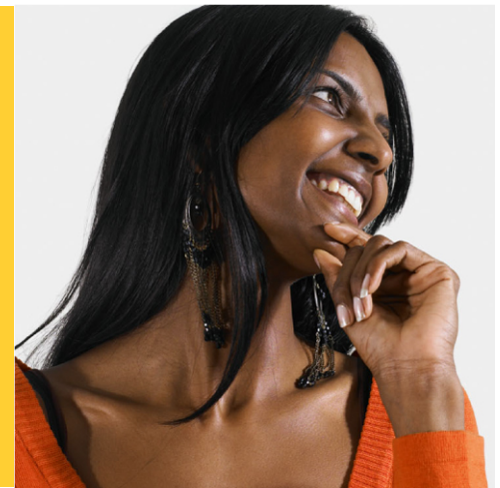


MPLS-based Metro Ethernet Networks

A Tutorial



Paresh Khatri
August 11, 2008

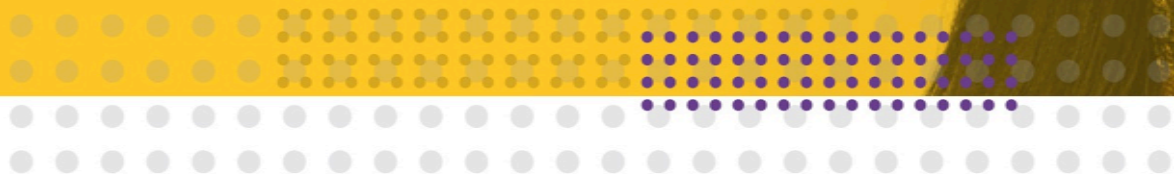
SANOOG

Agenda

1. Introduction
2. Introduction to Metro Ethernet Services
3. Traditional Metro Ethernet networks
4. Delivering Ethernet over MPLS
5. Summary
6. Questions

1

Introduction



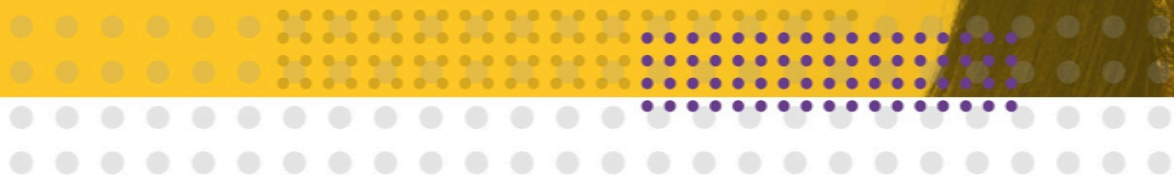
Introduction

Paresh Khatri (paresh.khatri@alcatel-lucent.com.au)

- Consulting Architect - CTO, APAC IP Division of Alcatel-Lucent
- 10 years of design/architecture experience, including 6 years with Service Providers
- Current role(s):
 - Within the architecture team for the Telstra (Australia) TITAN project, a 5-year IP/NGN transformation project
 - Consulting Architect within the APAC IPD CTO organisation
 - Key focus areas:
 - Large-scale IP/MPLS networks
 - L2/L3 VPNs
 - Carrier Ethernet
- Acknowledgements:
 - Some figures and text are provided courtesy of the Metro Ethernet Forum (MEF)

2

Introduction to Metro Ethernet Services



Agenda

2. Introduction to Metro Ethernet Services

2.1 Why Metro Ethernet ?

2.2 The Market for Metro Ethernet

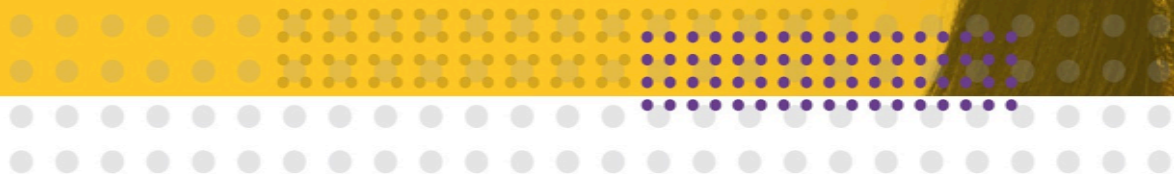
2.3 The work of the Metro Ethernet Forum

2.4 Carrier Ethernet Services defined by the MEF

2.5 The MEF Certification Program

2.1

Why Metro Ethernet ?



What is Metro Ethernet ?

- “... generally defined as the network that bridges or connects geographically separated enterprise LANs while also connecting across the WAN or backbone networks that are generally owned by service providers. The Metro Ethernet Networks provide connectivity services across Metro geography utilising Ethernet as the core protocol and enabling broadband applications”
from “Metro Ethernet Networks - A Technical Overview” from the Metro Ethernet Forum

Introduction to Metro Ethernet Services

SONOG

Traditional alternatives to Metro Ethernet:

- TDM-based services such as leased lines delivered over SDH/SONET networks
 - Sub-rate/E1/T1/E3/T3/STM-1...
 - NG-SDH implementations allow Ethernet transport over SDH/SONET networks
- Frame-relay services
- ATM services

Introduction to Metro Ethernet Services

SONOG

Why Metro Ethernet ?

- Benefits both providers and customers in numerous ways ...
- Packet traffic has now overtaken all other traffic types
 - 98% of all enterprise LAN traffic starts and ends on an Ethernet port
 - Network infrastructure needs to move away from the traditional circuit-switched networks to more efficiently deliver packet traffic
 - The use of circuit-switched technology (e.g. SDH/SONET) within metro networks is not the most efficient way of delivering packet traffic
- Need for rapid provisioning
 - As more and more sites get added to metropolitan networks, the time-to-provision has to be reduced
 - Building services on circuit-switched networks is notoriously slow

Why Metro Ethernet ? (continued)

- Reduced CAPEX/OPEX
 - Customers are demanding a lower cost per bit - traditional delivery methods cannot keep up with this requirement
 - Continued investment in legacy systems for the ultimate delivery of packet traffic is not cost-effective. Ethernet interfaces provide the lowest cost-per-bit access interface for connecting to customers
 - The Total Cost of Ownership (TCO) of service delivery is greatly reduced, ultimately benefiting both the provider and the customer
- Increased and flexible bandwidth options
 - Customers require higher-speed access tails with a variety of bandwidth options
 - Ethernet offers granularity from 10Mbps to 10Gbps
 - Sub-rate services are also possible
 - Work on 40G/100G Ethernet is underway in the IEEE (802.3ba)

Why Metro Ethernet ? (continued)

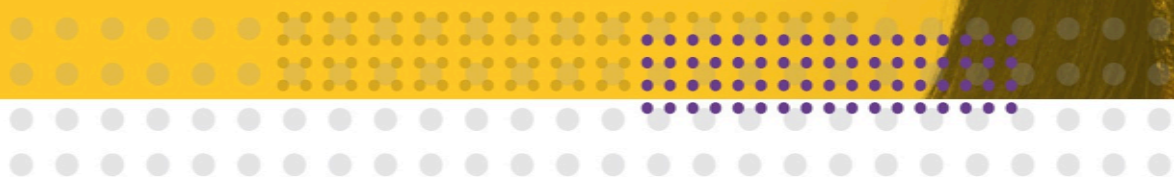
- Well-known interfaces and technology
 - Customers love and understand Ethernet - after all, they have been using it for years in their enterprise LANs
 - Technology is mature, yet continually improving - the last three bandwidth increases in Ethernet interfaces have improved the preceding rates by 1000% !!! . Ethernet retains the potential for continued growth.
- New revenue-generation opportunities
 - The applications and demands for Ethernet services are increasing - providers cannot afford to miss out !
 - Varied applications:
 - Inter-connection of Enterprise LANs
 - Radio Access Network (RAN) Backhaul
 - Emulation of leased lines
 - DSL Backhaul (DSL Forum TR-101)

Why Metro Ethernet ? (continued)

- Ethernet is the interface of choice
 - Ethernet is now becoming the interface of choice not just for customer access interfaces but for also provider trunk interfaces
 - CPE devices with Ethernet interfaces are readily available and at a very low cost point for customers
- Flexibility
 - The nature of Ethernet networks means that the provider now has an unprecedented amount of flexibility in constructing service offerings

2.2

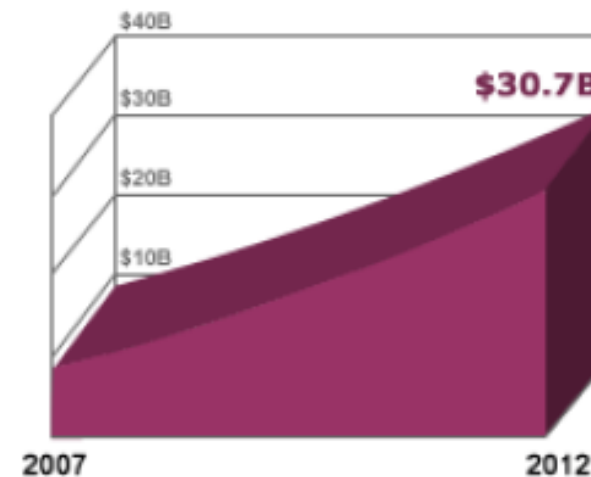
The Market for Metro Ethernet



The Market for Metro Ethernet Services

- Worldwide revenue for Business Ethernet services is expected to reach \$31 billion by 2012
- Robust enterprise customer demand is projected for the next five years, with double-digit annual growth across all geographic regional markets: Asia/Pacific, EMEA, U.S., and ROW.
- Service providers throughout the world are committed to Ethernet as the future ubiquitous standard for network service connectivity.
- Ethernet equipment vendors are actively enabling this important transition.
- Customer accessibility to fibre facilities is the top challenge to worldwide growth of Ethernet services.

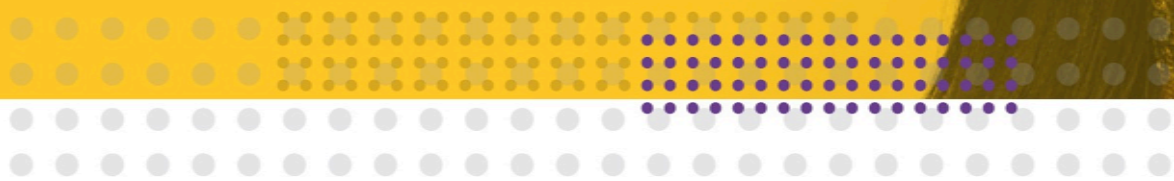
**Worldwide Ethernet Services
Business Applications**
Revenue in \$Billions



Copyright Vertical Systems Group - ENS
www.verticalsystems.com

2.3

The work of the Metro Ethernet Forum (MEF)



Metro Ethernet Forum



MEF Mission & Industry Leadership:

Accelerate the worldwide adoption of Carrier Ethernet networks and services



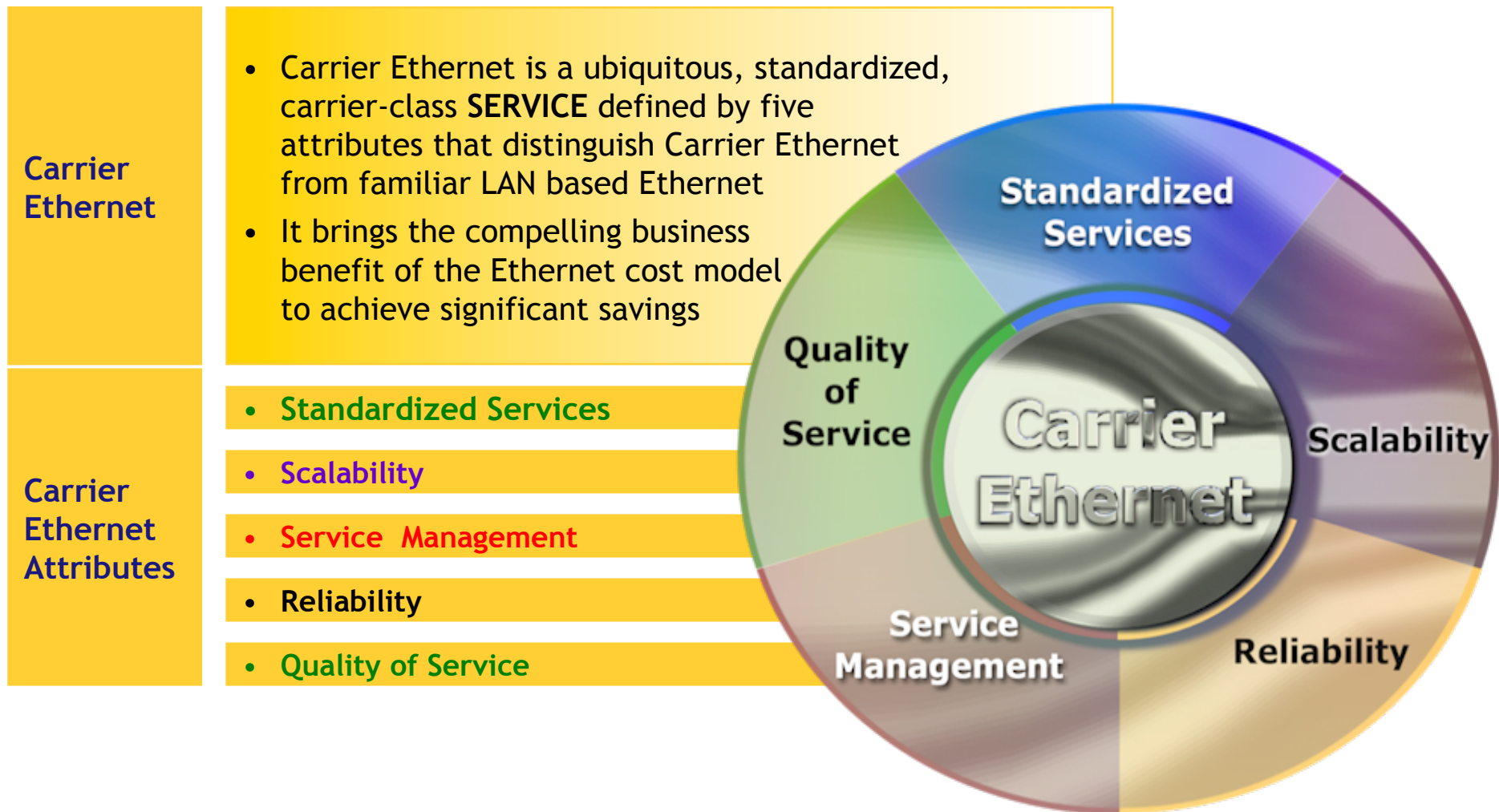
Specifications and Liaison

- MEF 2 Requirements and Framework for Ethernet Service Protection
- MEF 3 Circuit Emulation Service Definitions, Framework and Requirements in Metro Ethernet Networks
- MEF 4 Metro Ethernet Network Architecture Framework Part 1: Generic Framework
- MEF 5 Metro Ethernet Services Definitions Phase I
- MEF 7 EMS-IMS Information Model
- MEF 8 Implementation Agreement for the Emulation of PDH Circuits over Metro Ethernet Networks
- MEF 9 Abstract Test Suite for Ethernet Services at the UNI
- MEF 10 Ethernet Services Attributes Phase I
- MEF 11 User Network Interface (UNI) Requirements and Framework
- MEF 12 Metro Ethernet Network Architecture Framework Part 2: Ethernet Services Layer
- MEF 13 User Network Interface (UNI) Type 1 Implementation Agreement
- MEF 14 Abstract Test Suite for Traffic Management Phase 1
- MEF 15 Requirements for Management of Metro Ethernet Phase 1 Network Elements
- MEF 16 Ethernet Local Management Interface

Marketing Carrier Ethernet

Certification Program

The 5 Attributes of Carrier Ethernet



Attribute 1: Standardized Services

- Ubiquitous services provided locally & globally via standardised equipment.
- E-Line, E-LAN, E-Tree: provide transparent, private line, virtual private line and multi-point to multi-point LAN services.
- Requires no changes to customer LAN equipment or networks and accommodates existing network connectivity such as, time-sensitive, TDM traffic and signaling.
- Ideally suited to converged voice, video & data networks
- Wide choice and granularity of bandwidth and quality of service options

Attribute 2: Scalability

- The ability for millions to use a network service that is ideal for the widest variety of business, information, communications and entertainment applications with voice, video and data
- Spans Access & Metro to National & Global Services over a wide variety of physical infrastructures implemented by a wide range of Service Providers
- Scalability of bandwidth from 1Mbps to 10Gbps and beyond, in granular increments

Attribute 3: Reliability

- The ability for the network to detect & recover from incidents without impacting customers.
- Meeting the most demanding quality and availability requirements.
- Rapid recovery time when problems do occur; as low as 50ms.

Attribute 4: Quality of Service

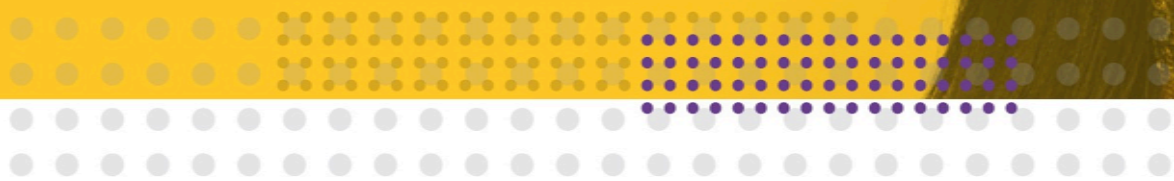
- Wide choice and granularity of bandwidth and quality of service options.
- Service Level Agreements (SLAs) that deliver end-to-end performance matching the requirements for voice, video and data over converged business and residential networks.
- Provisioning via SLAs that provide end-to-end performance based on committed information rate (CIR), frame loss, delay and delay variation characteristics.

Attribute 5: **Service Management**

- The ability to monitor, diagnose and centrally manage the network, using standards-based vendor independent implementations.
- Carrier-Class OAM.
- Rapid service provisioning.

2.4

Carrier Ethernet Services defined by the MEF



Introduction to Metro Ethernet Services

SONOG

What do we mean by Metro Ethernet services ?

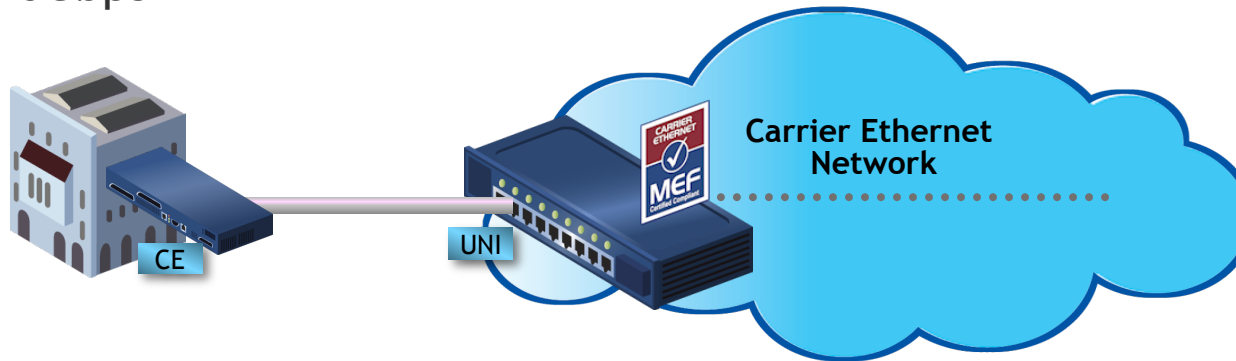
- Use of Ethernet access tails
- Provision of Ethernet-based services across the MAN
 - Point-to-point
 - Point-to-multipoint
 - Multipoint-to-multipoint
- However, the underlying infrastructure used to deliver Ethernet services does NOT have to be Ethernet !!!
- Referred to as Carrier Ethernet services by the Metro Ethernet Forum
 - The terms “Carrier Ethernet” and “Metro Ethernet” are used interchangeably in this presentation, but in the strict sense of the term, “Carrier Ethernet” refers to the carrier-grade evolution of “Metro Ethernet”

MEF Carrier Ethernet Terminology

SONOG

The User Network Interface (UNI)

- The UNI is the physical interface or port that is the demarcation between the customer and the service provider/Cable Operator/Carrier/MSO
- The UNI is always provided by the Service Provider
- The UNI in a Carrier Ethernet Network is a standard physical Ethernet Interface at operating speeds 10Mbps, 100Mbps, 1Gbps or 10Gbps



CE: Customer Equipment, UNI: User Network Interface.

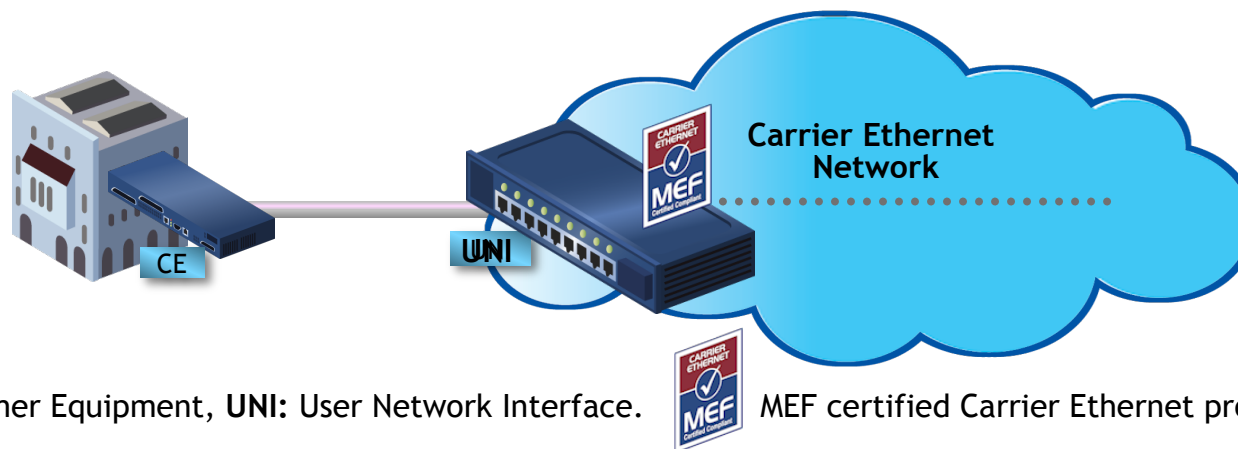


MEF certified Carrier Ethernet products

MEF Carrier Ethernet Terminology

The User Network Interface (UNI):

- MEF has defined two types of UNIs:
 - MEF UNI Type I
 - A UNI compliant with MEF 13
 - Manually Configurable
 - MEF UNI Type II
 - Automatically Configurable via E-LMI
 - Manageable via OAM

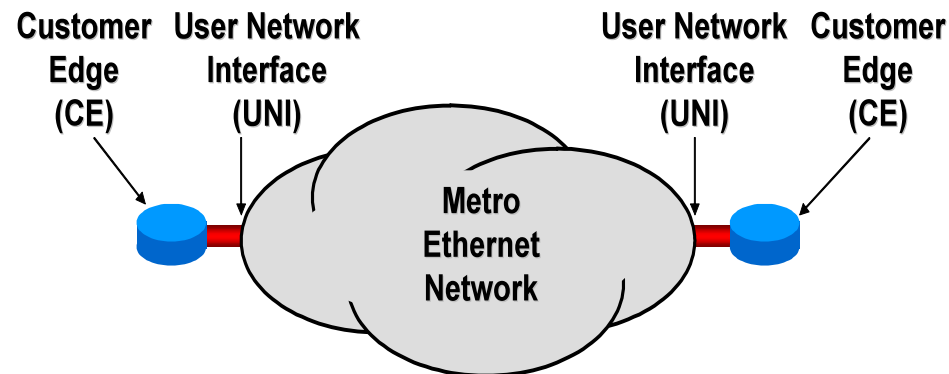


CE: Customer Equipment, UNI: User Network Interface.

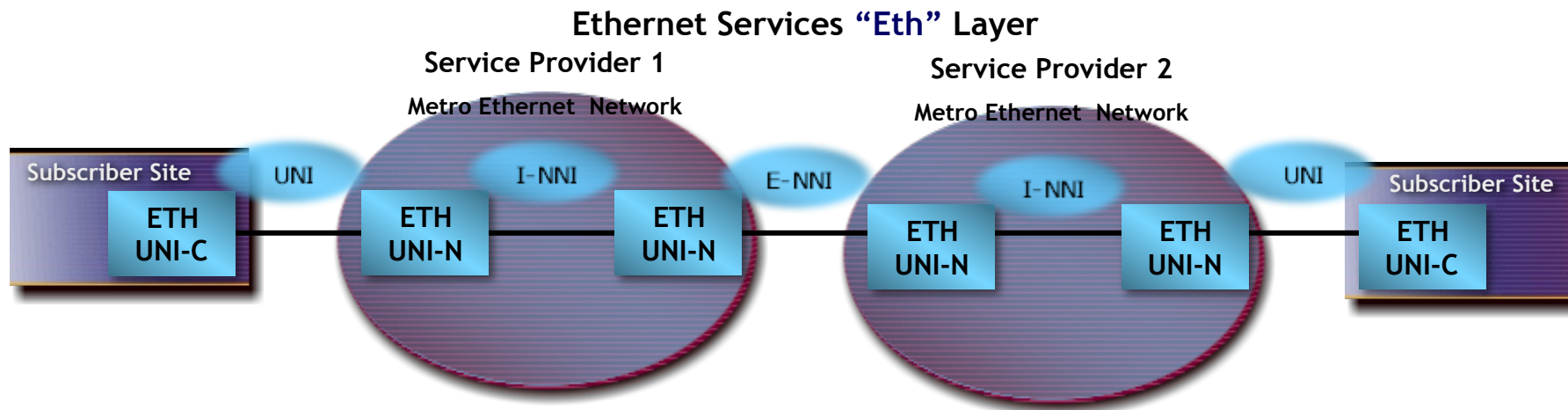
MEF certified Carrier Ethernet products

MEF Carrier Ethernet Terminology

- Customer Equipment (CE) attaches to the Metro Ethernet Network (MEN) at the UNI
 - Using standard Ethernet frames.
- CE can be
 - Router or bridge/switch - IEEE 802.1 bridge



MEF Ethernet Services Model



UNI: User Network Interface, UNI-C: UNI-customer side, UNI-N network side
NNI: Network to Network Interface, E-NNI: External NNI; I-NNI Internal NNI

MEF Carrier Ethernet Terminology

SONOG

Ethernet Virtual Connection (EVC)

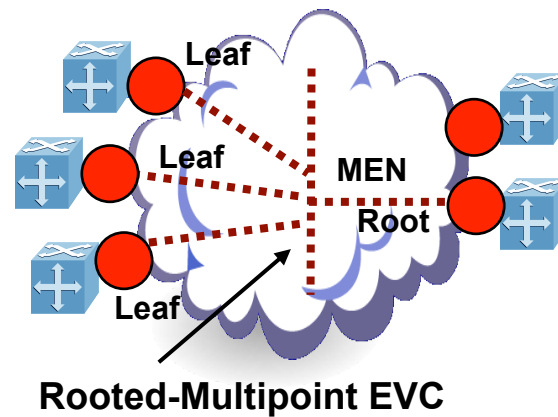
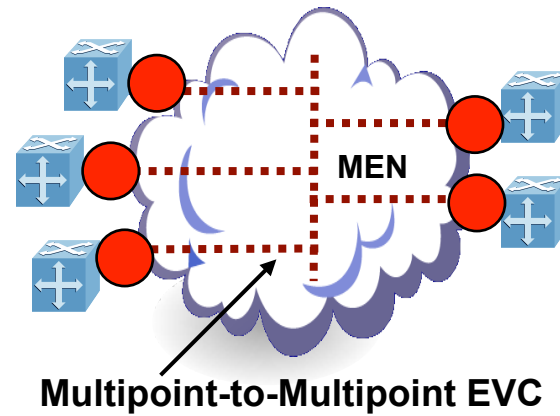
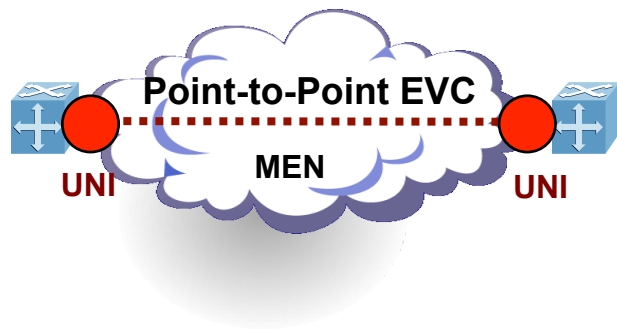
- An Ethernet Service Instantiation
 - Most commonly (but not necessarily) identified via a VLAN-ID
 - Like Frame Relay and ATM PVCs or SVCs
- Connects two or more subscriber sites (UNI's)
 - Can multiplex multiple EVCs on the same UNI
- An association of two or more UNIs
- Prevents data transfer between sites that are not part of the same EVC

MEF Carrier Ethernet Terminology



Ethernet Virtual Connection (EVC)

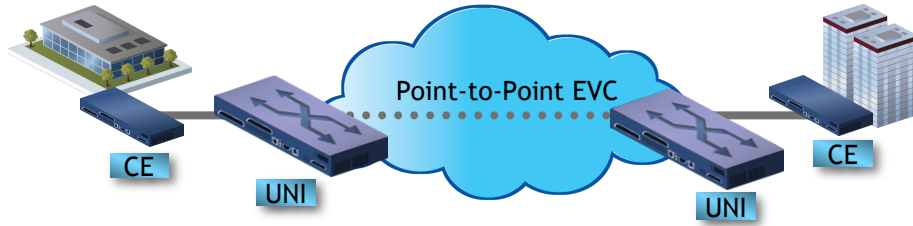
- Three types of EVC:



Basic Carrier Ethernet Services



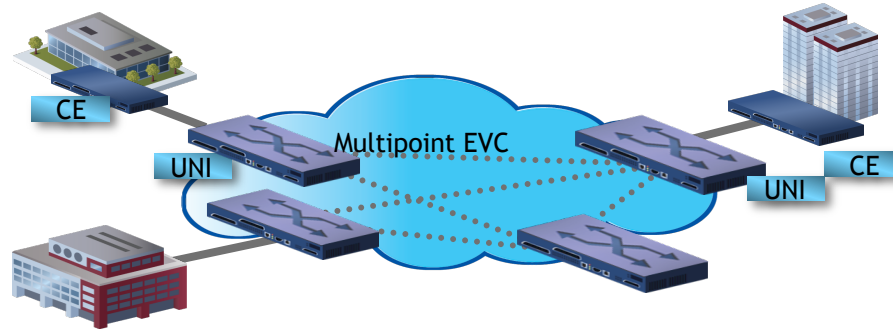
E-LINE



Point to Point
Service Type used to create

- Ethernet Private Lines
- Virtual Private Lines
- Ethernet Internet Access

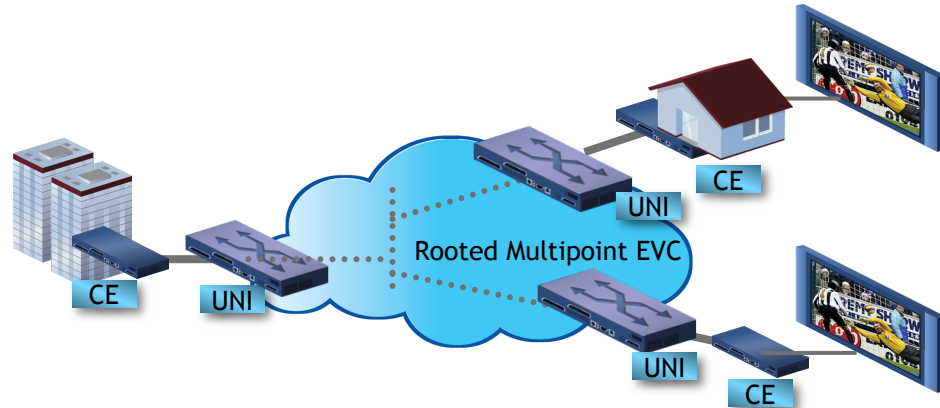
E-LAN



Multi-Point to Multi-Point
Service Type used to create

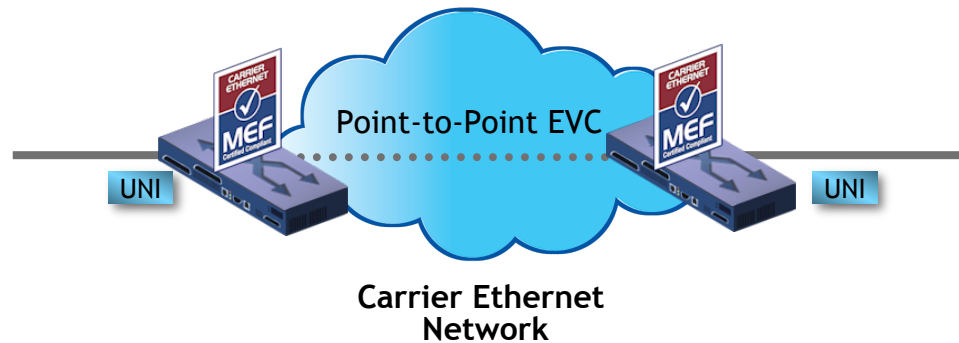
- Multipoint Layer 2 VPNs
- Transparent LAN Service
- Foundation for Multicast networks e.g. TLS, IPTV

E-TREE



Point to Multi-Point
• Efficient use of Service Provider ports

In a Carrier Ethernet network, data is transported across Point-to-Point and Multipoint-to-Multipoint EVCs according to the attributes and definitions of the E-Line and E-LAN services

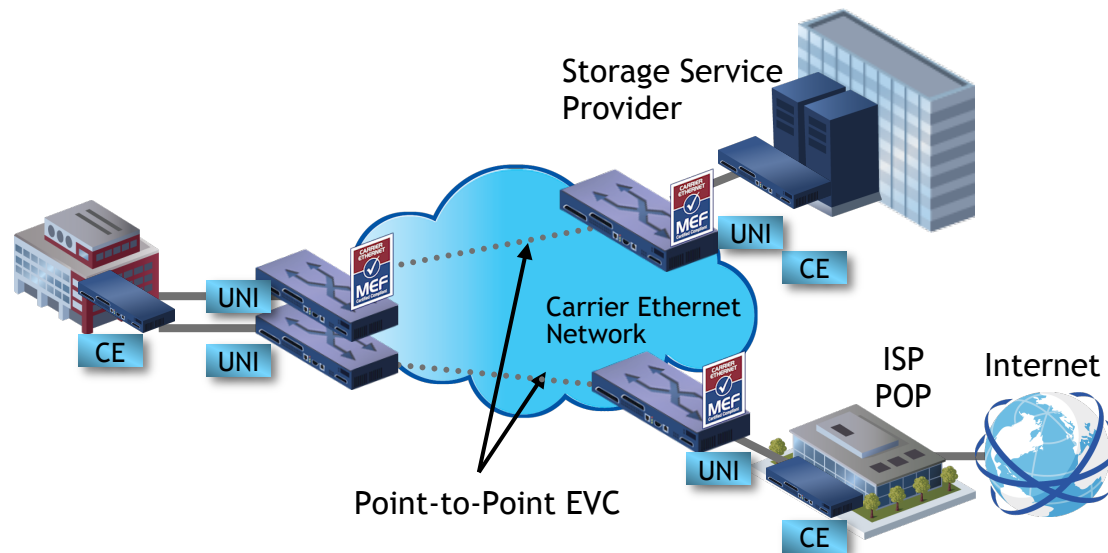


Services Using E-Line Service Type



Ethernet Private Line (EPL)

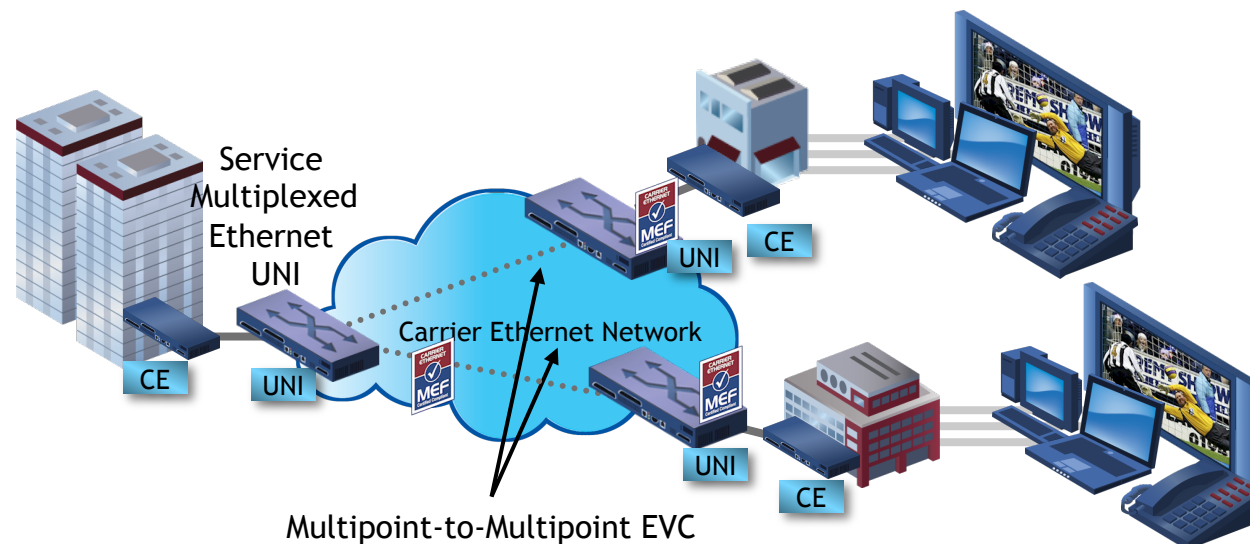
- Replaces a TDM Private line
- Dedicated UNIs for Point-to-Point connections
- Single Ethernet Virtual Connection (EVC) per UNI



Services Using E-Line Service Type

Ethernet Virtual Private Line (EVPL)

- Replaces Frame Relay or ATM services
- Supports Service Multiplexed UNI (i.e. multiple EVCs per UNI)
- Allows single physical connection (UNI) to customer premise equipment for multiple virtual connections
- This is a UNI that must be configurable to support Multiple EVCs per UNI

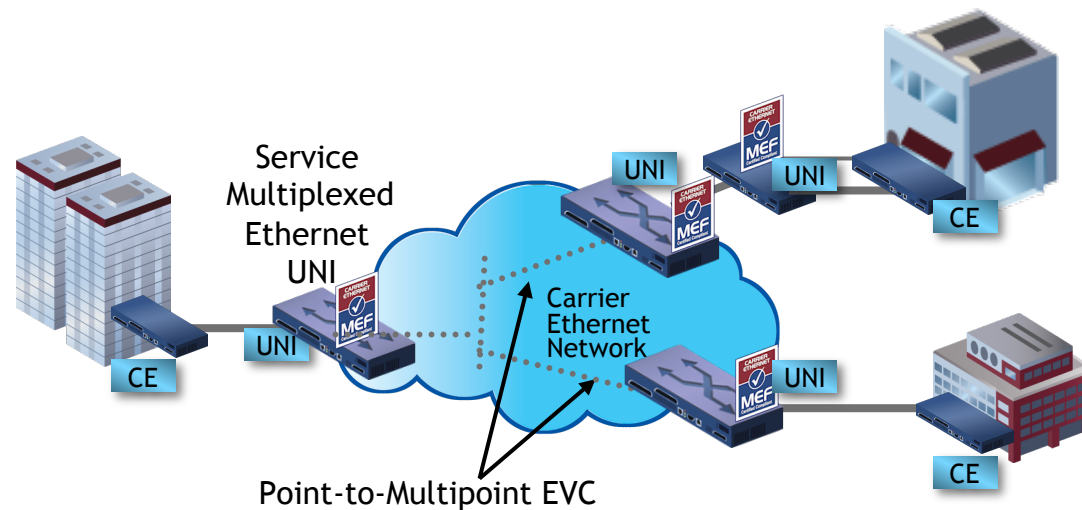


Services Using E-LAN Service Type



Ethernet Private LAN and Ethernet Virtual Private LAN Services

- Supports dedicated or service-multiplexed UNIs
- Supports transparent LAN services and multipoint VPNs

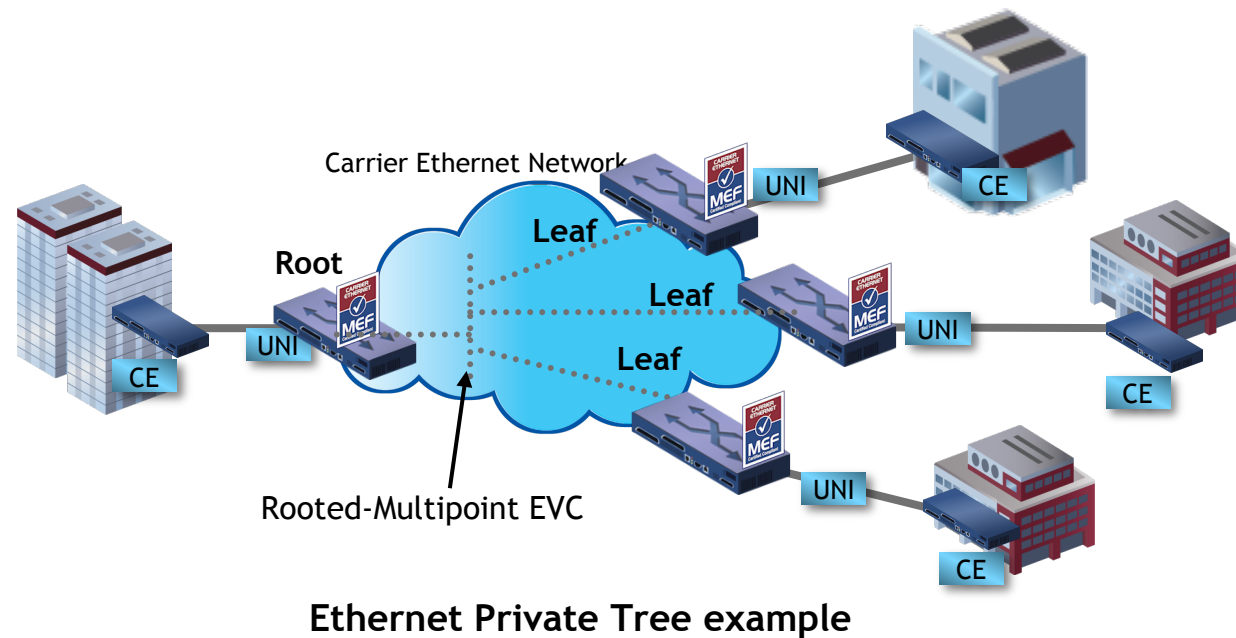


Services Using E-Tree Service Type



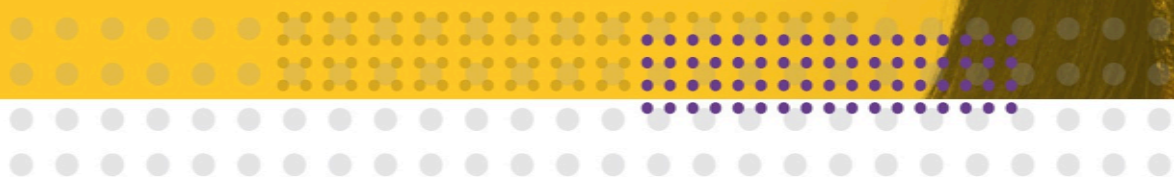
Ethernet Private Tree (EP-Tree) and Ethernet Virtual Private Tree (EVP-Tree) Services

- Enables Point-to-Multipoint Services with less provisioning than typical hub and spoke configuration using E-Lines
 - Provides traffic separation between users with traffic from one “leaf” being allowed to arrive at one of more “Roots” but never being transmitted to other “leaves”



2.4

The MEF Certification Program



The MEF Certification Program

Background

- Worldwide Ethernet Services based on large populations of multi-vendor equipment yet delivering a single set of services must be carefully engineered
- From user perspective, services must be plug and play

This is facilitated by the Certification Program

- The program consists of a series of thorough tests providing evidence for end-users, service providers and manufacturers alike, that products and services are compliant to published MEF specifications

Key Benefits of Certification

Key Benefits for the Enterprise

- Designed to assure business users that services behave and perform according to agreed service level specification and known international standards
- It provides IT departments with the knowledge to make informed decisions at greatly reduced risks
- Accelerates Carrier Ethernet deployment at reduced cost
- Provides common terminology to compare services
- *“Are you MEF Certified Compliant?”*
 - Now a common feature of RFQs.

Key Benefits of Certification

Key Benefits for Service Providers

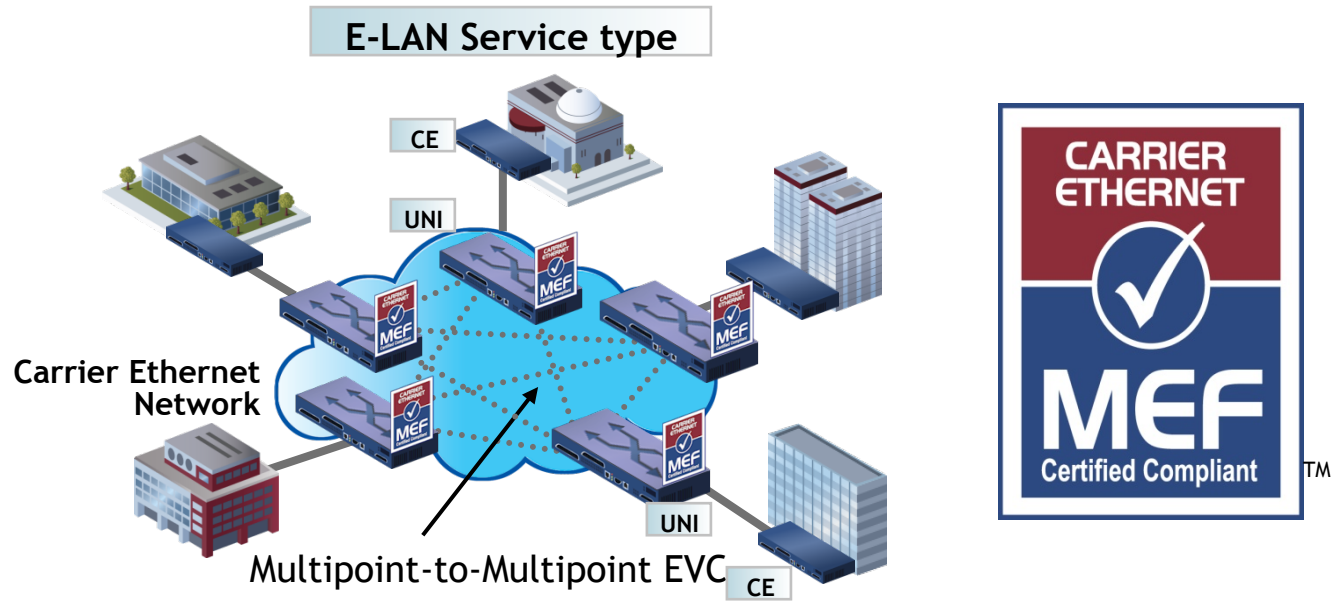
- Immediate assurance that vendors equipment complies to MEF Specifications
- Reduces service costs, charges and time on complex testing between vendors, especially on global installations
- Establishes solid foundation for Carrier Ethernet ubiquity, & interoperability
- Removes confusion caused by proprietary service names and overlapping options
- Conformance to MEF 9 allows customers to specify their service requirements by referencing independent, international standards

Key Benefits of Certification

Key Benefits for Equipment Manufacturers

- Created a globally recognized interoperability standard improves 'approval' process,
- Increases tender opportunities and competitiveness
- Dramatically reduces testing costs, time-to-market and installation time.
- Creates an independent validation of function and conformance and is creating a strong partnership with service provider customers
- Provides a performance and behaviour benchmark
- Creates a platform for future management standards testing

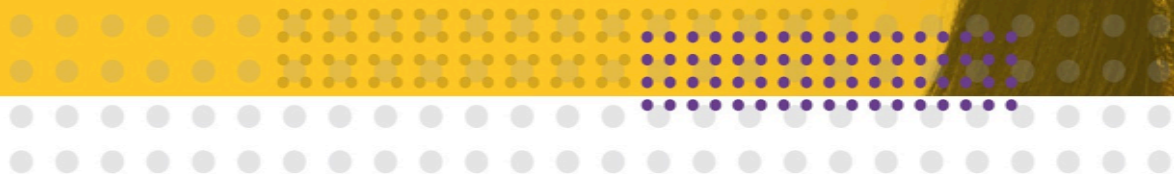
Look for the logo



MEF certified Carrier Ethernet products UNI: User Network Interface, CE: Customer Equipment

3

Traditional Metro Ethernet networks



Agenda

3. Traditional Metro Ethernet Networks

3.1 Customer Delineation

3.2 Forwarding Mechanism

3.3 Resiliency and Redundancy

3.4 Recent Developments

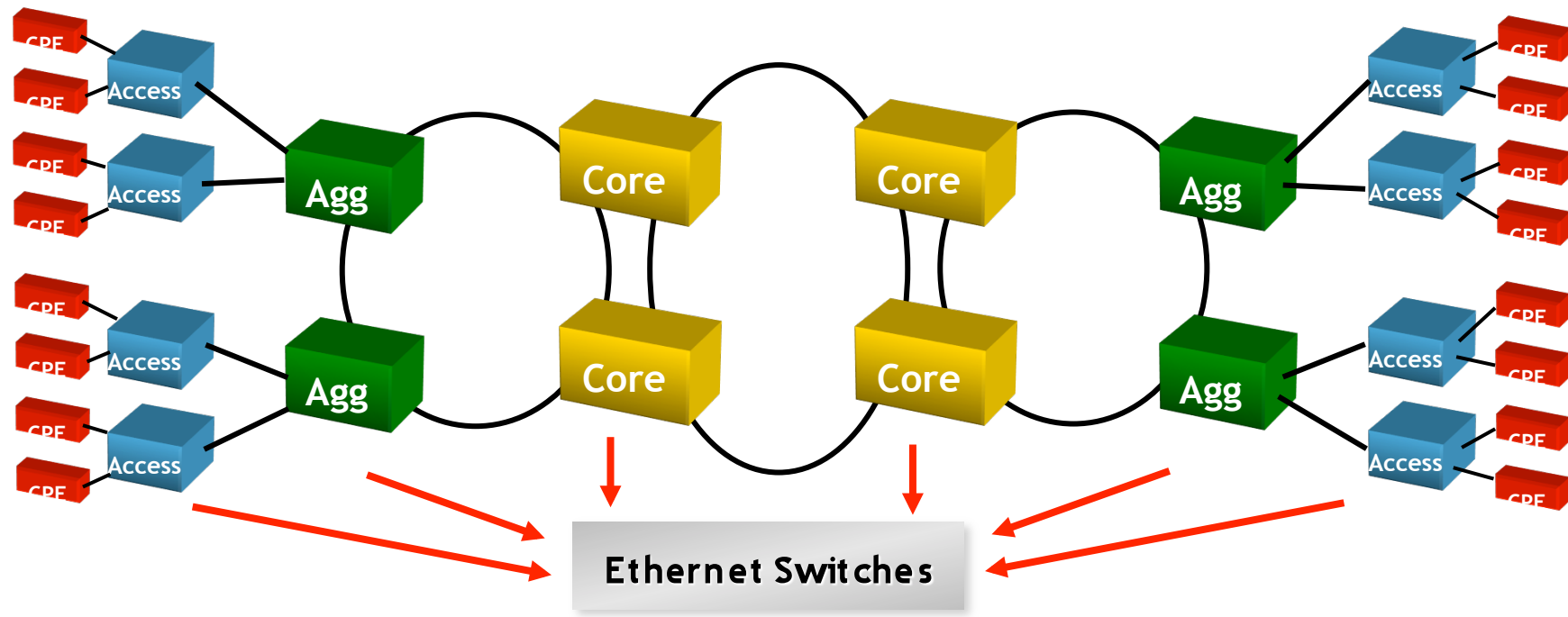
3.5 Summary

Traditional Metro Ethernet Networks



Traditional methods of Ethernet delivery:

- Ethernet switching/bridging networks (802.1q)
 - Customers delimited by VLAN IDs
 - VLAN IDs globally significant
 - Resiliency provided using variants of the Spanning Tree Protocol



3.1

Customer Delineation

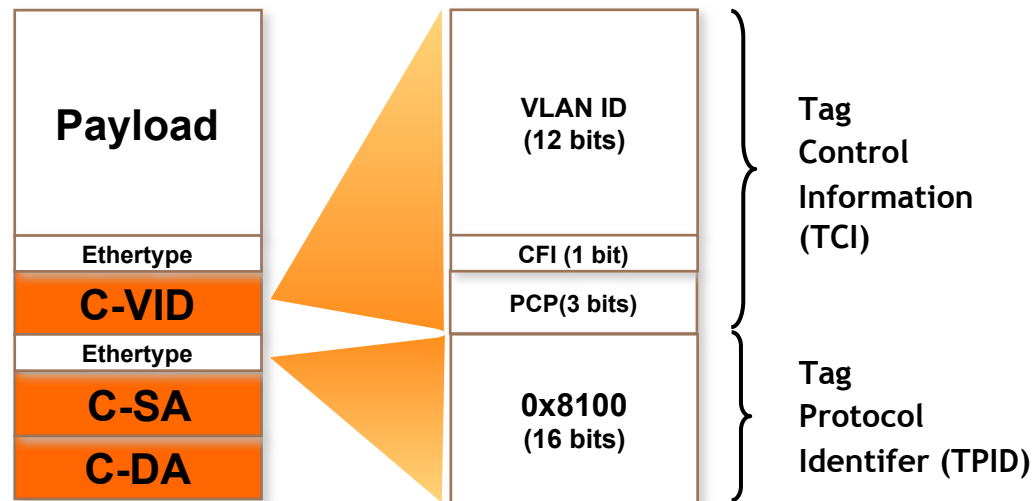


Traditional Metro Ethernet Networks



Customer delineation:

- Ethernet switching/bridging networks
- First generation was based on 802.1q switches
 - One obvious limitation was the VLAN ID space - the 12-bit VLAN ID allows a maximum of 4094 VLANs (VLANs 0 and 4095 are reserved). This limited the total number of services in any one switching/bridging domain.
 - The other problem was that of VLAN assignment - customers could not arbitrarily use any VLAN IDs since the VLAN IDs were globally significant !

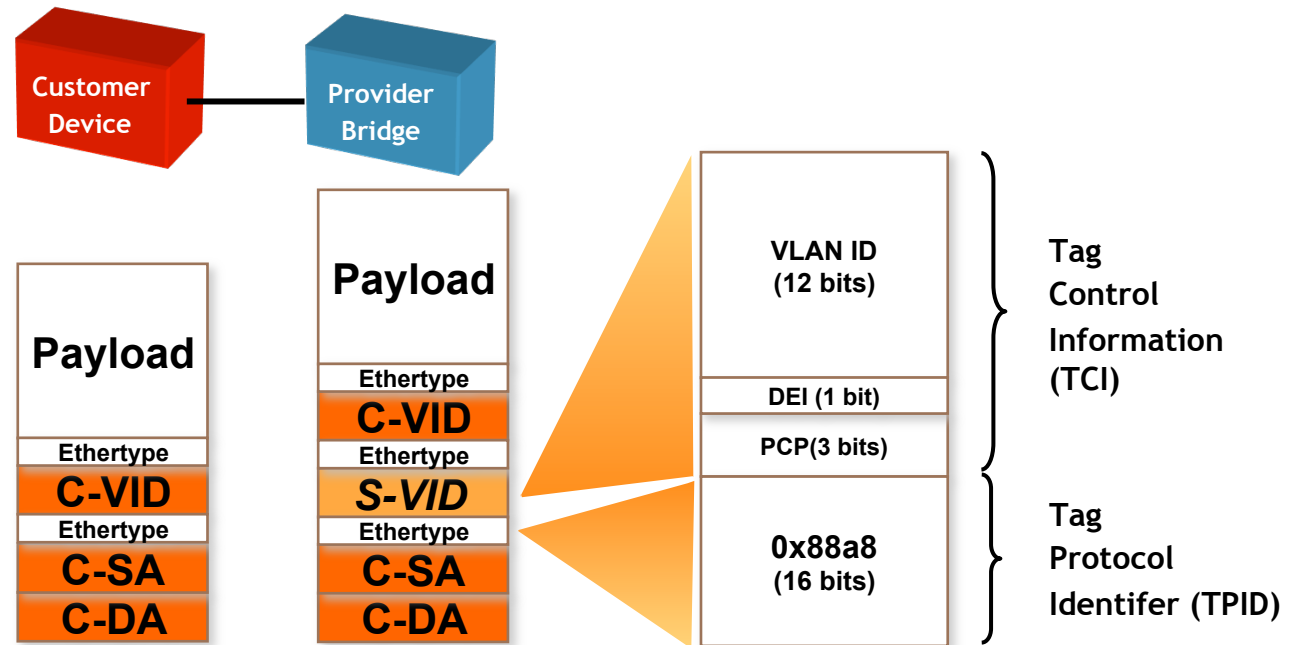


Traditional Metro Ethernet Networks



Customer delineation :

- Q-in-Q (aka VLAN stacking, aka 802.1ad) comes to the rescue !
 - Q-in-Q technology, which has now been standardised by the IEEE as 802.1ad (Provider Bridging), allowed the addition of an additional tag to customer Ethernet frames - the S-tag. The S-tag (Service Tag) was imposed by the Service Provider and therefore, it became possible to carry customer tags (C-tags) transparently through the network.



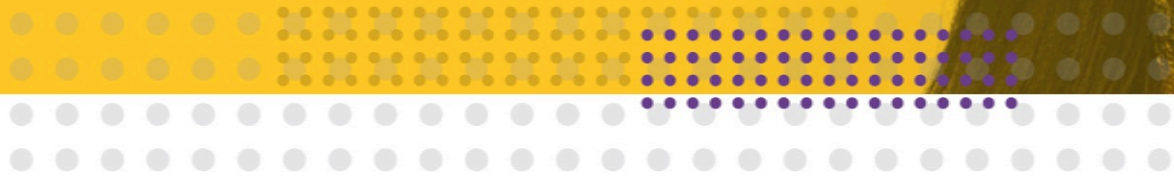
Traditional Metro Ethernet Networks

SONOG

Customer delineation:

- Some important observations about Q-in-Q:
 - This is not a new encapsulation format; it simply results in the addition of a second tag to the customer Ethernet frame.
 - There is no change to the customer destination or source MAC addresses
 - The number of distinct service instances within each Provider Bridging domain is still limited by the S-VLAN ID space i.e. 4094 S-VLANs. The difference is that customer VLANs can now be preserved and carried transparently across the provider network.

3.2 Forwarding Mechanism

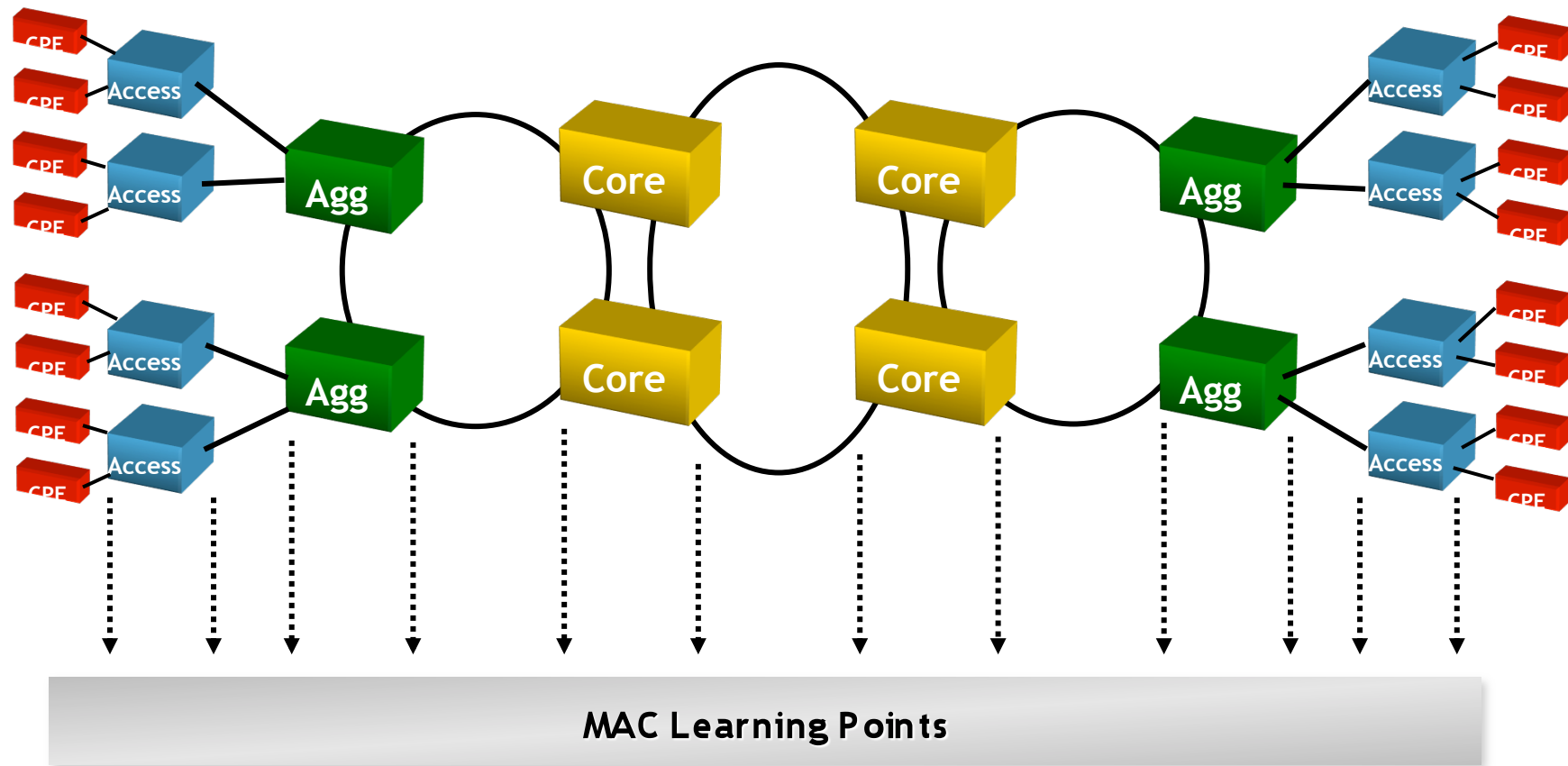


Traditional Metro Ethernet Networks



Forwarding Mechanism:

- Dynamic learning methods used to build forwarding databases

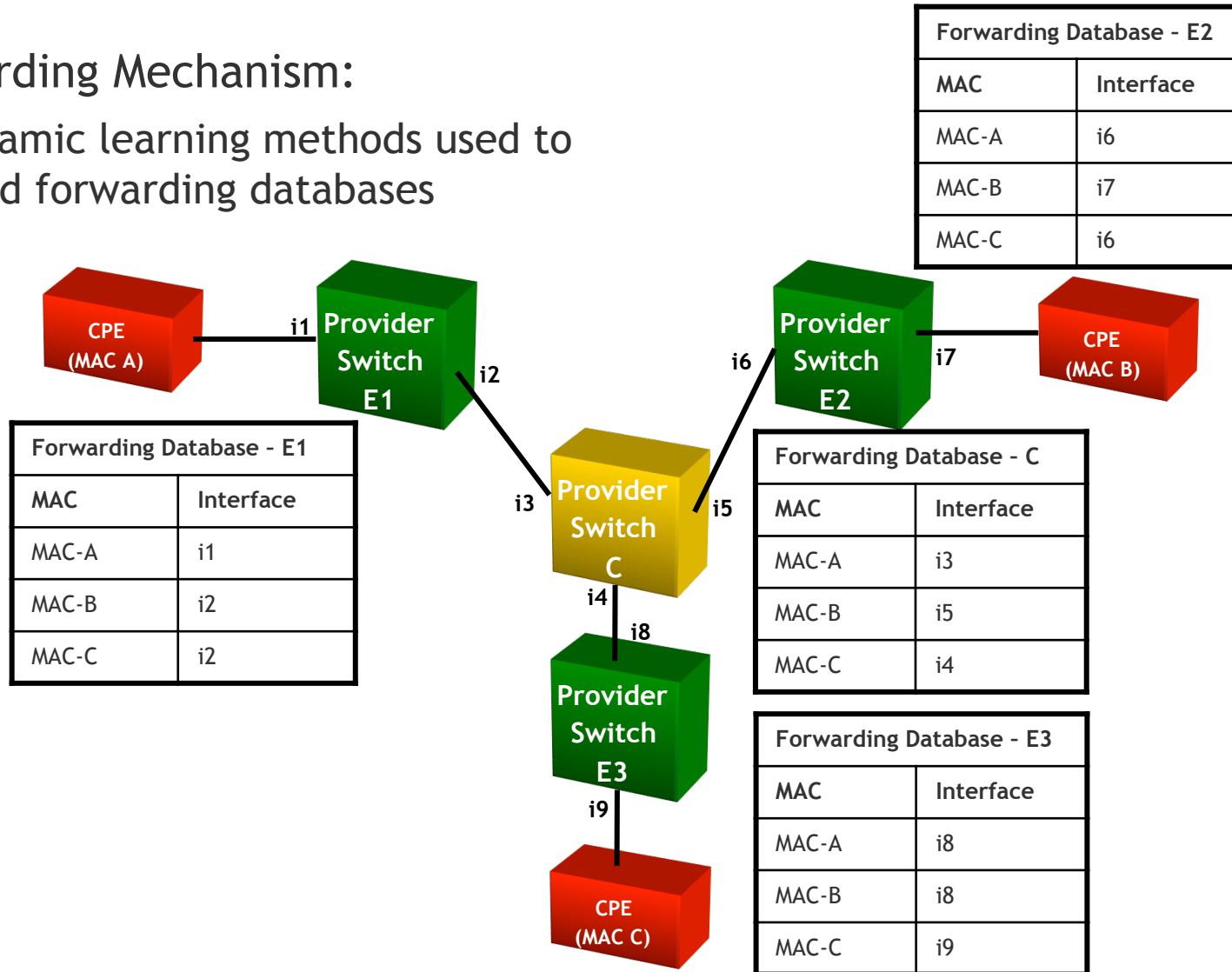


Traditional Metro Ethernet Networks



Forwarding Mechanism:

- Dynamic learning methods used to build forwarding databases

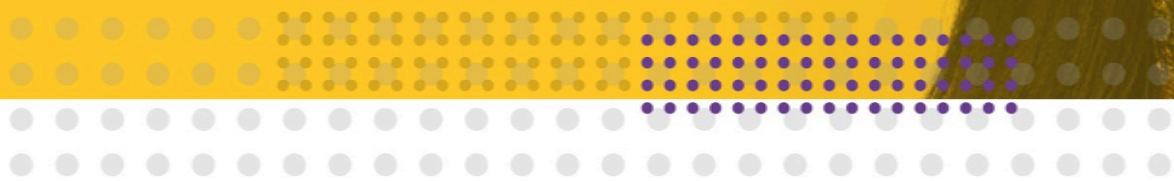


Forwarding Mechanism:

- Dynamic learning methods used to build forwarding databases
 - Data-plane process - there are no control-plane processes for discovering endpoint information
- In the worst case, ALL switches have forwarding databases that include ALL MAC addresses. This is true even for switches in the core of the network (Switch C in preceding example).
 - Switches have limited resources for storing MAC addresses. This poses severe scaling issues in all parts of the network. VLAN-stacking does not help with this problem.
 - On topology changes, forwarding databases are flushed and addresses need to be re-learned. While these addresses are re-learned, traffic to unknown destinations is flooded through the network, resulting in wasted bandwidth.

3.3

Resiliency and Redundancy



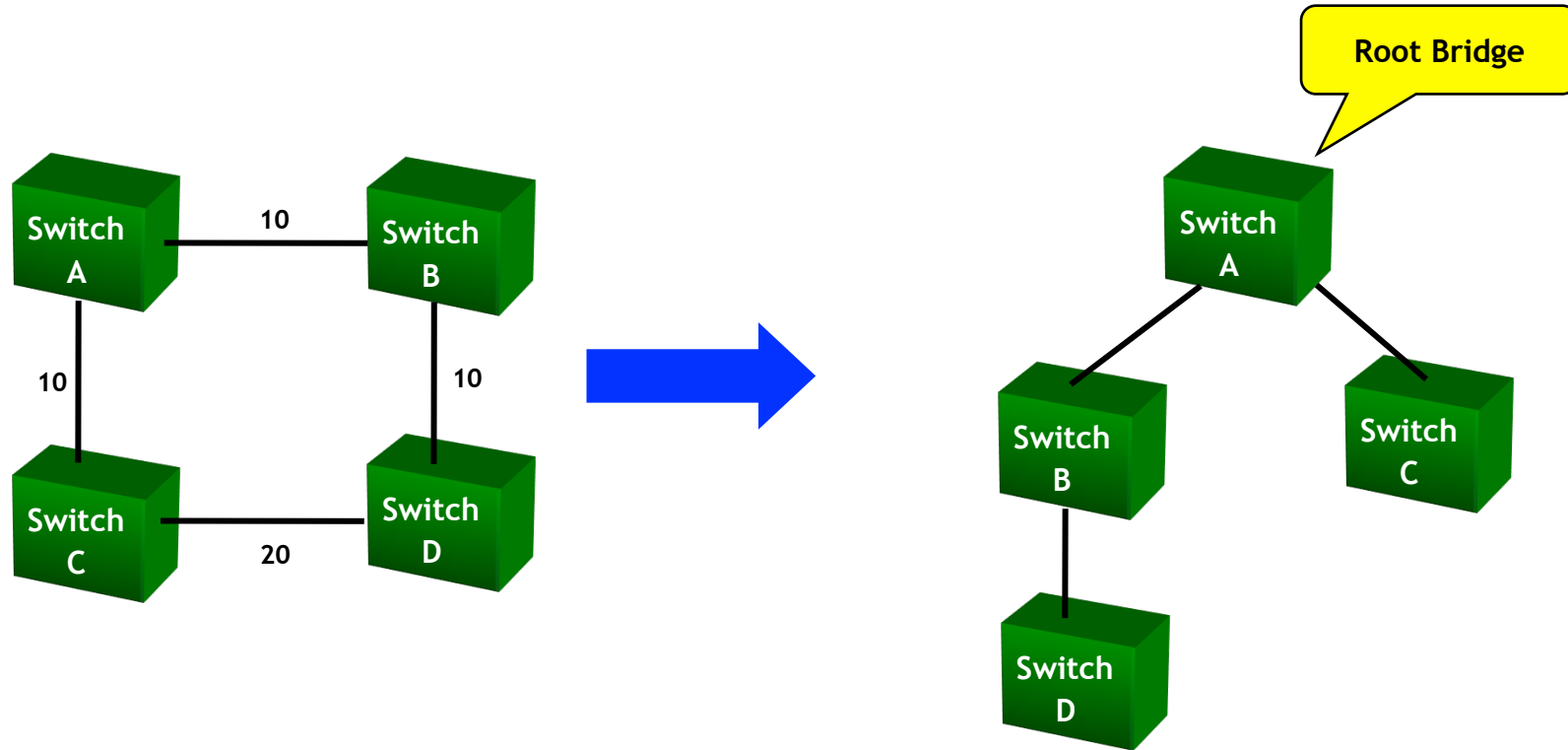
Resiliency and Redundancy

- Redundancy is needed in any network offering Carrier-grade Ethernet BUT loops are bad !!
- The Spanning Tree Protocol (STP) is used to break loops in bridged Ethernet networks
 - There have been many generations of the STP over the years
 - All of these variants work by removing redundant links so that there is one, and only one, active path from each switch to every other switch i.e. all loops are eliminated. In effect, a minimum cost tree is created by the election of a root bridge and the subsequent determination of shortest-path links to the root bridge from every other bridge
 - Bridges transmit special frames called Bridge Protocol Data Units (BPDUs) to exchange information about bridge priority, path costs etc.
- High Availability is difficult to achieve in traditional Metro Ethernet networks.

Traditional Metro Ethernet Networks



Building the Spanning Tree ...



Rudimentary Traffic-Engineering Capabilities

Traditional Metro Ethernet Networks



First generation of STP (IEEE802.1d-1998):

- Had a number of significant shortcomings:
 - Convergence times - the protocol is timer-based with times in the order of 10s of seconds. After network topology changes (failure or addition of links), it could take up to 50s for the network to re-converge
 - The protocol was VLAN-unaware, which meant that in an IEEE 802.1q network, all VLANs had to share the same spanning tree. This meant that there were network links that would not be utilised at all since they were placed into a blocked state.
 - Many vendors implemented their own, proprietary extensions to the protocol to allow the use of a separate STP instance per VLAN, allowing better link utilisation within the network
 - There were many conditions which resulted in the inadvertent formation of loops in the network. Given the flooding nature of bridged Ethernet, and the lack of a TTL-like field in Ethernet frames, looping frames could loop forever.
 - There are numerous well-publicised instances of network meltdowns in Enterprise and Service Provider networks
 - A lot of service providers have been permanently scarred by the catastrophic effects of STP loops !

Traditional Metro Ethernet Networks

SONOG

Newer generations of STP (IEEE802.1d-2004 - Rapid STP aka 802.1w):

- Some major improvements:
 - Dependence on timers is reduced. Negotiation protocols have been introduced to allow rapid transitioning of links to a forwarding state
 - The Topology Change process has been re-designed to allow faster recovery from topology changes
 - Optimisations for certain types of direct and indirect link failures
 - Convergence times are now down to sub-second in certain special cases but a lot of failure cases still require seconds to converge !
- But...
 - The protocol was still VLAN-unaware, which meant that the issue of under-utilised links was still present

Traditional Metro Ethernet Networks

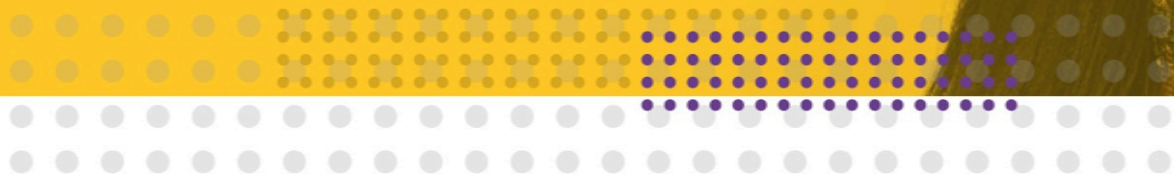
SONOG

Newer generations of STP (IEEE802.1q-2003 - Multiple STP aka 802.1s):

- Built on top of RSTP
- Added VLAN awareness:
 - Introduces the capability for the existence of multiple STP instances within the same bridged network
 - Allows the association of VLANs to STP instances, in order to provide a (relatively) small number of STP instances, instead of using an instance per VLAN.
 - Different STP instances can have different topologies, which allows much better link utilisation
- BUT
 - The stigma associated with past failures is hard to remove...
 - The protocol is fairly complicated, compared to its much simpler predecessors

3.4

Recent Developments



Traditional Metro Ethernet Networks

SONOG

Provider Backbone Bridging

- Takes IEEE 802.1ad to the next level
- MAC-in-MAC technology:
 - Customer Ethernet frames are encapsulated in a provider Ethernet frame
- Alleviates the MAC explosion problem
 - Core switches no longer need to learn customer MAC addresses
- Does not address the STP issue, however.

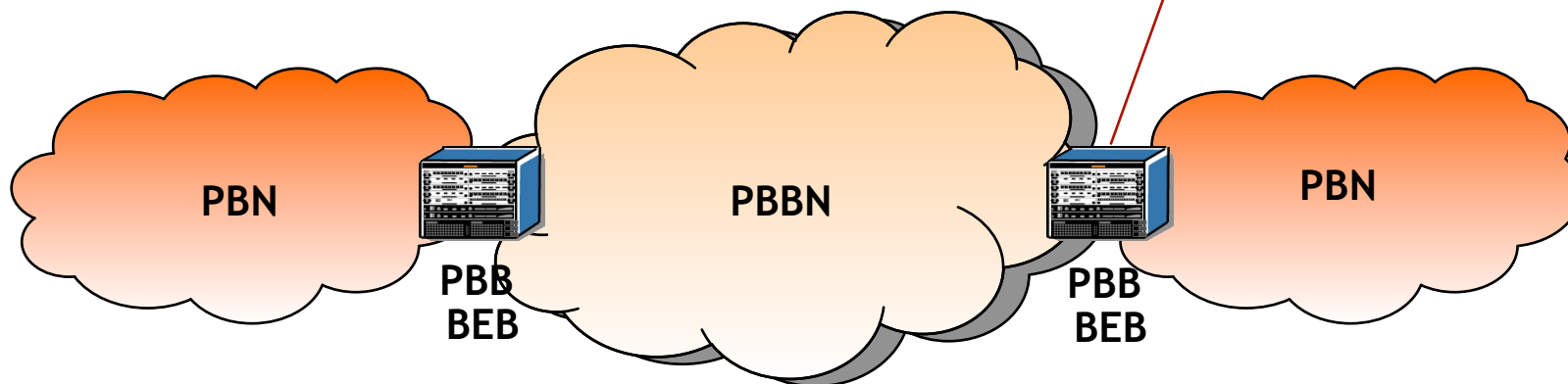
Provider Backbone Bridging (PBB)

Ethernet Technology being standardized in IEEE 802.1ah Task Group

- Designed to interconnect Provider Bridge Networks (PBN - IEEE 802.1ad)
- Adds a Backbone Header to a Customer/QinQ Ethernet Frame
 - Provider Addressing for Backbone Forwarding
 - New extended tag for Service Virtualization
- Standardization ongoing

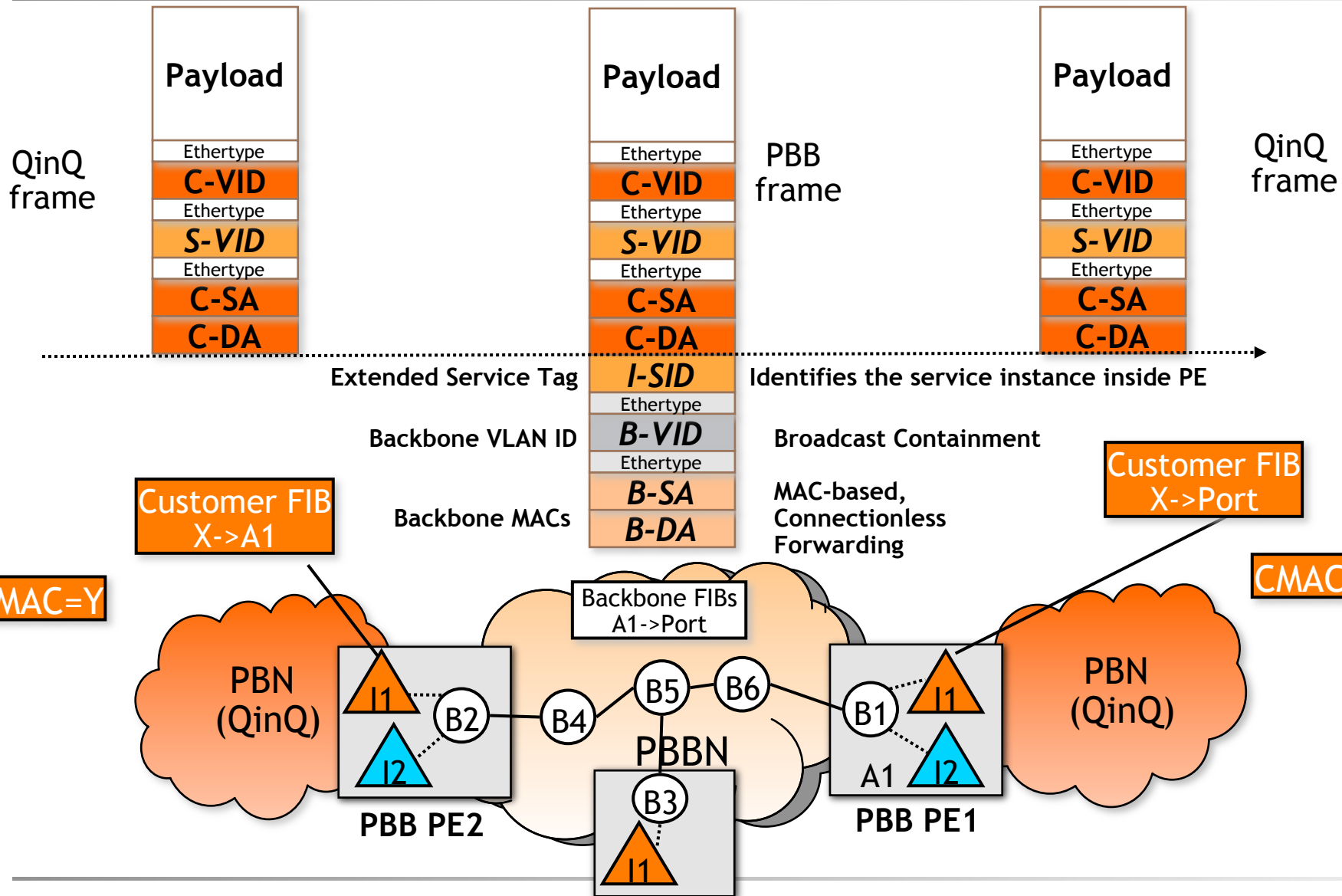
BEB:
Backbone Edge Bridge

Forward frames based
on backbone MAC
addresses



**PBBN is Ethernet based:
Connectionless Forwarding based on MAC Learning & Forwarding,
Loop Avoidance based on STP,
VLAN ID for Broadcast Containment**

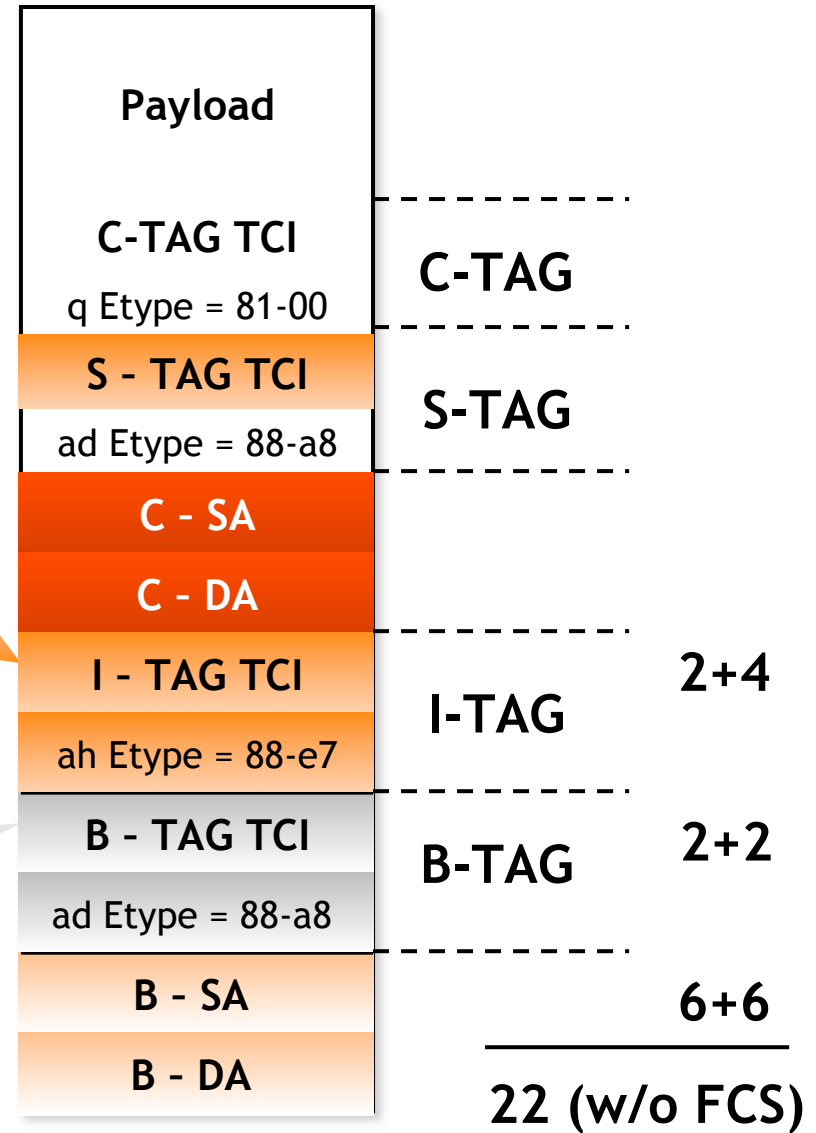
IEEE 802.1ah Model for PBB - I and B Components



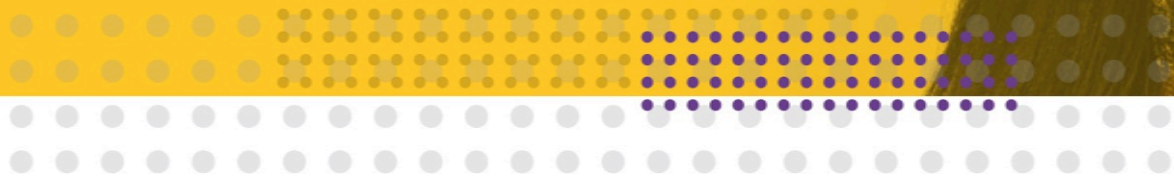
802.1ah Provider Backbone Bridge Encapsulation



- I-PCP = Customer Priority
- I-DEI = Drop Eligibility
- NCA = No Customer Addresses
- I-SID = Service Instance ID



3.5 Summary



Traditional Metro Ethernet Networks

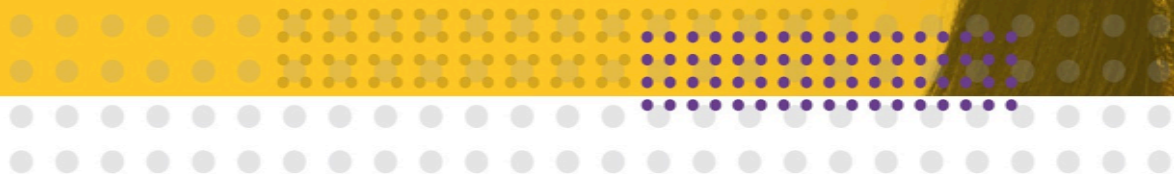
SONOG

Summary of Issues:

- High Availability is difficult to achieve in networks running the Spanning Tree Protocol
- Scalability - IEEE 802.1q/802.1ad networks run into scalability limitations in terms of the number of supported services
 - Customer Ethernet frames are encapsulated in a provider Ethernet frame
- QoS - only very rudimentary traffic-engineering can be achieved in bridged Ethernet networks.
- A lot of deployed Ethernet switching platforms lack carrier-class capabilities required for the delivery of Carrier Ethernet services

4

Delivering Ethernet over MPLS



Agenda

4. Delivering Ethernet over MPLS

4.1 Why MPLS ?

4.2 The Pseudowire Reference Model

4.3 Ethernet Virtual Private Wire Service

4.4 Ethernet Virtual Private LAN Service

4.5 Scaling VPLS

4.6 VPLS Topologies

4.1

Why MPLS ?



Delivering Ethernet over MPLS

SONOG

Why MPLS ?

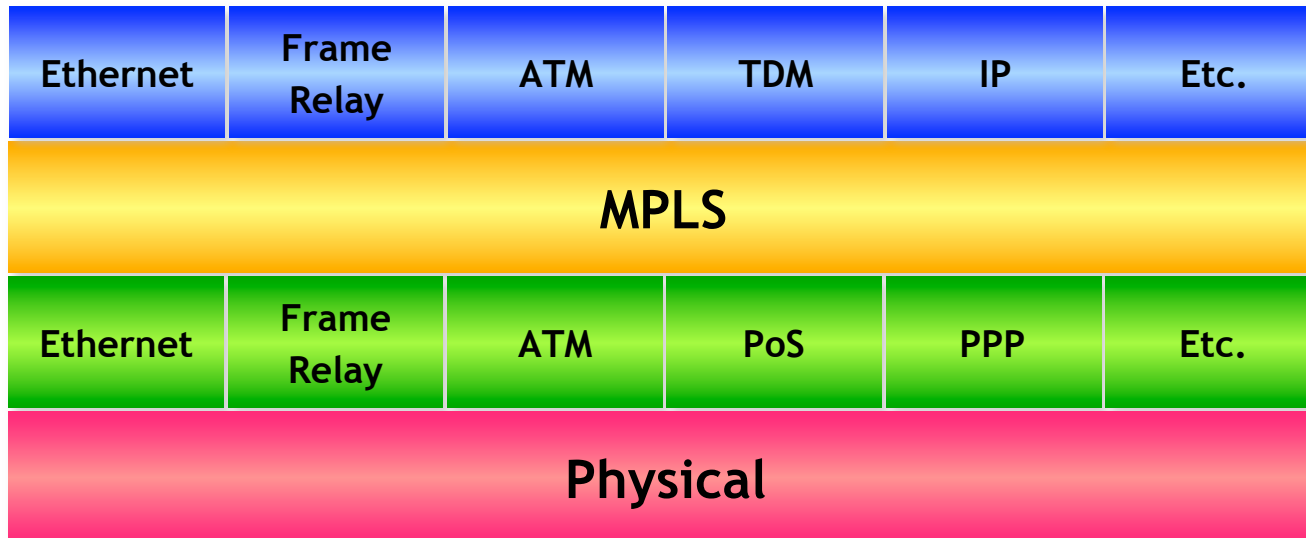
- Convergence: From “MPLS over everything” to “Everything over MPLS” !
 - One network, multiple services
- Excellent virtualisation capabilities
 - Today’s MPLS network can transport IP, ATM, Frame Relay and even TDM !
- Scalability
 - MPLS is used in some of the largest service provider networks in the world
- Advanced Traffic Engineering capabilities using RSVP-TE
- Rapid recovery based on MPLS Fast ReRoute (FRR)
 - Rapid restoration around failures by local action at the Points of Local Repair (PLRs)
 - Sub-50ms restoration on link/node failures is a key requirement for carriers who are used to such performance in their SONET/SDH networks
- Feature-richness
 - MPLS has 10 years of development behind it and continues to evolve today
- Layer 3 VPNs have already proven themselves as the killer app for MPLS - there is no reason why this success cannot be emulated by Layer 2 VPNs

MPLS is truly MultiProtocol



The “Multiprotocol” nature of MPLS:

- MPLS is multiprotocol in terms of both the layers above and below it !
- The ultimate technology for convergence

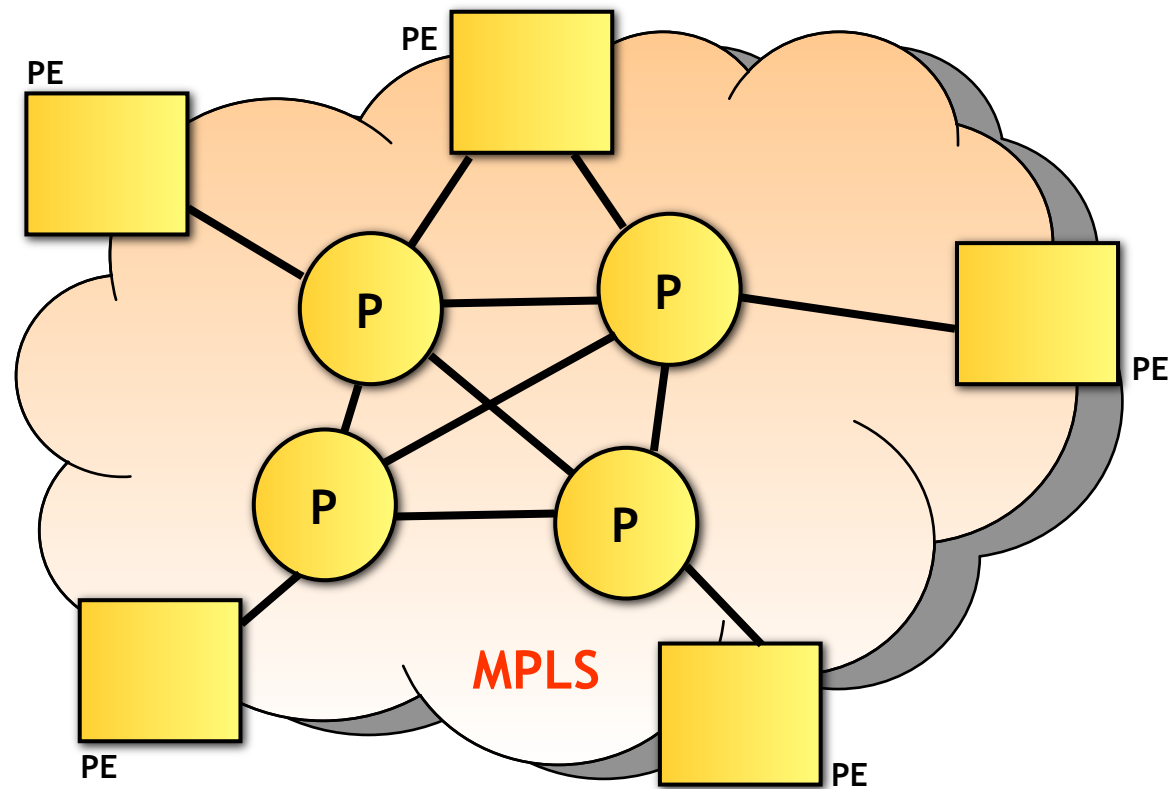


MPLS Virtualisation



The virtualisation capabilities of MPLS:

- One common network supports multiple, different overlaid services

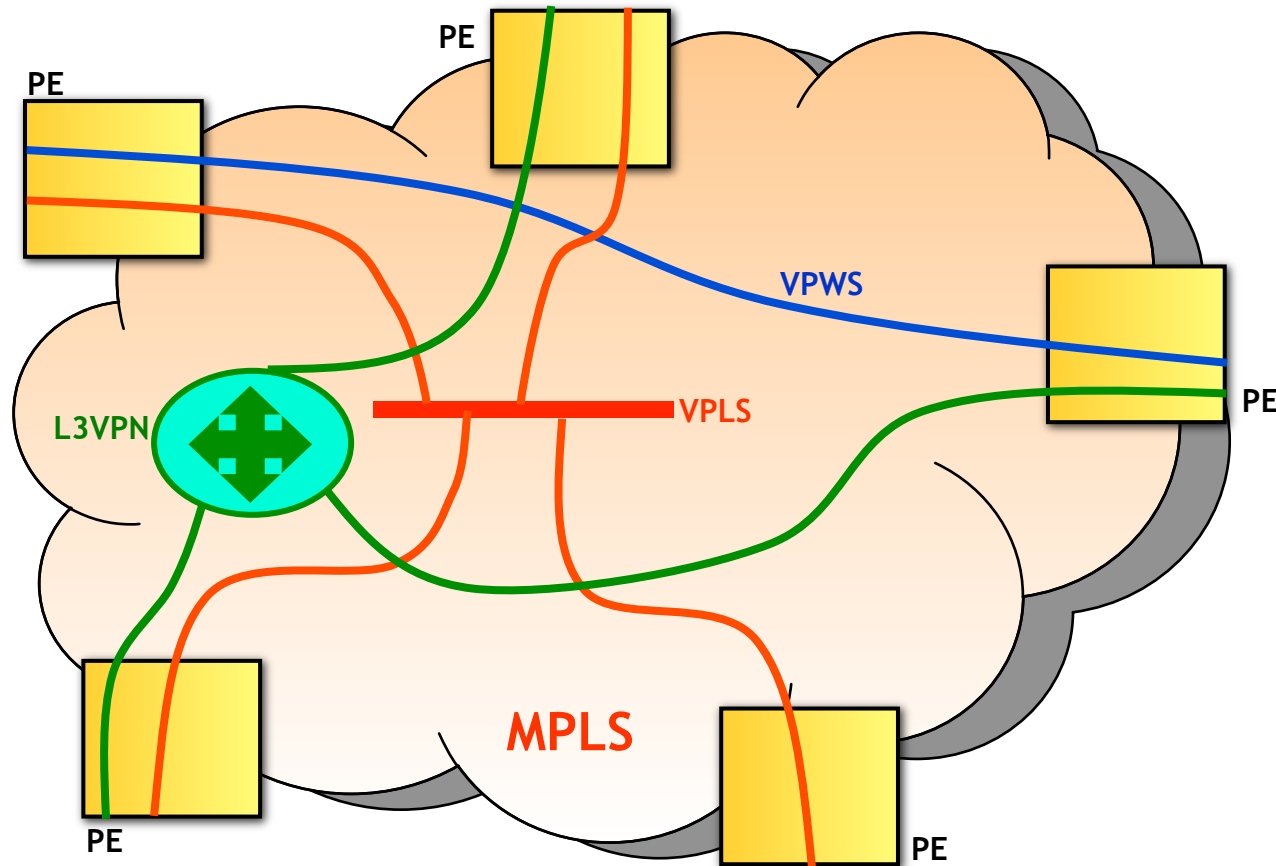


MPLS Virtualisation



The virtualisation capabilities of MPLS:

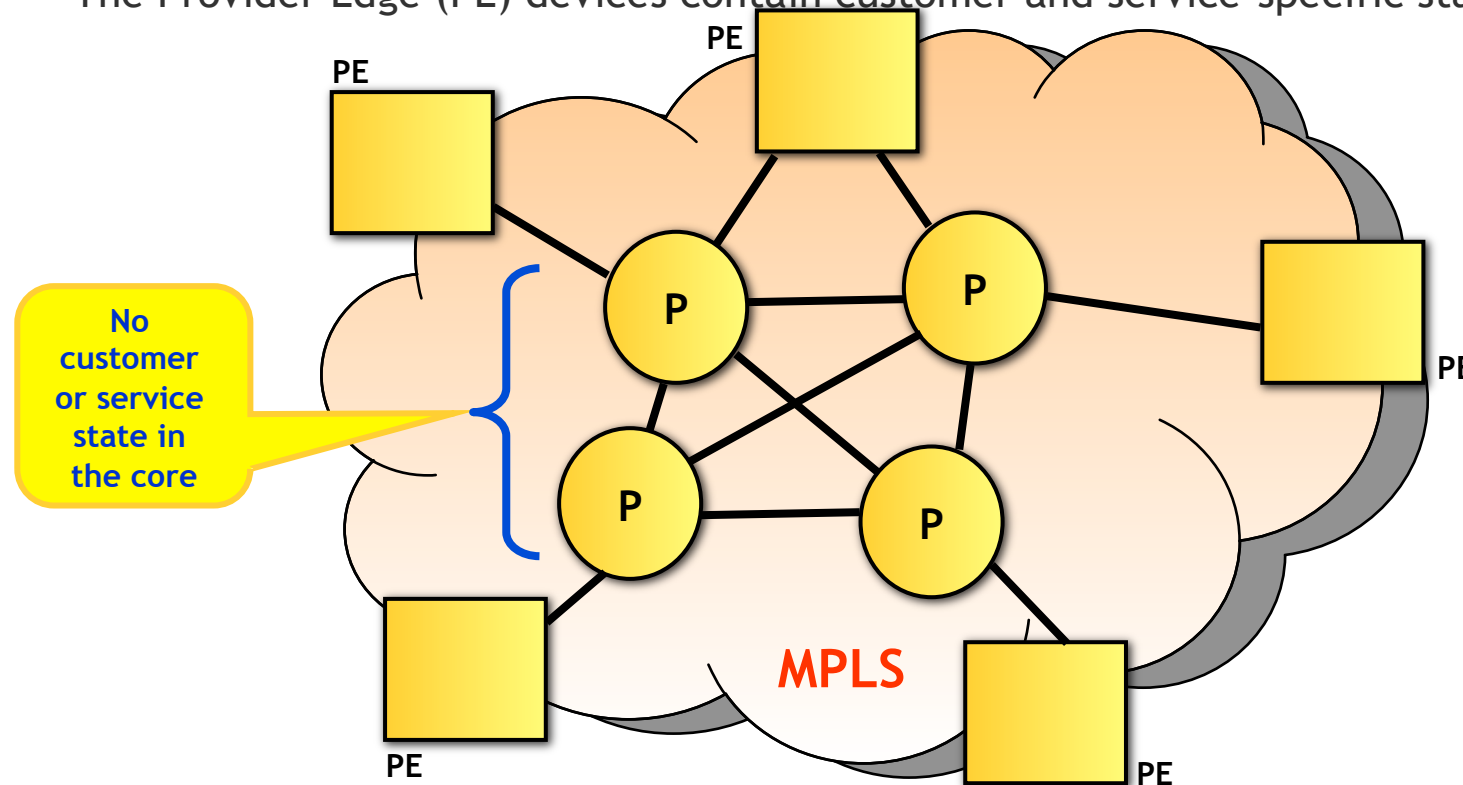
- One common network supports multiple, different overlaid services



MPLS Scalability

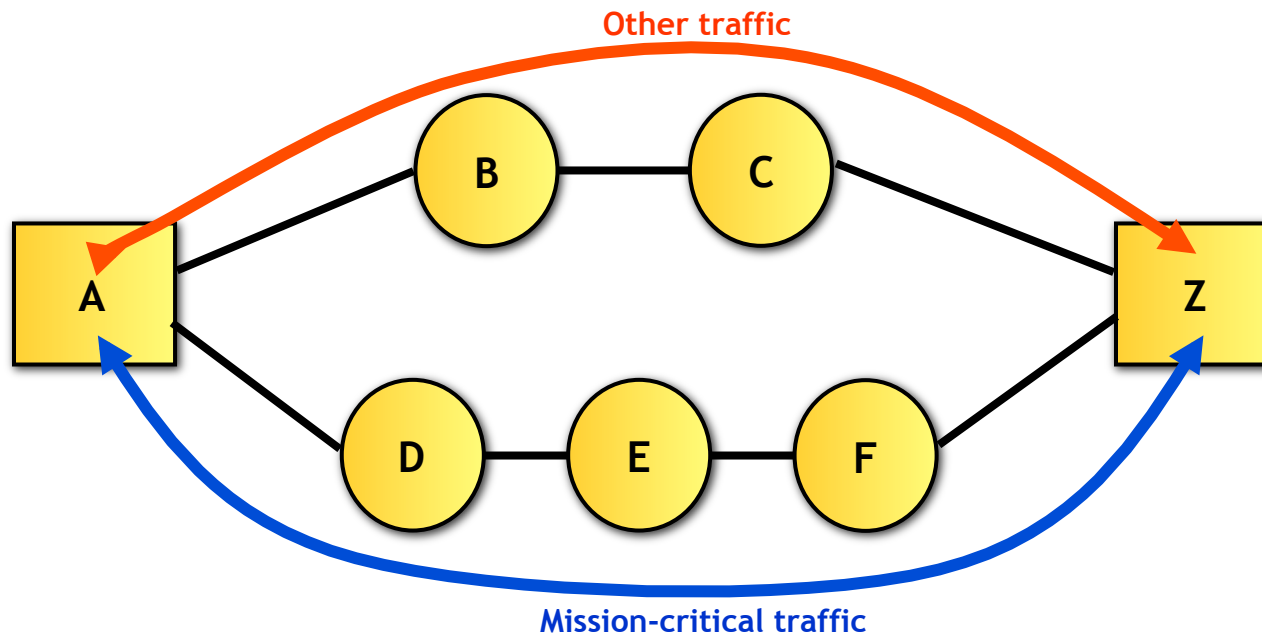
MPLS Scalability:

- Service state is kept only on the Provider Edge devices
- The Provider (P) devices simply contain reachability information to each other and all PEs in the network
- The Provider Edge (PE) devices contain customer and service-specific state



Traffic-Engineering capabilities

- The Problem: consider example below - all mission-critical traffic between nodes A and Z has to use the path A-D-E-F-Z, while all other traffic uses the path A-B-C-Z.

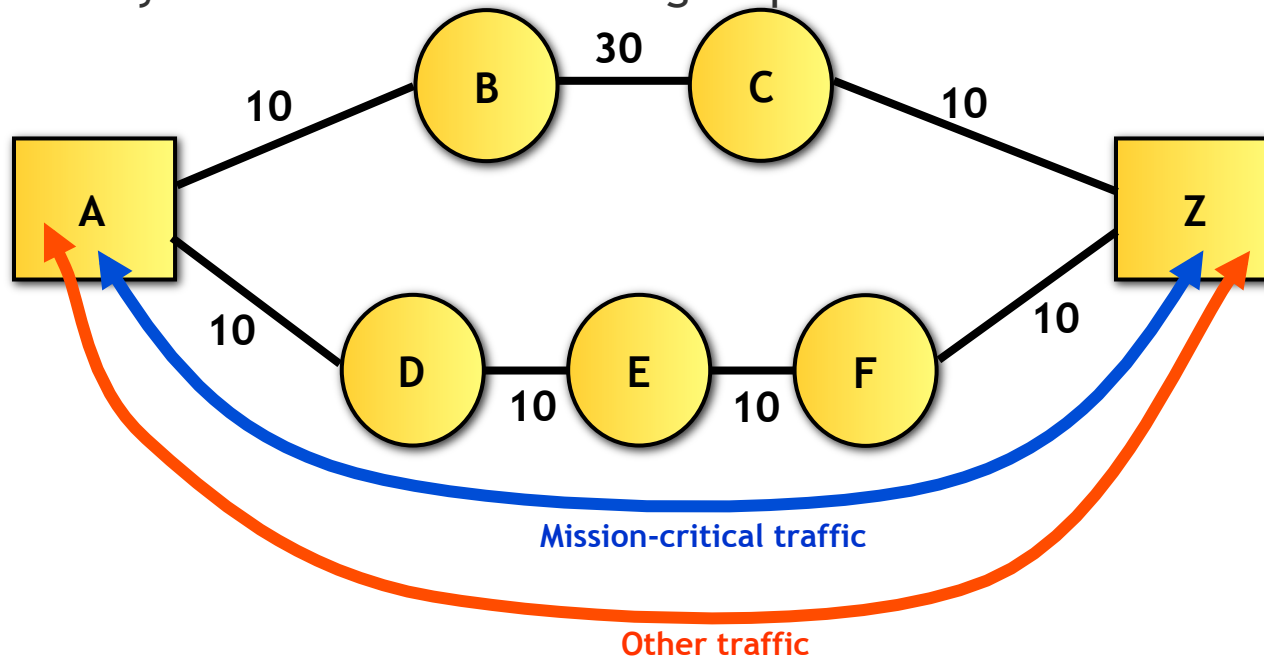


The IGP-based solution

- Use link metrics to influence traffic path
- It's all or nothing - Traffic cannot be routed selectively

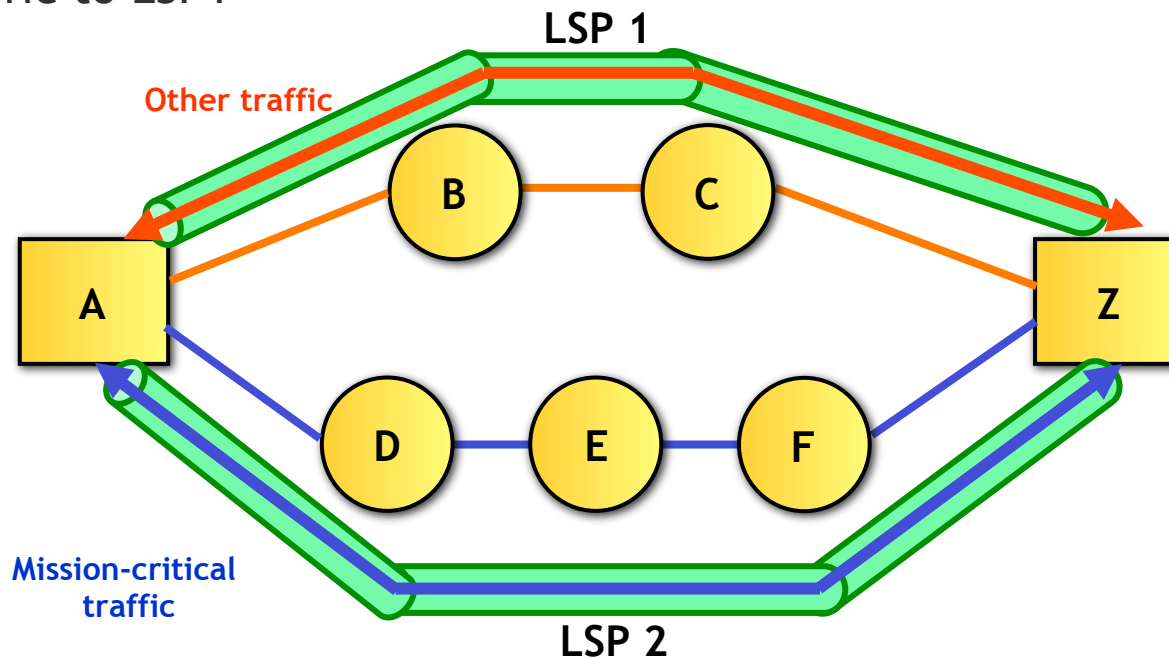
Other solutions

- Policy-based routing - will work but is cumbersome to manage and has to be carefully crafted to avoid routing loops



The MPLS solution

- Use constrained path routing to build Label Switched Paths (LSPs)
 - Constrain LSP1 to use only the “orange” physical links
 - Constrain LSP2 to use only the “blue” physical links
 - At the PEs, map the mission-critical traffic to LSP2 and...
 - ...all other traffic to LSP1

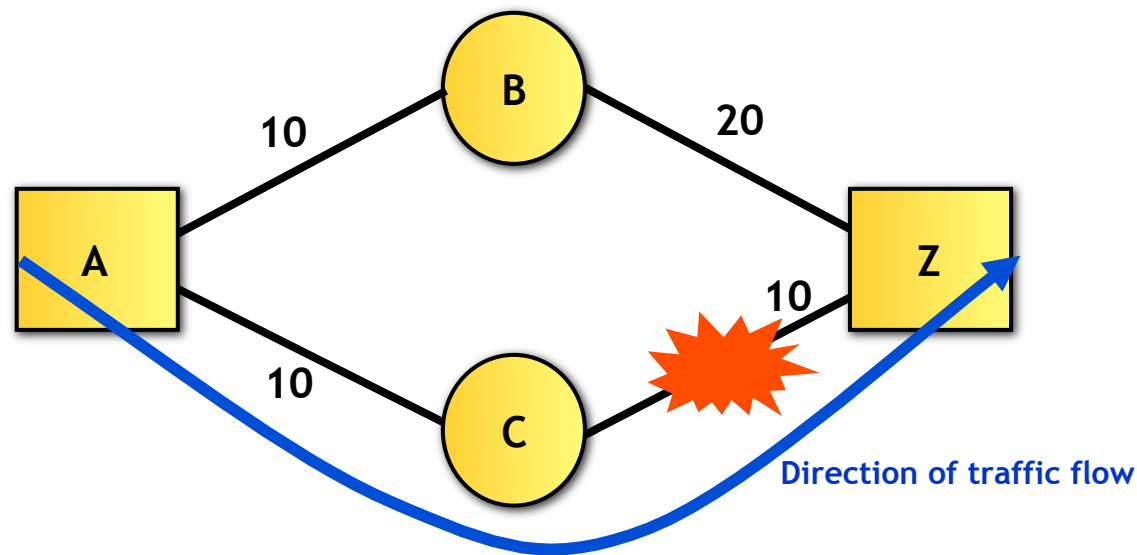


Recovery from failures - typical IGP

- Step 1 - Detection of the failure
 - One or more routers detect that a failure (link or node) has occurred
- Step 2 - Propagation of failure notification
 - The router(s) detecting the failure inform other routers in the domain about the failure
- Step 3 - Recomputation of Paths/Routes
 - All routers which receive the failure notification now have to recalculate new routes/paths by running SPF algorithms etc
- Step 4 - Updating of the Forwarding Table
 - Once new routes are computed, they are downloaded to the routers' forwarding table, in order to allow them to be used
- All of this takes time...

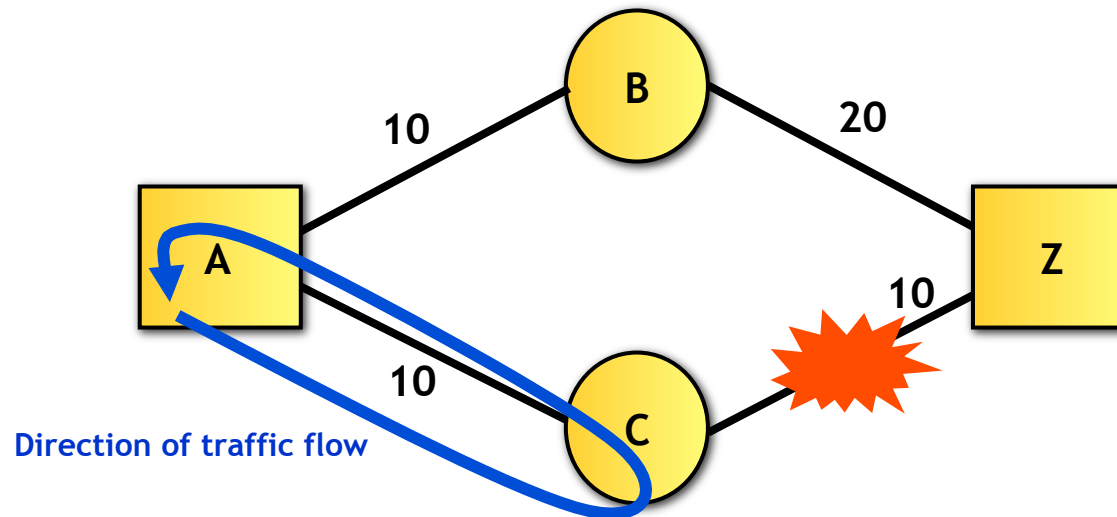
Failure and Recovery Example - IGP-based

- What happens immediately after the link between C and Z fails ?
 - Step 1 - Assuming a loss of signal (or similar physical indication) nodes C and Z immediately detect that the link is down
 - Node A does not know that the link is down yet and keeps sending traffic destined to node Z to Node C. Assuming that node C has not completed step 4 yet, this traffic is dropped.



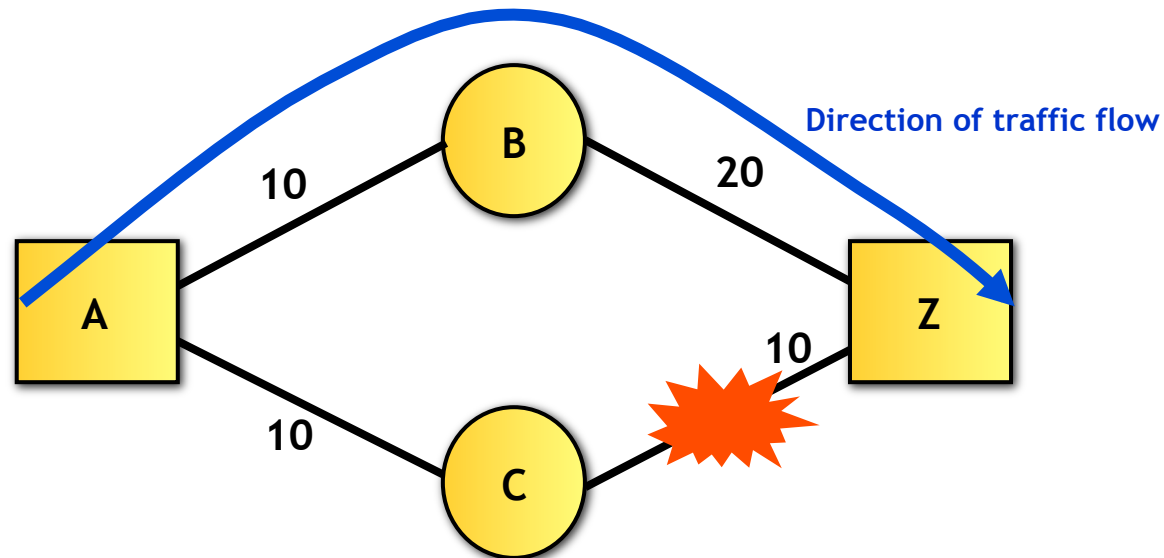
Failure and Recovery Example (continued) - IGP-based

- Node C (and node Z) will be the first to recalculate its routing table and update its forwarding table (step 4).
 - In the meantime, Node A does not know that the link is down yet and keeps sending traffic destined to node Z to Node C. Given that node C has completed step 4, it now believes (quite correctly) that the best path to Z is via node A. BUT - node A still believes that the best path to node Z is via node C so it sends the traffic right back to node C. We have a transient loop (micro-loop)
 - The loop resolves itself as soon as node A updates its forwarding table but in the meantime, valuable packets have been dropped



Failure and Recovery Example (continued)

- Node A and all other nodes eventually update their forwarding tables and all is well again.
- But the damage is already done. . .

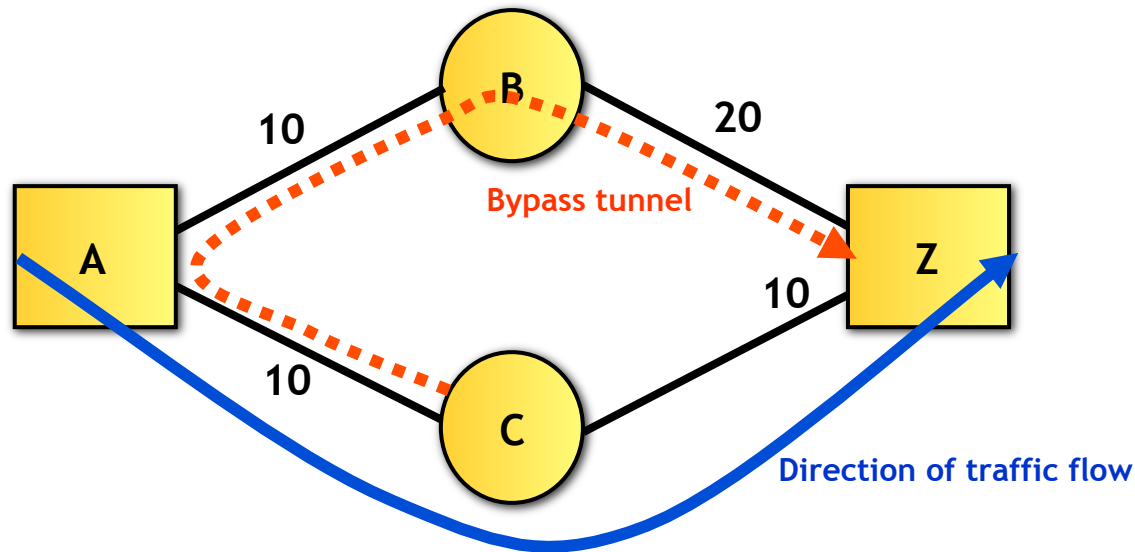


Recovery from failures - how can MPLS help ?

- RSVP-TE Fast Re-Route (FRR) pre-computes detours around potential failure points such as next-hop nodes and links
- When link or node failures occur, the routers (Points of Local Repair) directly connected to the failed link rapidly (sub-50ms) switch all traffic onto the detour paths.
- The network eventually converges and the head-end router (source of the traffic) switches traffic onto the most optimal path. Until that is done, traffic flows over the potentially sub-optimal detour path BUT the packet loss is kept to a minimum

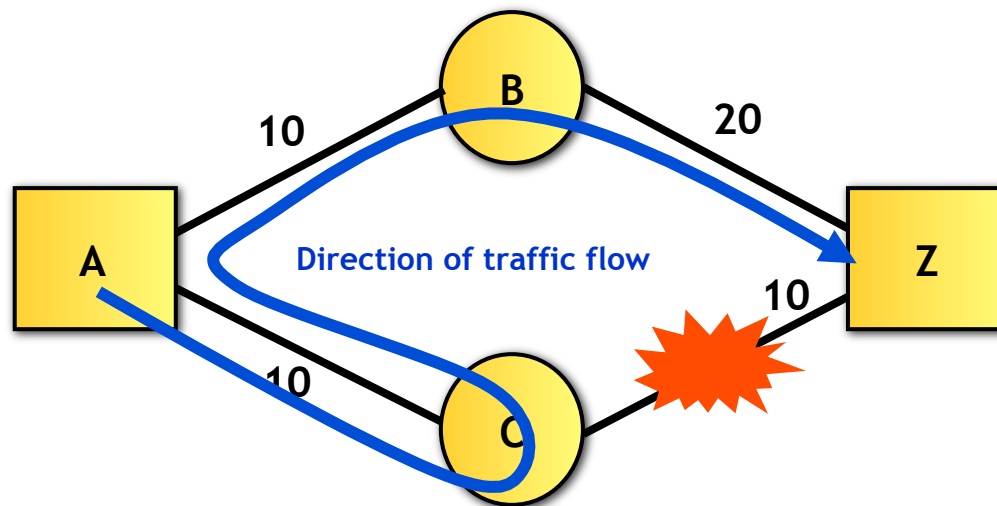
Failure and Recovery Example - with MPLS FRR

- Node C pre-computes and builds a detour around link C-Z



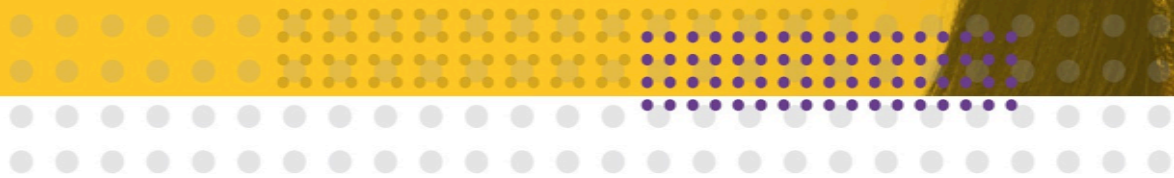
Failure and Recovery Example - with MPLS FRR

- When link C-Z fails, node C reroutes traffic onto the detour tunnel
- Traffic does a U-turn but still makes it to the destination



4.2

The Pseudowire Reference Model



The Pseudowire Reference Model

Pseudowires:

- Key enabling technology for delivering Ethernet services over MPLS
- Specified by the **pwe3** working group of the IETF
- Originally designed for Ethernet over MPLS (EoMPLS) - initially called Martini tunnels
- Now extended to many other services - ATM, FR, Ethernet, TDM
- Encapsulates and transports service-specific PDUs/Frames across a Packet Switched Network (PSN) tunnel
- The use of pseudowires for the emulation of point-to-point services is referred to as Virtual Private Wire Service (VPWS)
- IETF definition (RFC3985):

"...a mechanism that emulates the essential attributes of a telecommunications service (such as a T1 leased line or Frame Relay) over a PSN. PWE3 is intended to provide only the minimum necessary functionality to emulate the wire with the required degree of faithfulness for the given service definition."

PWE3 Terminology

Pseudowire Terminology

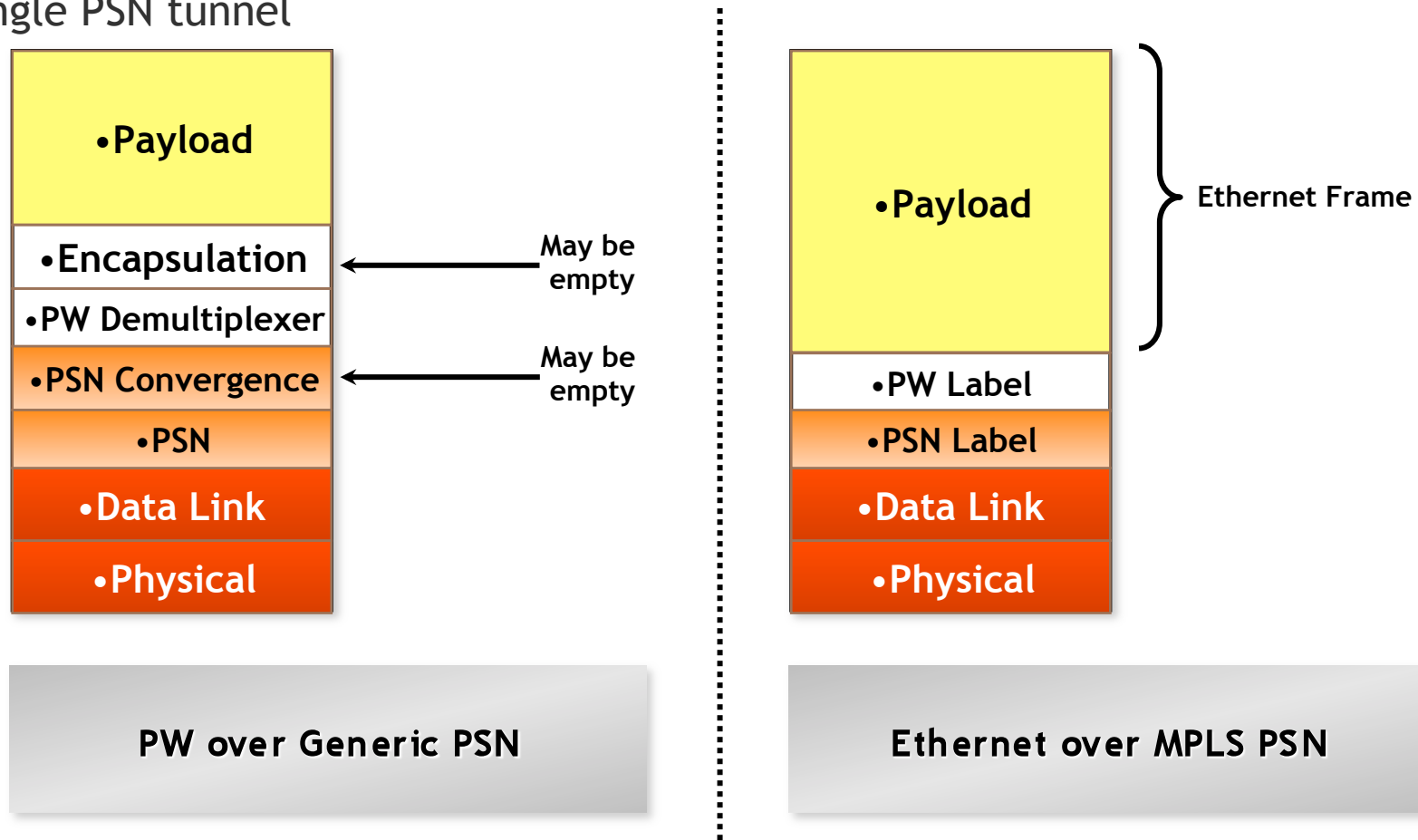
- Attachment circuit (AC)
 - The physical or virtual circuit attaching a CE to a PE.
- Customer Edge (CE)
 - A device where one end of a service originates and/or terminates.
- Forwarder (FWRD)
 - A PE subsystem that selects the PW to use in order to transmit a payload received on an AC.
- Packet Switched Network (PSN)
 - Within the context of PWE3, this is a network using IP or MPLS as the mechanism for packet forwarding.
- Provider Edge (PE)
 - A device that provides PWE3 to a CE.
- Pseudo Wire (PW)
 - A mechanism that carries the essential elements of an emulated service from one PE to one or more other PEs over a PSN.
- PSN Tunnel
 - A tunnel across a PSN, inside which one or more PWs can be carried.
- PW Demultiplexer
 - Data-plane method of identifying a PW terminating at a PE.

Pseudowire Protocol Layering



Pseudowire - Protocol Layering:

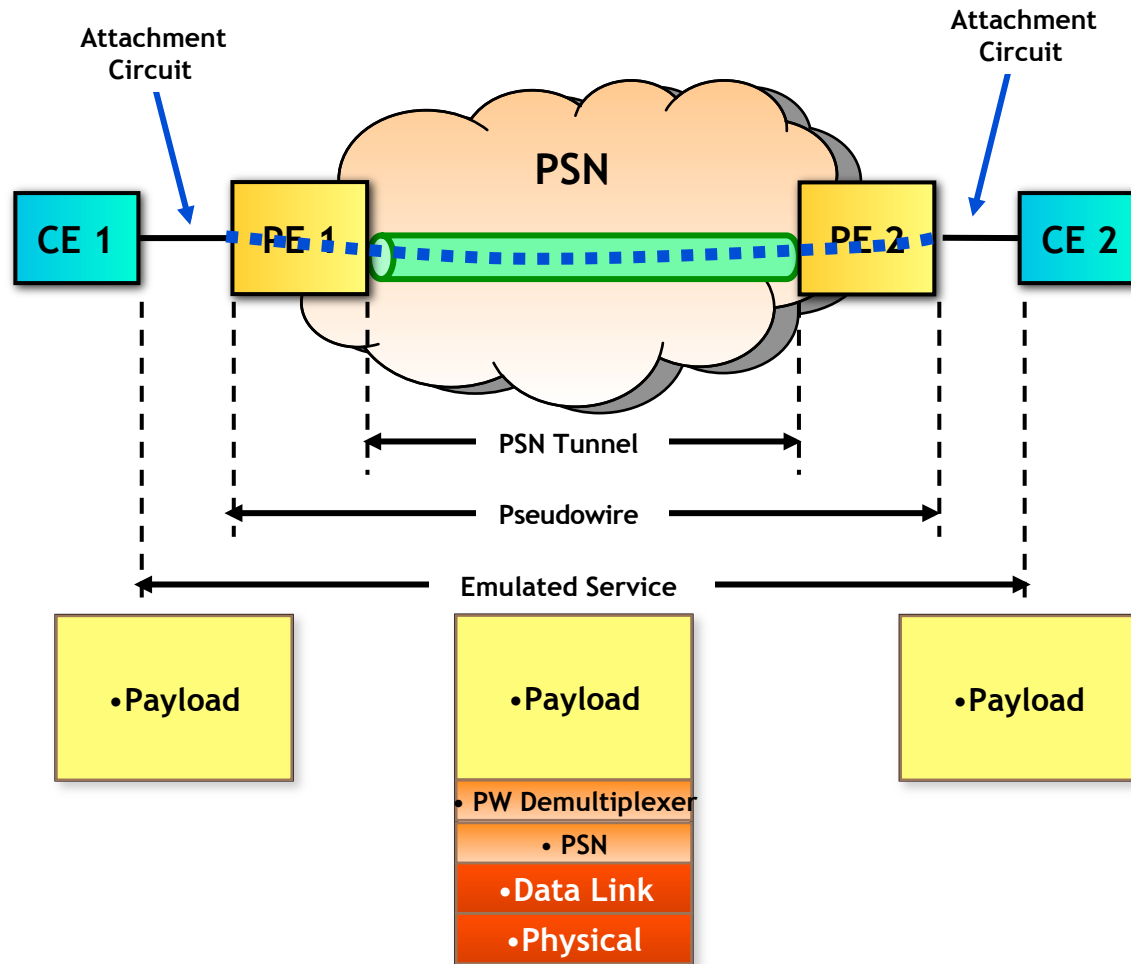
- The PW demultiplexing layer provides the ability to deliver multiple PWs over a single PSN tunnel



PWE3 Reference Model

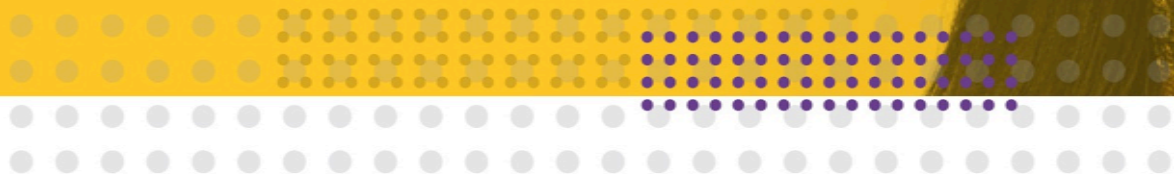
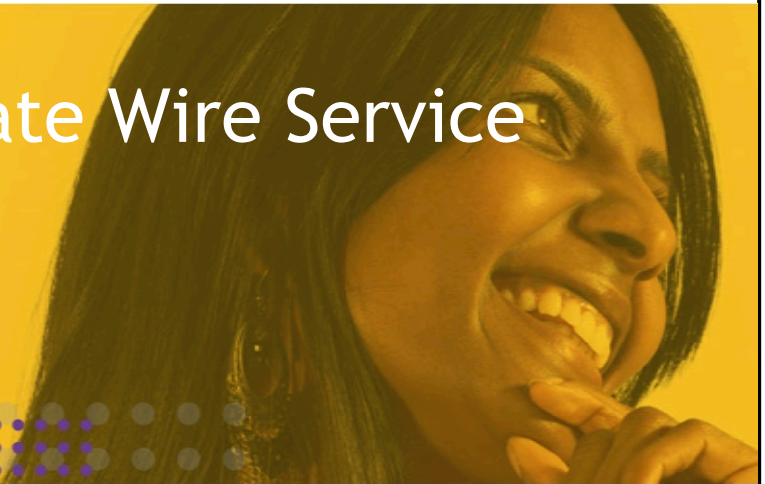


Generic PWE3 Architectural Reference Model:



4.3

Ethernet Virtual Private Wire Service (VPWS)



Ethernet Pseudowires:

- Encapsulation specified in RFC4448 - “Encapsulation Methods for Transport of Ethernet over MPLS Networks”
- Ethernet pseudowires carry Ethernet/802.3 Protocol Data Units (PDUs) over an MPLS network
- Enables service providers to offer “emulated” Ethernet services over existing MPLS networks
- RFC4448 defines a point-to-point Ethernet pseudowire service
- Operates in one of two modes:
 - Tagged mode - In tagged mode, each frame **MUST** contain at least one 802.1Q VLAN tag, and the tag value is meaningful to the two PW termination points.
 - Raw mode - On a raw mode PW, a frame **MAY** contain an 802.1Q VLAN tag, but if it does, the tag is not meaningful to the PW termination points, and passes transparently through them.

Ethernet Virtual Private Wire Service

SONOG

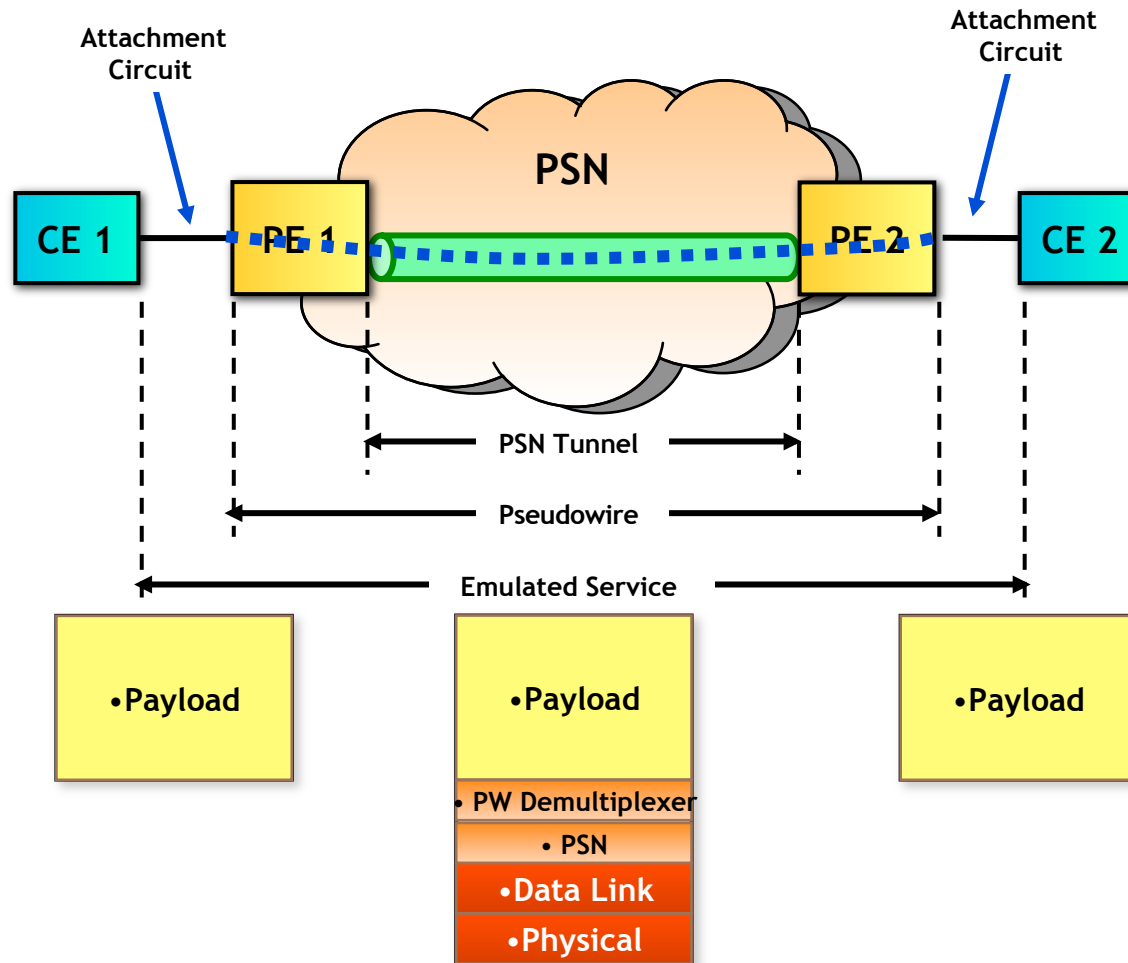
Ethernet Pseudowires (continued):

- Two types of services:
 - “port-to-port” - all traffic ingressing each attachment circuit is transparently conveyed to the other attachment circuit, where each attachment circuit is an entire Ethernet port
 - “Ethernet VLAN to VLAN” - all traffic ingressing each attachment circuit is transparently conveyed to the other attachment circuit, where each attachment circuit is a VLAN on an Ethernet port
 - In this service instance, the VLAN tag may be stripped on ingress and then re-imposed on egress.
 - Alternatively, the VLAN tag may be stripped on ingress and a completely different VLAN ID imposed on egress, allowing VLAN re-write
 - The VLAN ID is locally significant to the Ethernet port

PWE3 Reference Model for Ethernet VPWS



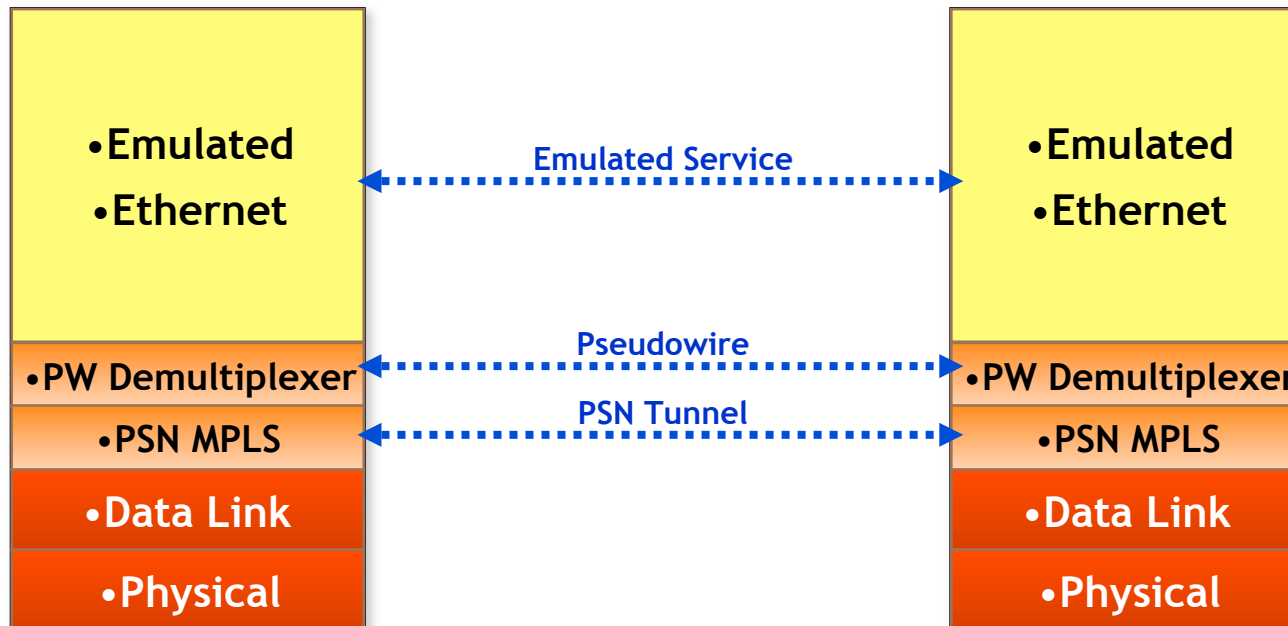
PWE3 Architectural Reference Model for Ethernet Pseudowires



Ethernet Virtual Private Wire Service



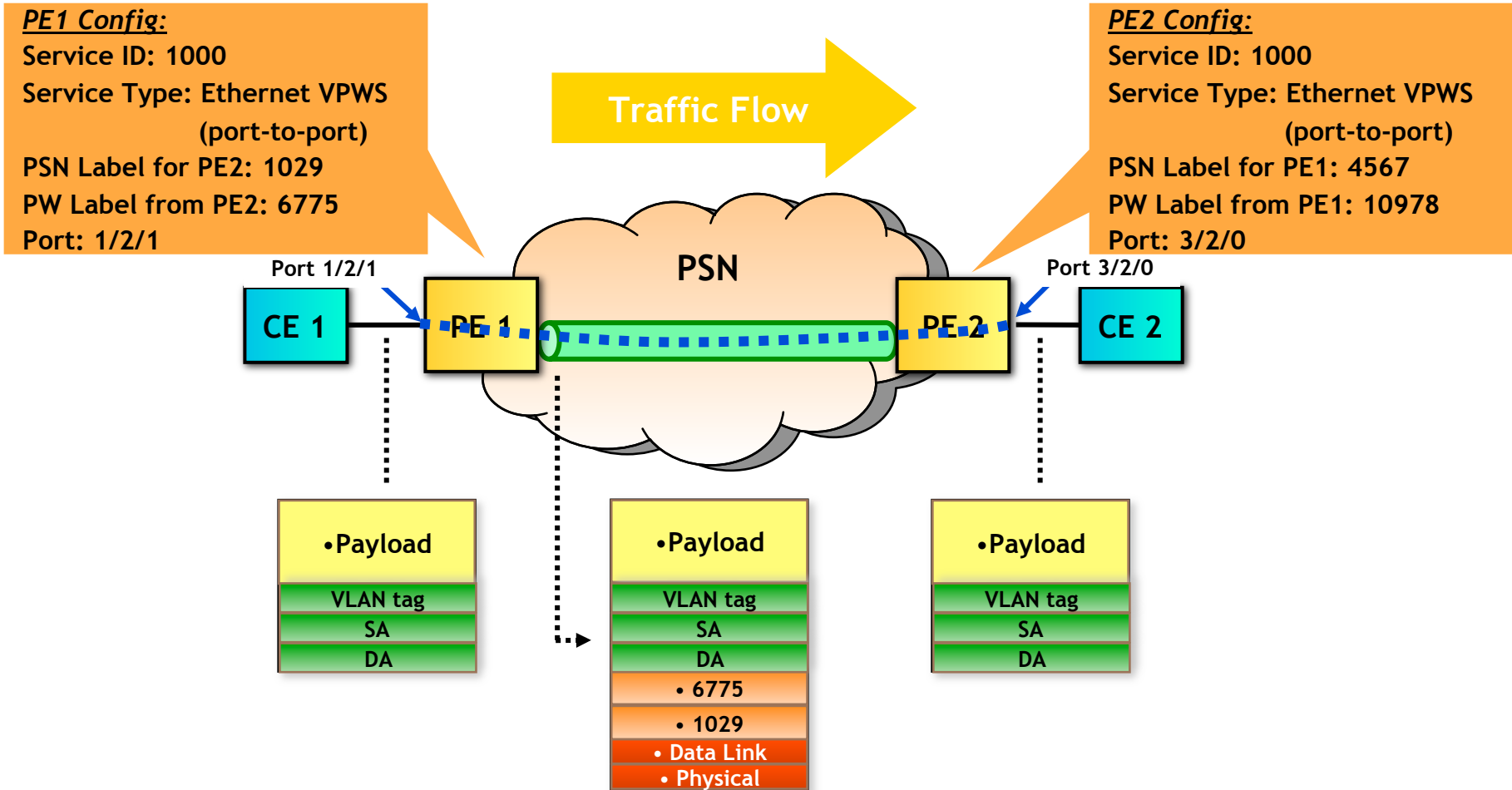
Ethernet PWE3 Protocol Stack Reference Model:



Ethernet VPWS Example 1



Example 1: Ethernet VPWS port-to-port (traffic flow from CE1 to CE2)



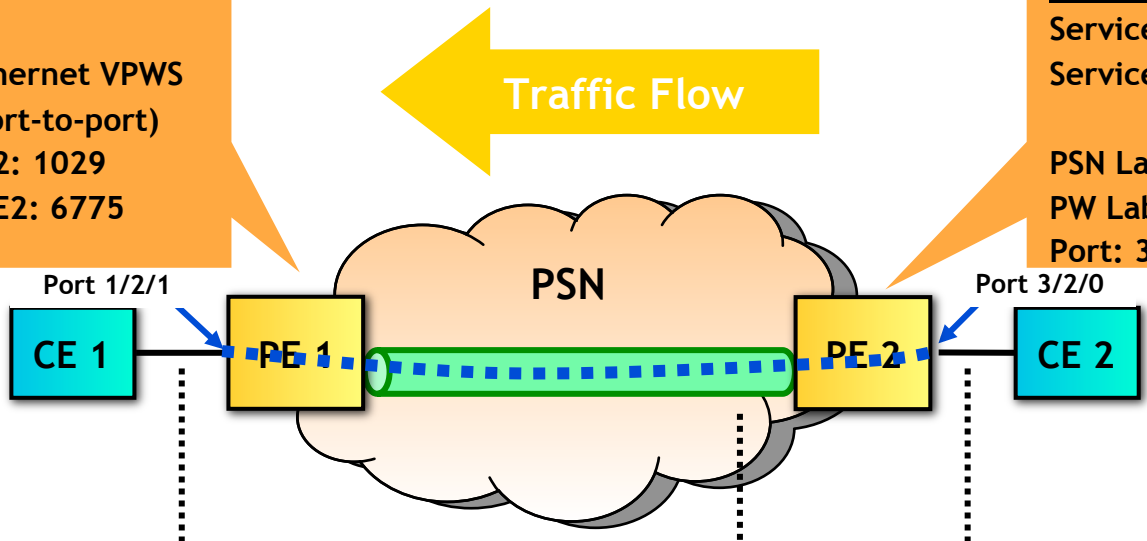
Ethernet VPWS Example 1



Example 1: Ethernet VPWS port-to-port (traffic flow from CE2 to CE1)

PE1 Config:
 Service ID: 1000
 Service Type: Ethernet VPWS
 (port-to-port)
 PSN Label for PE2: 1029
 PW Label from PE2: 6775
 Port: 1/2/1

PE2 Config:
 Service ID: 1000
 Service Type: Ethernet VPWS
 (port-to-port)
 PSN Label for PE1: 4567
 PW Label from PE1: 10978
 Port: 3/2/0



| |
|----------|
| •Payload |
| VLAN tag |
| SA |
| DA |

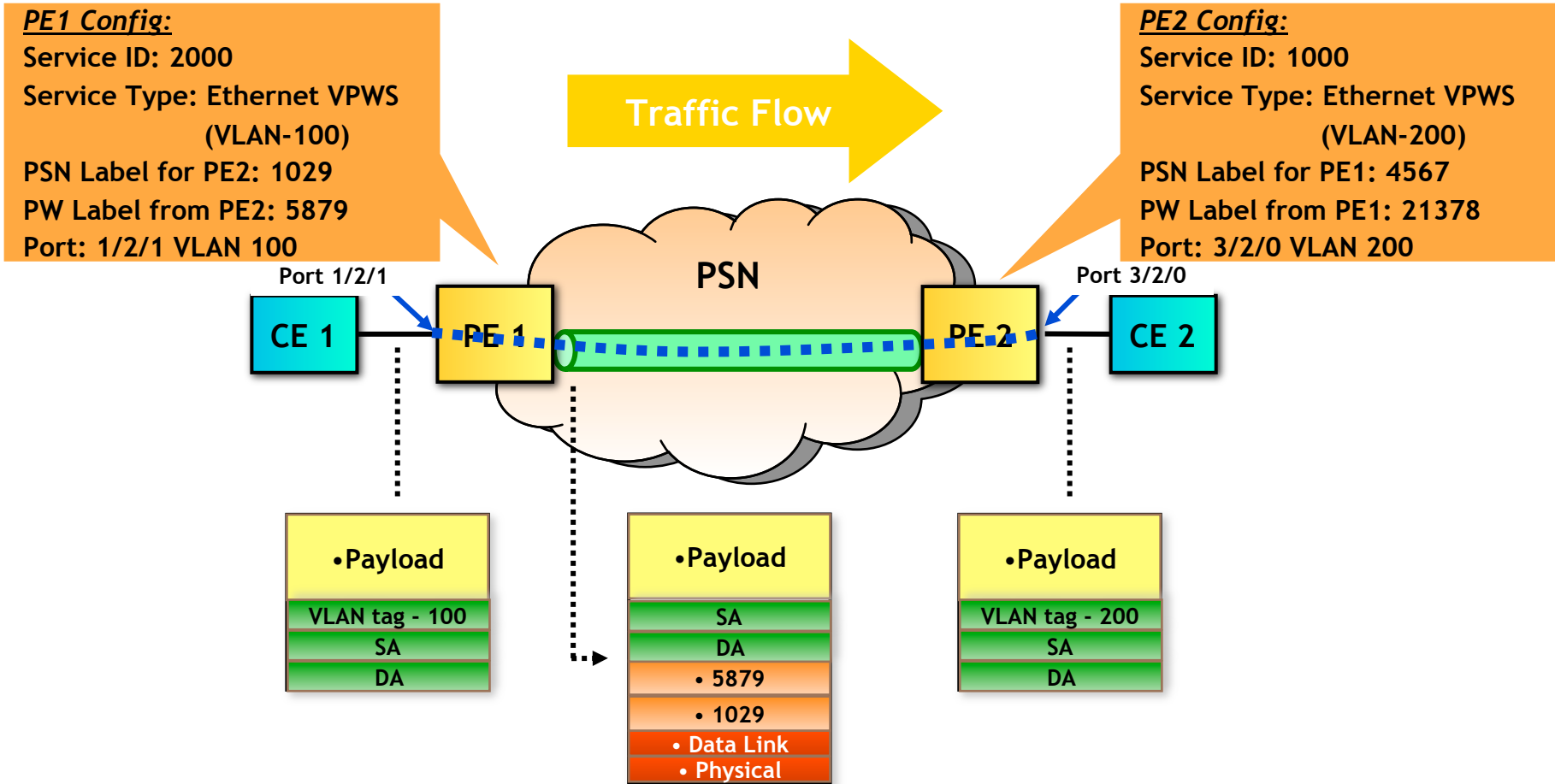
| |
|-------------|
| •Payload |
| VLAN tag |
| SA |
| DA |
| • 10978 |
| • 4567 |
| • Data Link |
| • Physical |

| |
|----------|
| •Payload |
| VLAN tag |
| SA |
| DA |

Ethernet VPWS Example 2



Example 2: Ethernet VPWS VLAN-based (traffic flow from CE1 to CE2)



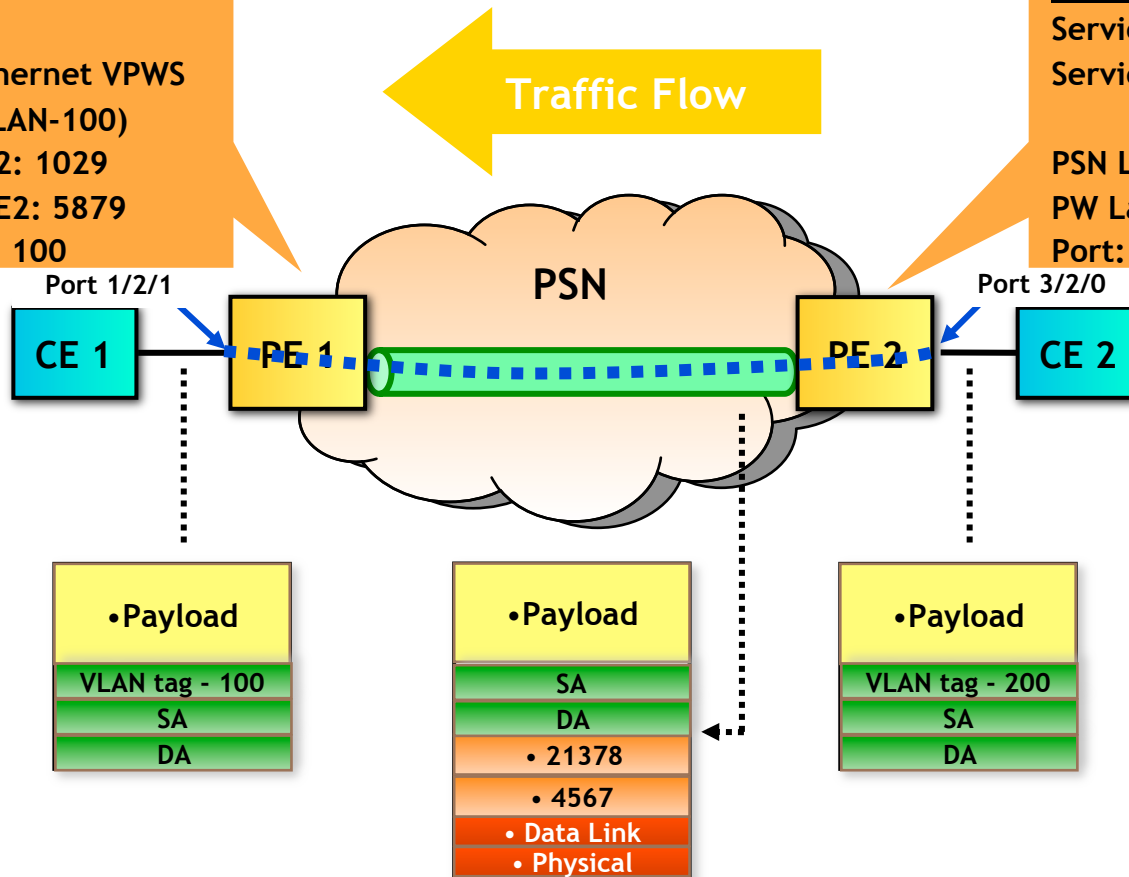
Ethernet VPWS Example 2



Example 2: Ethernet VPWS VLAN-based (traffic flow from CE2 to CE1)

PE1 Config:
 Service ID: 2000
 Service Type: Ethernet VPWS (VLAN-100)
 PSN Label for PE2: 1029
 PW Label from PE2: 5879
 Port: 1/2/1 VLAN 100

PE2 Config:
 Service ID: 1000
 Service Type: Ethernet VPWS (VLAN-200)
 PSN Label for PE1: 4567
 PW Label from PE1: 21378
 Port: 3/2/0 VLAN 200



Ethernet Pseudowires - Setup and Maintenance:

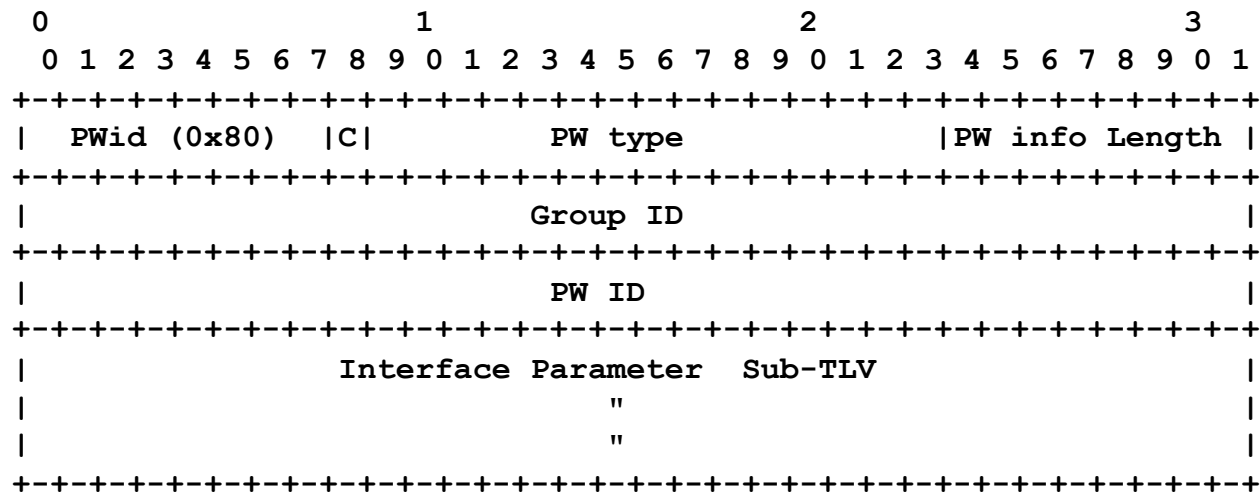
- Signalling specified in RFC4447 - “Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)”
- The MPLS Label Distribution Protocol, LDP [RFC5036], is used for setting up and maintaining the pseudowires
 - PW label bindings are distributed using the LDP downstream unsolicited mode
 - PEs establish an LDP session using the LDP Extended Discovery mechanism a.k.a Targeted LDP or tLDP
- The PSN tunnels are established and maintained separately by using any of the following:
 - The Label Distribution Protocol (LDP)
 - The Resource Reservation Protocol with Traffic Engineering (RSVP-TE)
 - Static labels

Ethernet Virtual Private Wire Service



Ethernet Pseudowires - Setup and Maintenance:

- LDP distributes FEC to label mappings using the PWid FEC Element (popularly known as FEC Type 128)
- Both pseudowire endpoints have to be provisioned with the same 32-bit identifier for the pseudowire to allow them to obtain a common understanding of which service a given pseudowire belongs to.



Ethernet Pseudowires - Setup and Maintenance:

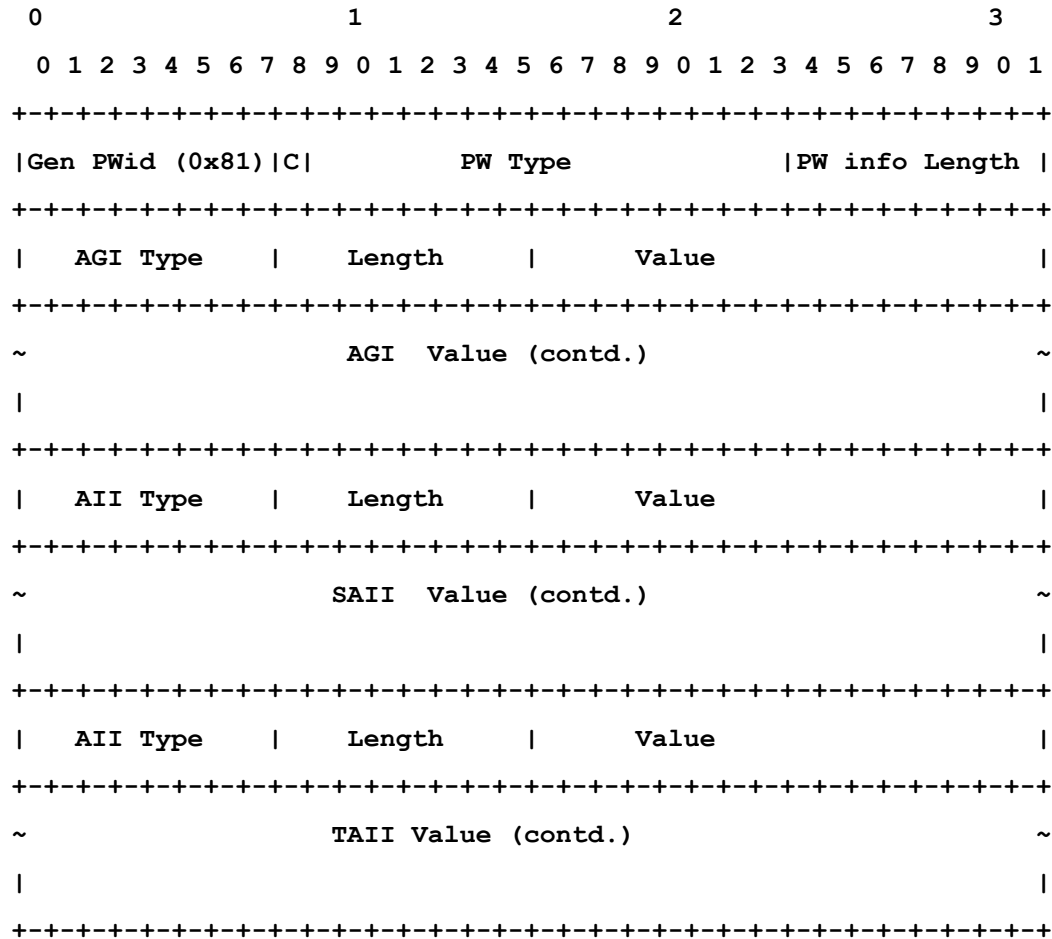
- A new TLV, the Generalized Pwid FEC Element (popularly known as FEC Type 129) has also been developed but is not widely deployed as yet
- The Generalized Pwid FEC element requires that the PW endpoints be uniquely identified; the PW itself is identified as a pair of endpoints. In addition, the endpoint identifiers are structured to support applications where the identity of the remote endpoints needs to be auto-discovered rather than statically configured.

Ethernet Virtual Private Wire Service



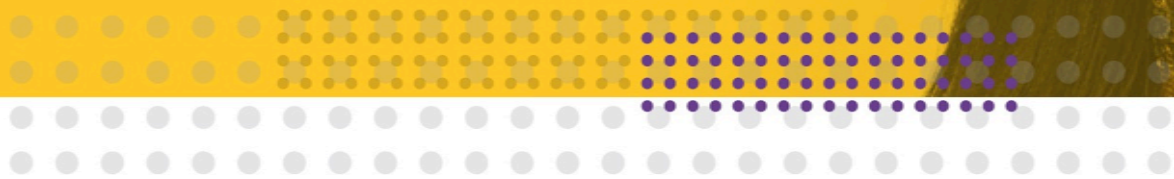
Ethernet Pseudowires - Setup and Maintenance:

- The Generalized Pwid FEC Element (popularly known as FEC Type 129)



4.4

Ethernet Virtual Private LAN Service (VPLS)



Ethernet Virtual Private LAN Service

SONOG

Ethernet VPLS:

- Two variants
 - RFC4762 - Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling. We will concentrate on this variant in the rest of this tutorial
 - RFC4761 - Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling

Ethernet Virtual Private LAN Service

SONOG

Definition:

- A VPLS creates an emulated **private** LAN segment for a given set of users.
- It creates a Layer 2 broadcast domain that is fully capable of learning and forwarding on Ethernet MAC addresses and that is closed to a given set of users. Multiple VPLS services can be supported from a single Provider Edge (PE) node.
- The primary motivation behind VPLS is to provide connectivity between geographically dispersed customer sites across MANs and WANs, as if they were connected using a LAN.
- The main intended application for the end-user can be divided into the following two categories:
 - Connectivity between customer routers: LAN routing application
 - Connectivity between customer Ethernet switches: LAN switching application

VPLS Benefits

SONOG

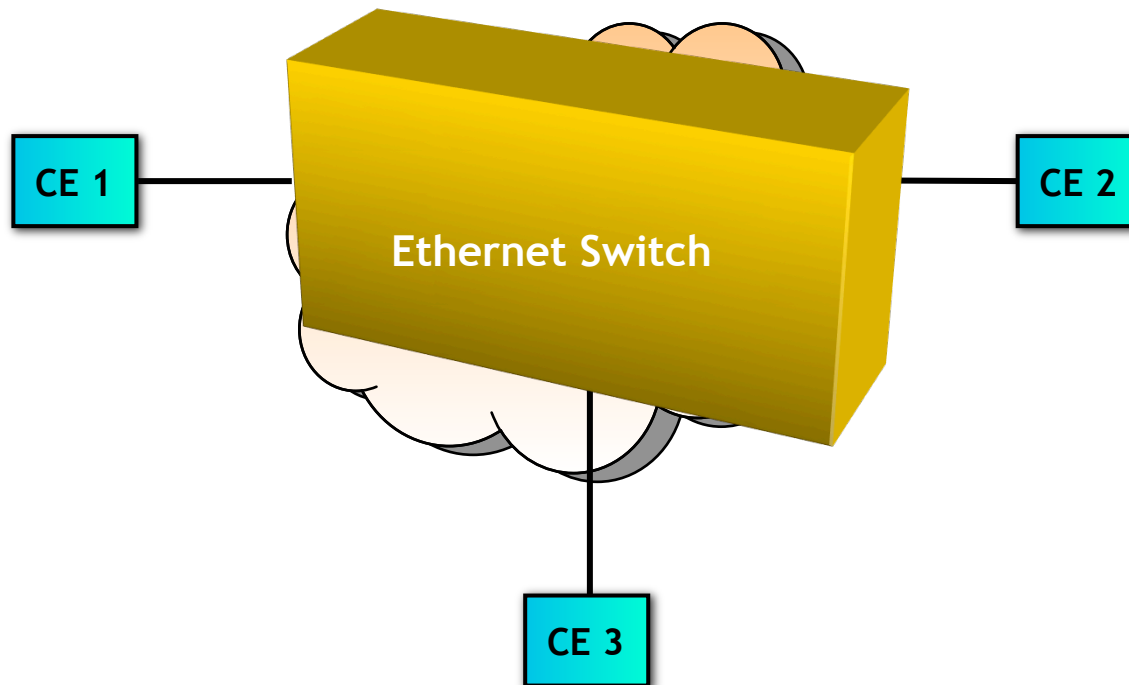
Benefits for the customer:

- **Simplicity**
 - Behaves like an “ethernet switch in the sky”
 - No routing interaction with the provider
 - Clear demarcation between subscriber and provider
 - Layer 3 agnostic
- **Scalable**
 - Provider configures site connectivity only
 - Hierarchy reduces number of sites touched
- **Multi-site connectivity**
 - On the fly connectivity via Ethernet bridging

VPLS Topological Model



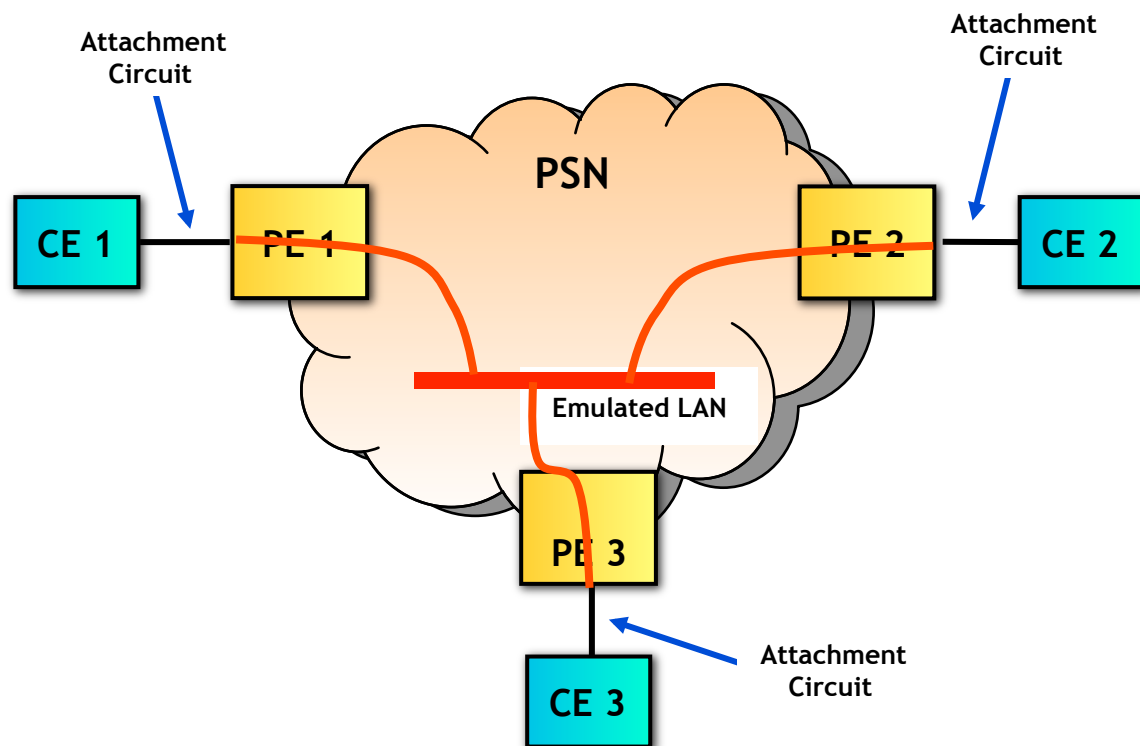
Topological Model for VPLS (customer view)



VPLS Topological Model



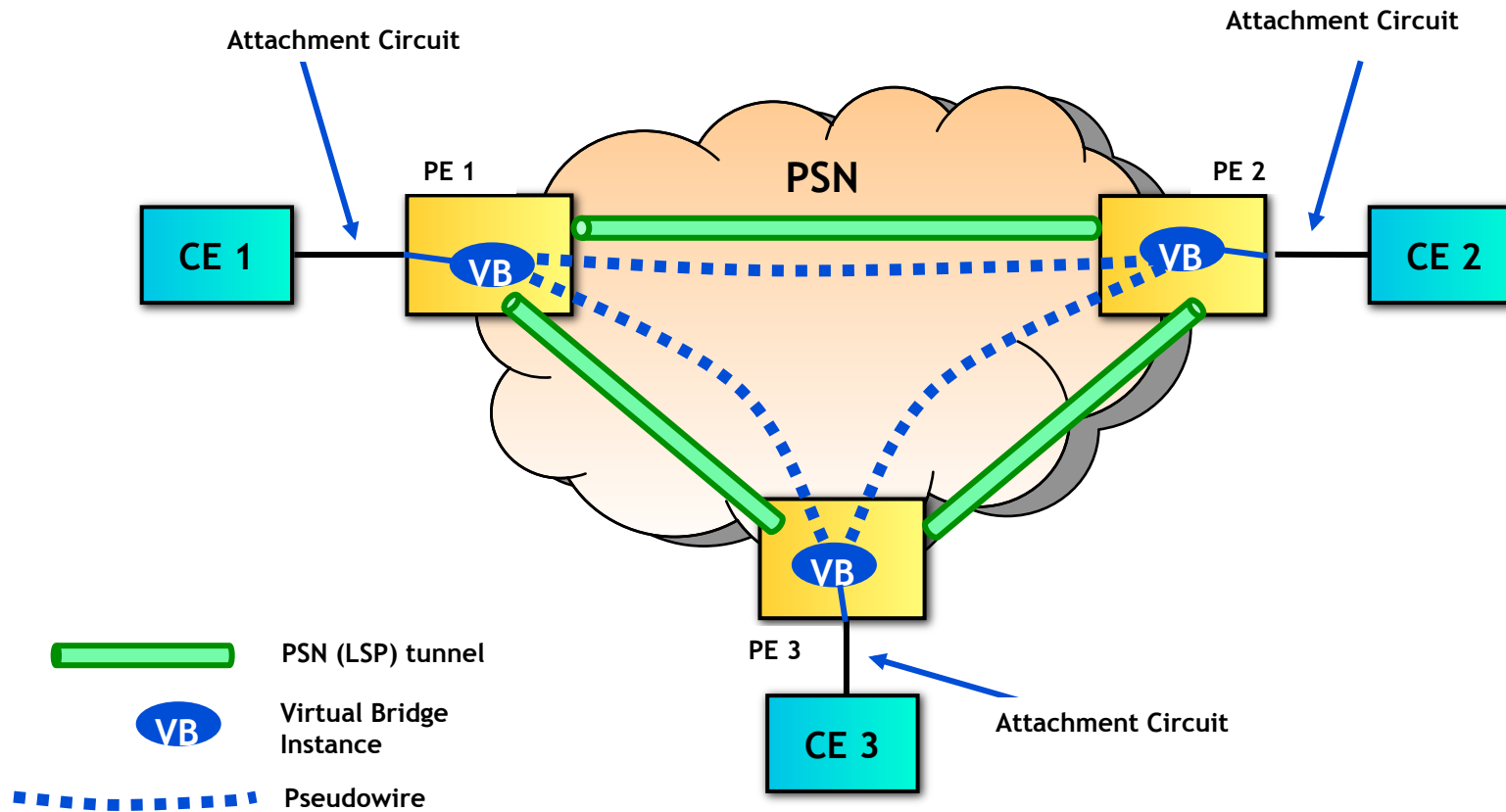
Topological Model for VPLS



Constructing VPLS Services



PSN Tunnels and Pseudowire Constructs for VPLS:



Provider Edge Functions:

- PE interfaces participating in a VPLS instance are able to flood, forward, and filter Ethernet frames, like a standard Ethernet bridged port
- Many forms of Attachment Circuits are acceptable, as long as they carry Ethernet frames:
 - Physical Ethernet ports
 - Logical (tagged) Ethernet ports
 - ATM PVCs carrying Ethernet frames
 - Ethernet Pseudowire
- Frames sent to broadcast addresses and to unknown destination MAC addresses are flooded to all ports:
 - Attachment Circuits
 - Pseudowires to all other PE nodes participating in the VPLS service
- PEs have the capability to associate MAC addresses with Pseudowires

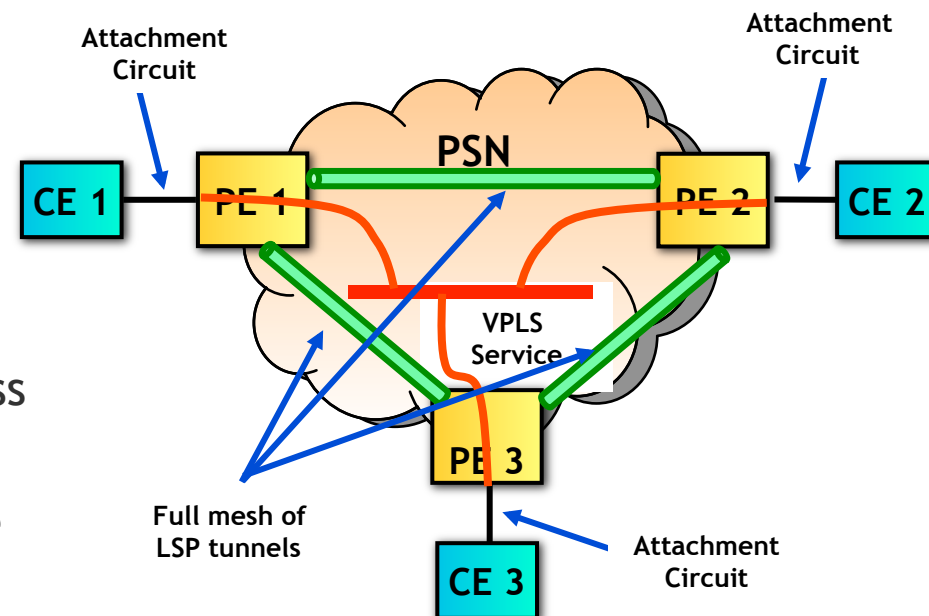
Provider Edge Functions (continued):

- Address learning:
 - Unlike BGP VPNs [RFC4364], reachability information is not advertised and distributed via a control plane.
 - Reachability is obtained by standard learning bridge functions in the data plane.
 - When a packet arrives on a PW, if the source MAC address is unknown, it is associated with the PW, so that outbound packets to that MAC address can be delivered over the associated PW.
 - When a packet arrives on an AC, if the source MAC address is unknown, it is associated with the AC, so that outbound packets to that MAC address can be delivered over the associated AC.

VPLS Signalling

VPLS Mechanics:

- Bridging capable PE routers are connected with a full mesh of MPLS LSP tunnels
- Per-Service VC labels are negotiated using RFC 4447 techniques
- Replicates unknown/broadcast traffic in a service domain
- MAC learning over tunnel & access ports
- Separate FIB per VPLS for private communication



VPLS Signalling

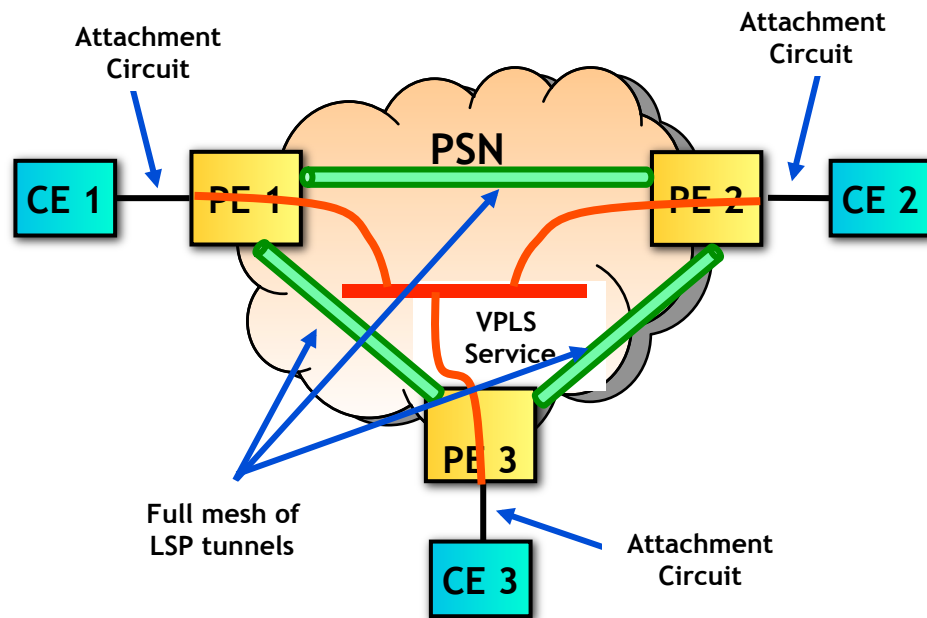


Tunnel establishment

- LDP:
 - MPLS paths based on IGP reachability
 - RSVP: traffic engineered MPLS paths with bandwidth & link constraints, and fast reroute alternatives

Pseudowire establishment

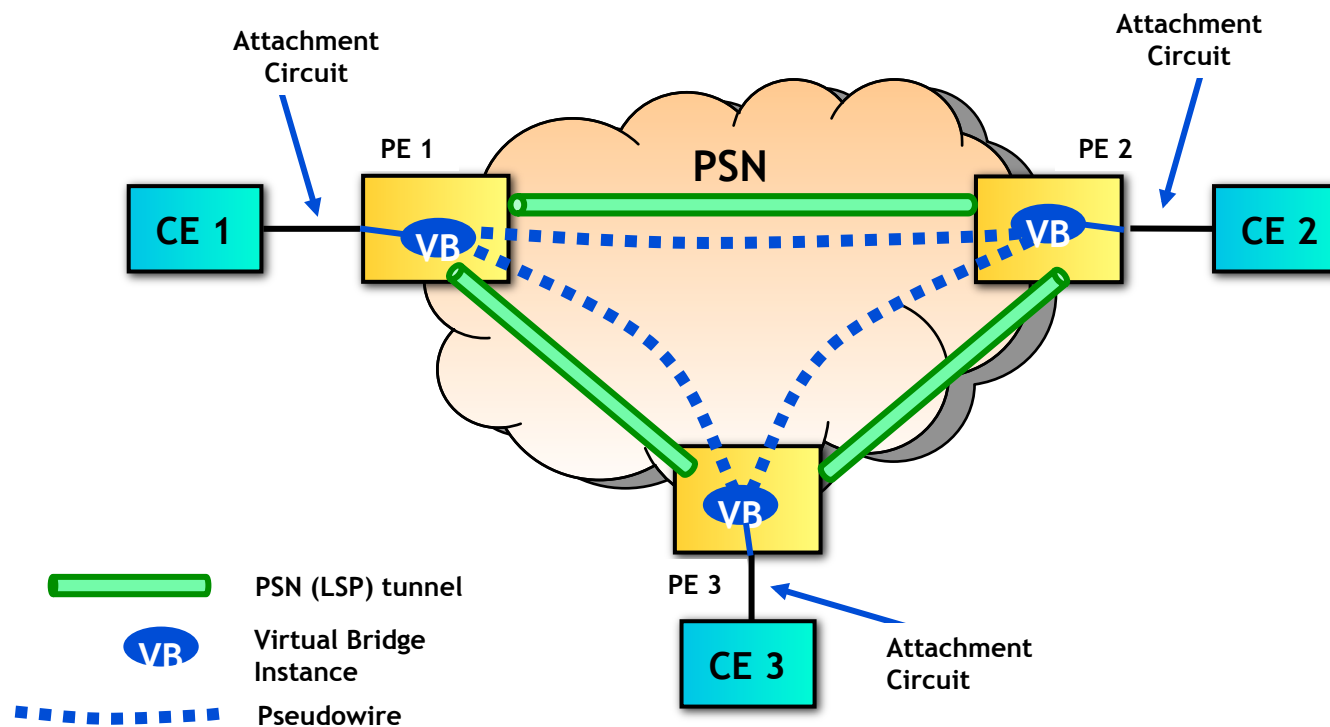
- LDP: point-to-point exchange of PW ID, labels, MTU



VPLS Signalling

A full mesh of pseudowires is established between all PEs participating in the VPLS service:

- Each PE initiates a targeted LDP session to the far-end System IP (loopback) address
- Tells far-end what PW label to use when sending packets for each service



Why a full mesh of pseudowires?

- If the topology of the VPLS is not restricted to a full mesh, then it may be that for two PEs not directly connected via PWs, they would have to use an intermediary PE to relay packets
- A loop-breaking protocol, such as the Spanning Tree Protocol, would be required
- With a full-mesh of PWs, every PE is now directly connected to every other PE in the VPLS via a PW; there is no longer any need to relay packets
- The loop-breaking rule now becomes the "split horizon" rule, whereby a PE **MUST NOT** forward traffic from one PW to another in the same VPLS mesh
 - Does this remind you of a similar mechanism used in IP networks ? The ibgp full-mesh !

Ethernet Pseudowires - Setup and Maintenance:

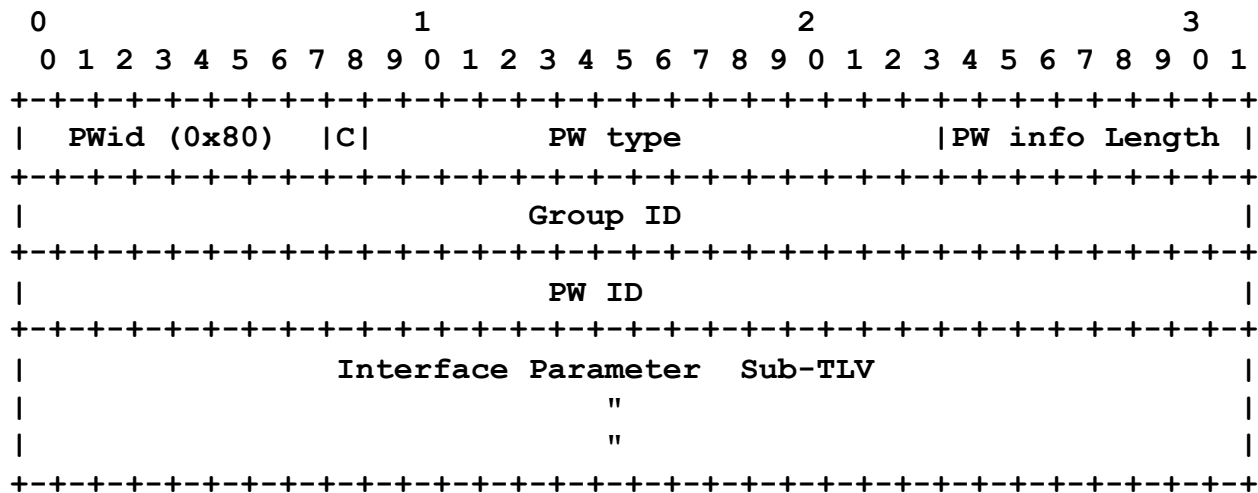
- Signalling specified in RFC4447 - “Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)”
- The MPLS Label Distribution Protocol, LDP [RFC5036], is used for setting up and maintaining the pseudowires
 - PW label bindings are distributed using the LDP downstream unsolicited mode
 - PEs establish an LDP session using the LDP Extended Discovery mechanism a.k.a Targeted LDP or tLDP
- The PSN tunnels are established and maintained separately by using any of the following:
 - The Label Distribution Protocol (LDP)
 - The Resource Reservation Protocol with Traffic Engineering (RSVP-TE)
 - Static labels

VPLS Pseudowire Signalling



Ethernet Pseudowires - Setup and Maintenance:

- LDP distributes FEC to label mappings using the PWid FEC Element (popularly known as FEC Type 128)
- Both pseudowire endpoints have to be provisioned with the same 32-bit identifier for the pseudowire to allow them to obtain a common understanding of which service a given pseudowire belongs to.



Ethernet Pseudowires - Setup and Maintenance:

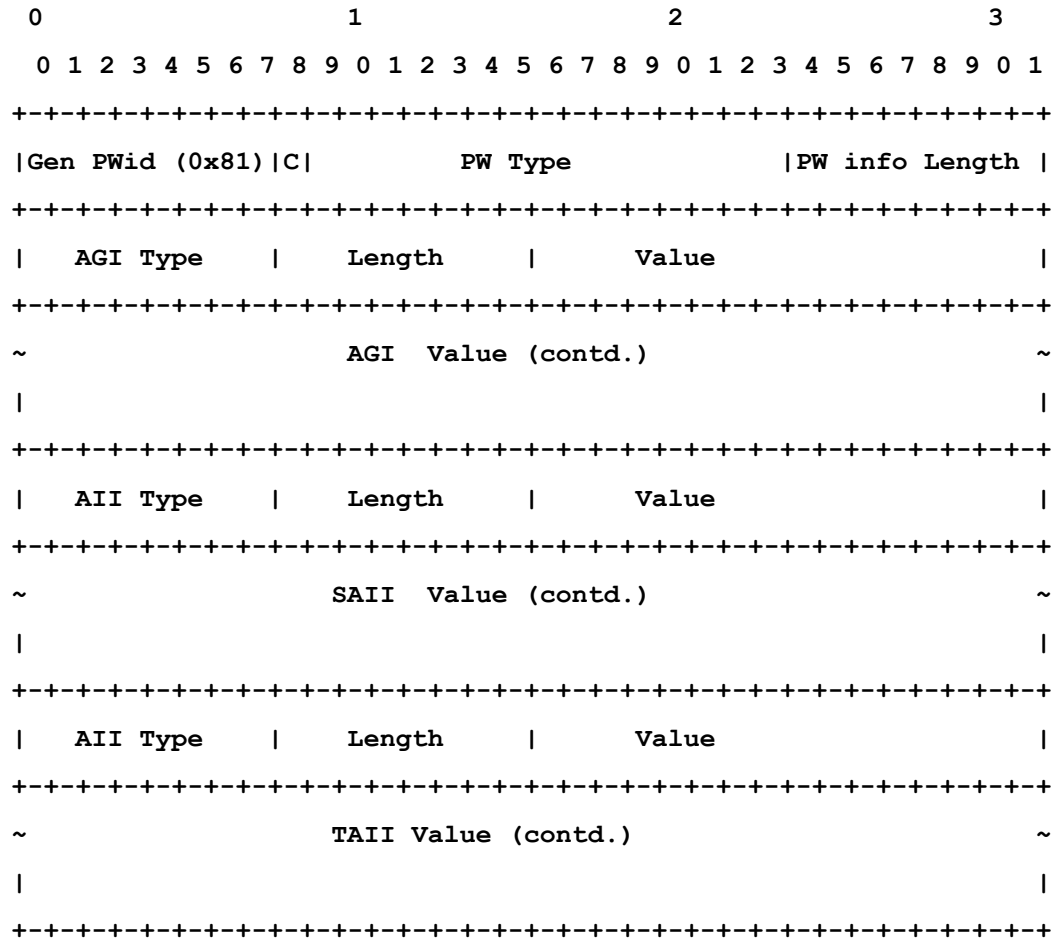
- A new TLV, the Generalized Pwid FEC Element (popularly known as FEC Type 129) has also been developed but is not widely deployed as yet
- The Generalized Pwid FEC element requires that the PW endpoints be uniquely identified; the PW itself is identified as a pair of endpoints. In addition, the endpoint identifiers are structured to support applications where the identity of the remote endpoints needs to be auto-discovered rather than statically configured.

VPLS Pseudowire Signalling



Ethernet Pseudowires - Setup and Maintenance:

- The Generalized PWid FEC Element (popularly known as FEC Type 129)

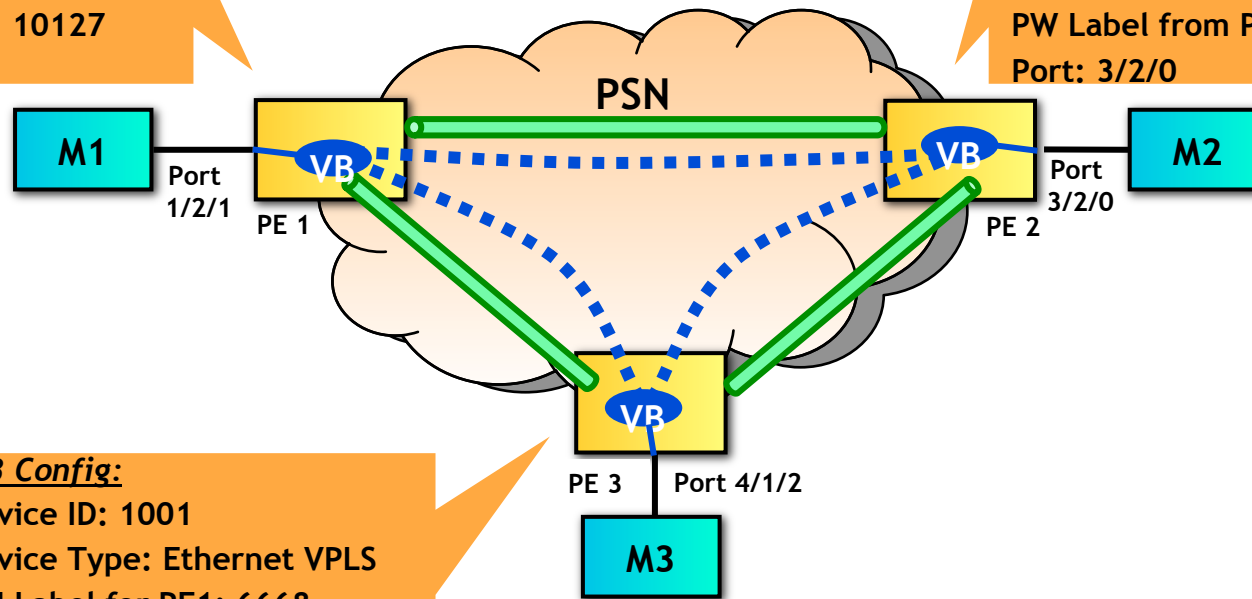


Ethernet VPLS Signalling Example



PE1 Config:
Service ID: 1001
Service Type: Ethernet VPLS
PSN Label for PE2: 1029
PSN Label for PE3: 9178
PW Label from PE2: 6775
PW Label from PE3: 10127
Port: 1/2/1

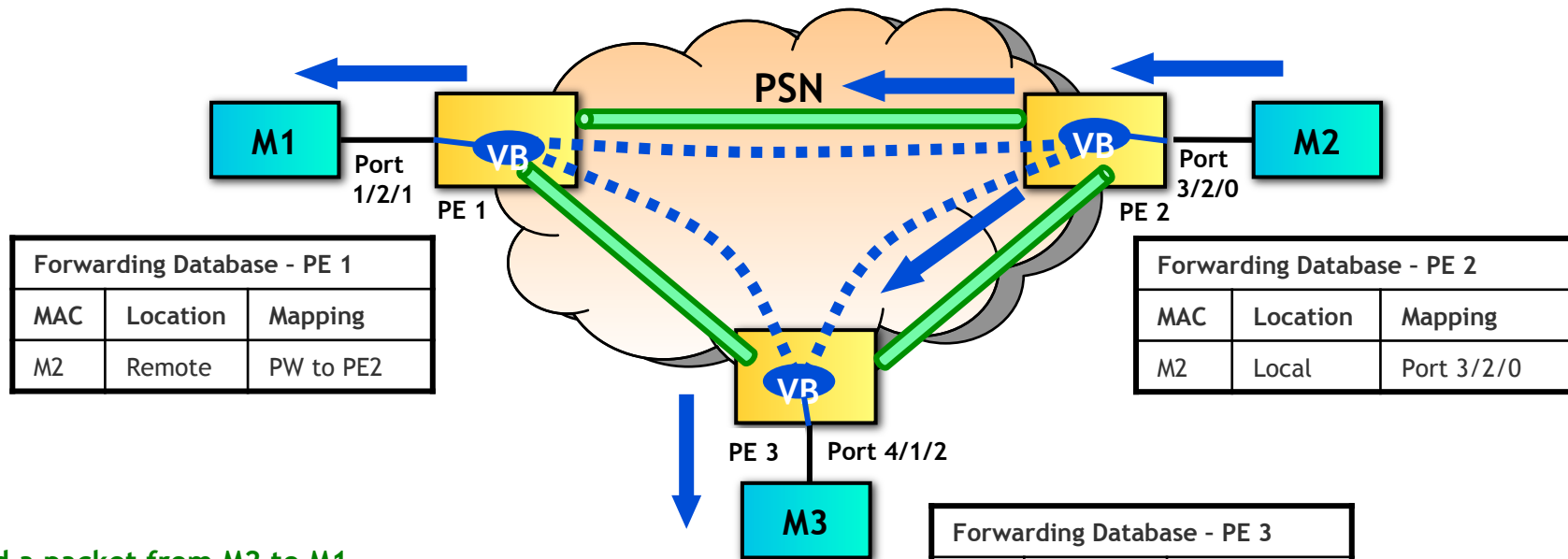
PE2 Config:
Service ID: 1001
Service Type: Ethernet VPLS
PSN Label for PE1: 4567
PSN Label for PE3: 11786
PW Label from PE1: 10978
PW Label from PE3: 4757
Port: 3/2/0



PE3 Config:
Service ID: 1001
Service Type: Ethernet VPLS
PSN Label for PE1: 6668
PSN Label for PE2: 12812
PW Label from PE1: 4568
PW Label from PE3: 10128
Port: 4/1/2

VPLS Packet Walkthrough and MAC Learning Example

Packet Walkthrough for VPLS Service-id 1001



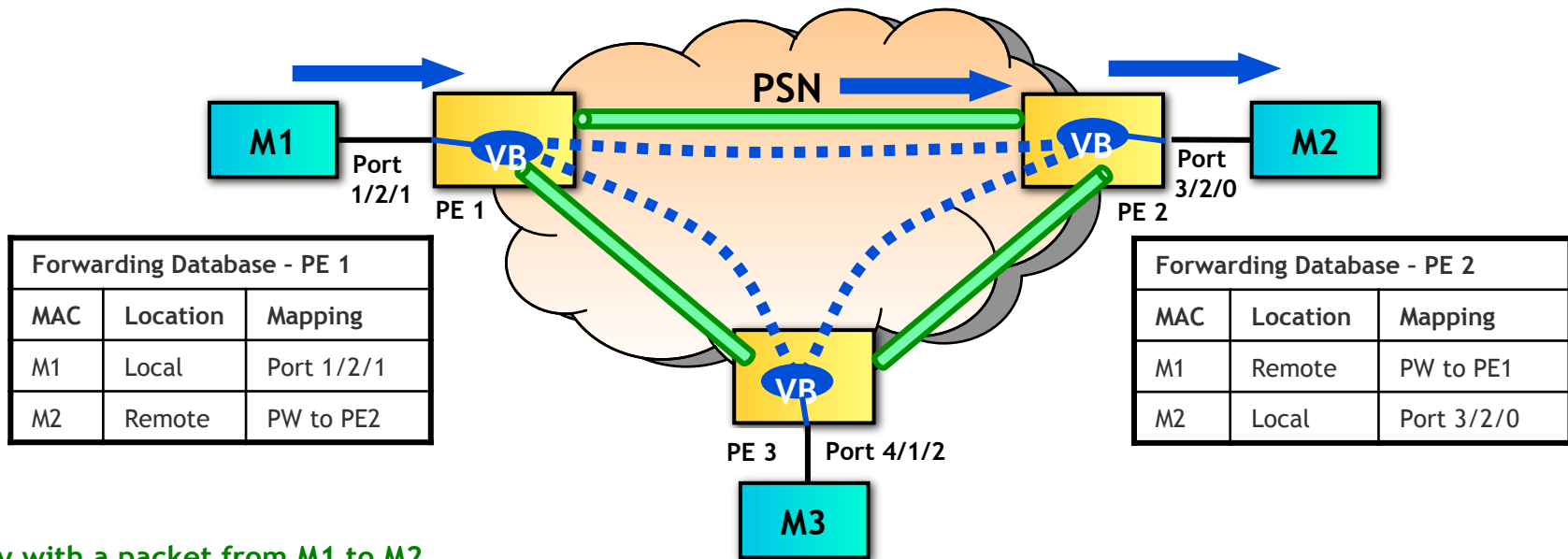
Send a packet from M2 to M1

- PE2 learns that M2 is reached on Port 3/2/0
- PE2 floods to PE1 with PW-label 10978 and PE3 with PW-label 4757
- PE1 learns from the PW-label 10978 that M2 is behind PE2
- PE1 sends on Port 1/2/1
- PE3 learns from the PW-label 4757 M2 is behind PE2
- PE3 sends on Port 4/1/2
- M1 receives packet

VPLS Packet Walkthrough and MAC Learning Example (cont.)



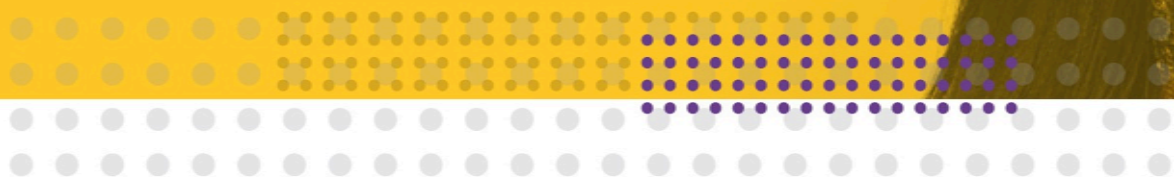
Packet Walkthrough for VPLS Service-id 1001



Reply with a packet from M1 to M2

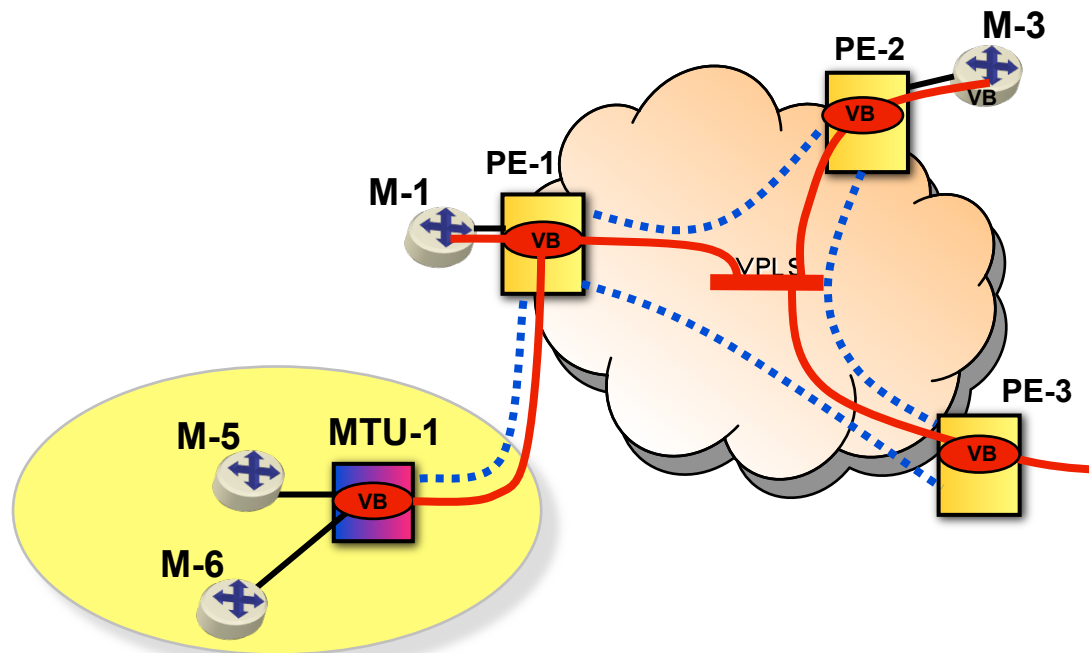
- PE1 learns M1 is on Port 1/2/1
- PE1 knows that M2 is reachable via PE2
- PE1 sends to PE2 using PW-label 6775
- PE2 knows that M2 is reachable on Port 3/2/0 and so it sends it out that port
- M2 receives packet

4.5 Scaling VPLS



Hierarchical-VPLS (H-VPLS)

- Introduces hierarchy in the base VPLS solution to provide scaling & operational advantages
- Extends the reach of a VPLS using spokes, i.e., point-to-point pseudowires or logical ports



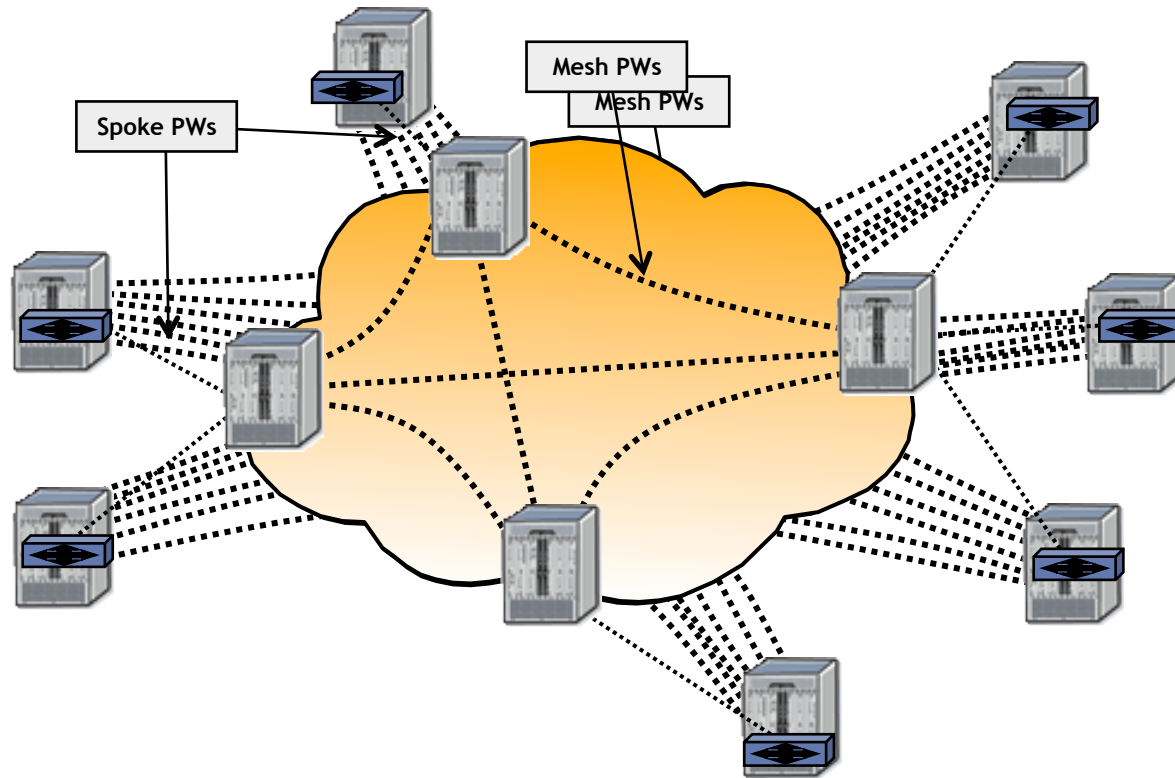
How is a spoke useful?

- Scales signalling
 - Full-mesh between MTUs is reduced to full-mesh between PEs and single PW between MTU and PE
- Scales replication
 - Replication at MTU is not required
 - Replication is reduced to what is necessary between PEs
- Simplifies edge devices
 - Keeps cost down because PEs can be replaced with MTUs
- Enables scalable inter-domain VPLS
 - Single spoke to interconnect domains

Scalability: Signalling



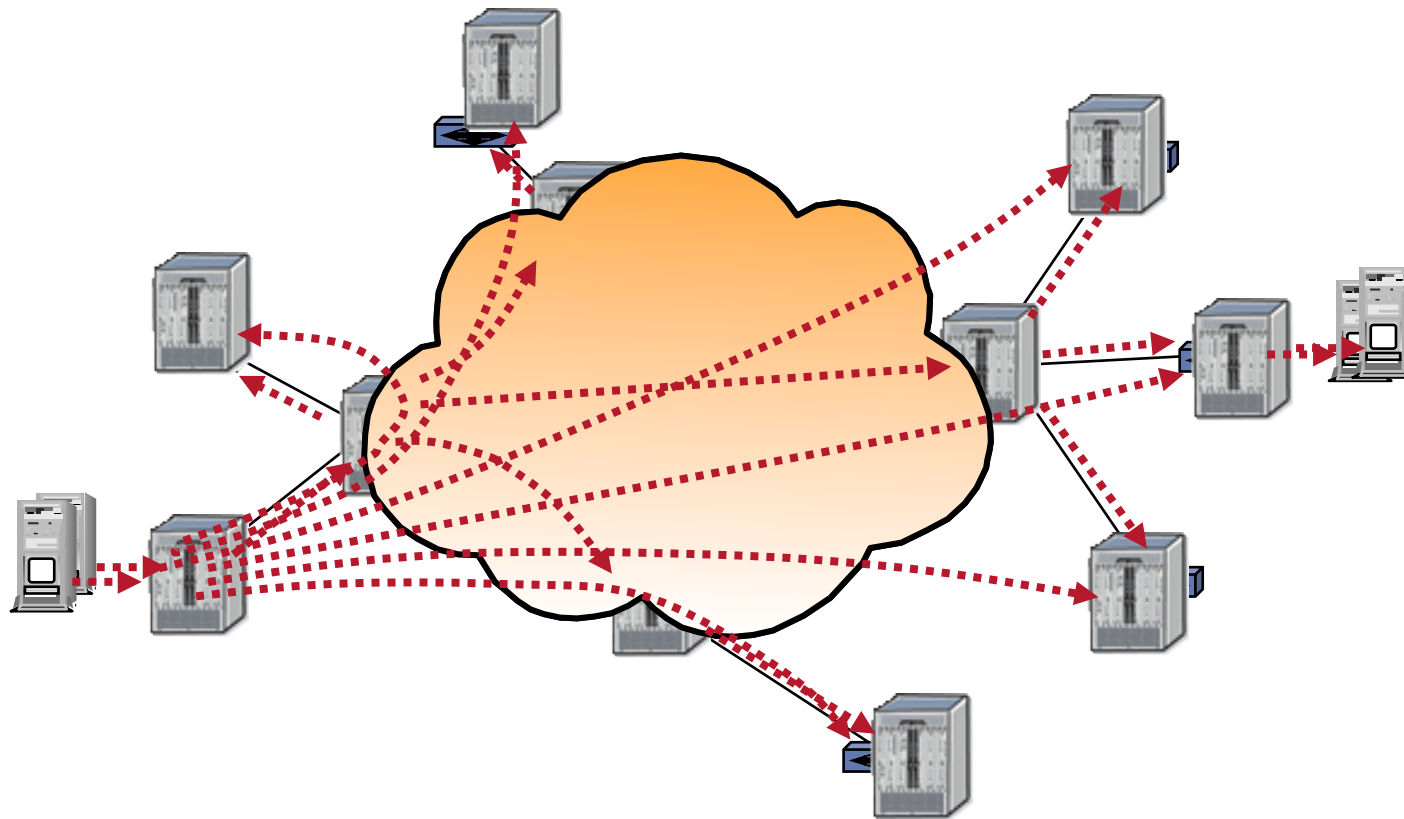
Full-mesh between PEs is reduced to full-mesh between PEs and single spoke between MTU and PE



Scalability: Replication



Flat architecture replication is reduced to distributed replication

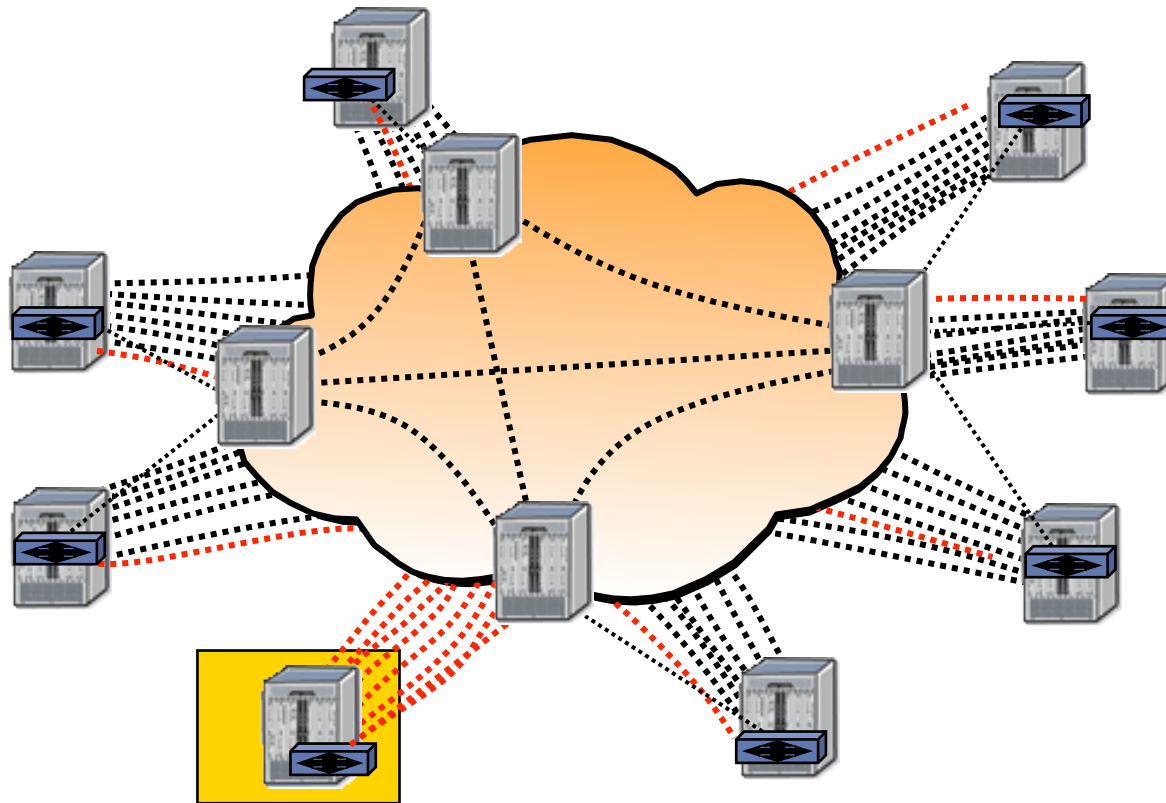


Scalability: Configuration

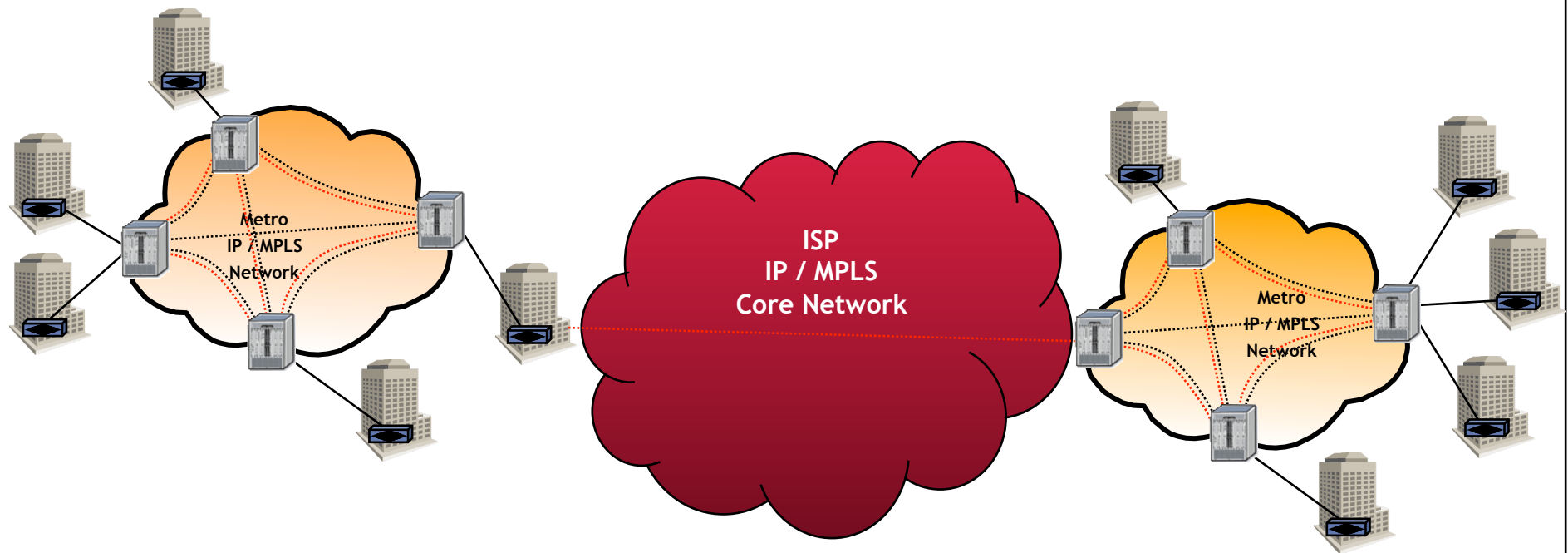


Full mesh configuration

is significantly reduced



Topological Extensibility: Metro Interconnect

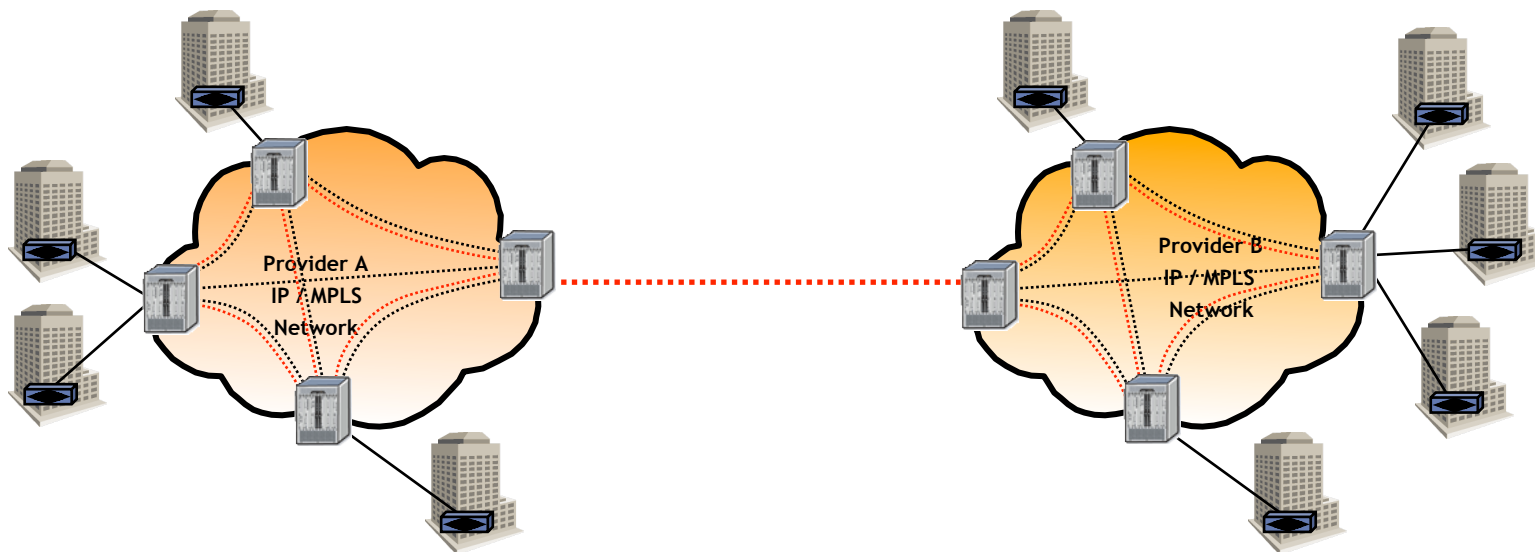


Topological Extensibility: Inter-AS Connectivity

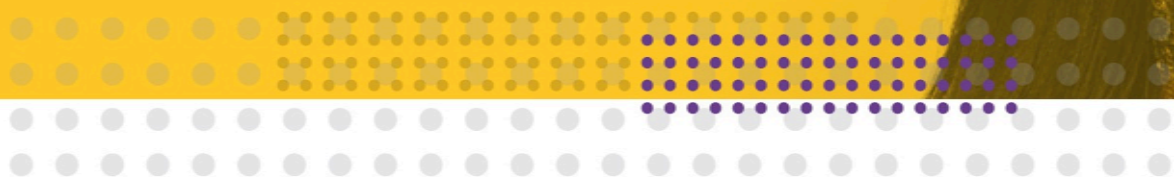
SONOG

Provider hand-off can be

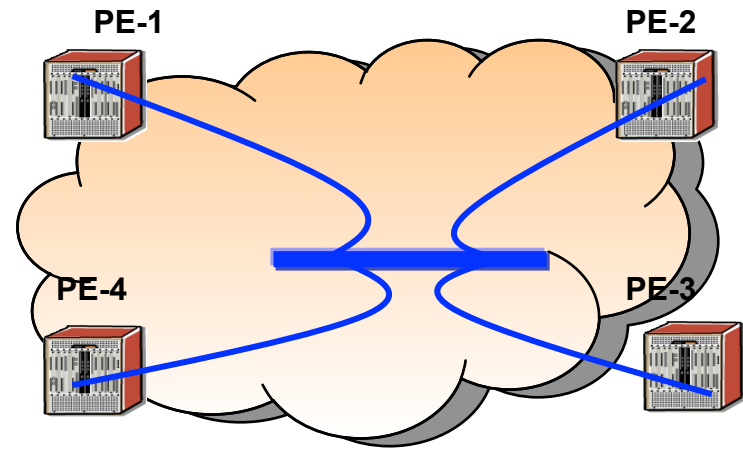
- q-tagged or q-in-q port
- Pseudowire spoke



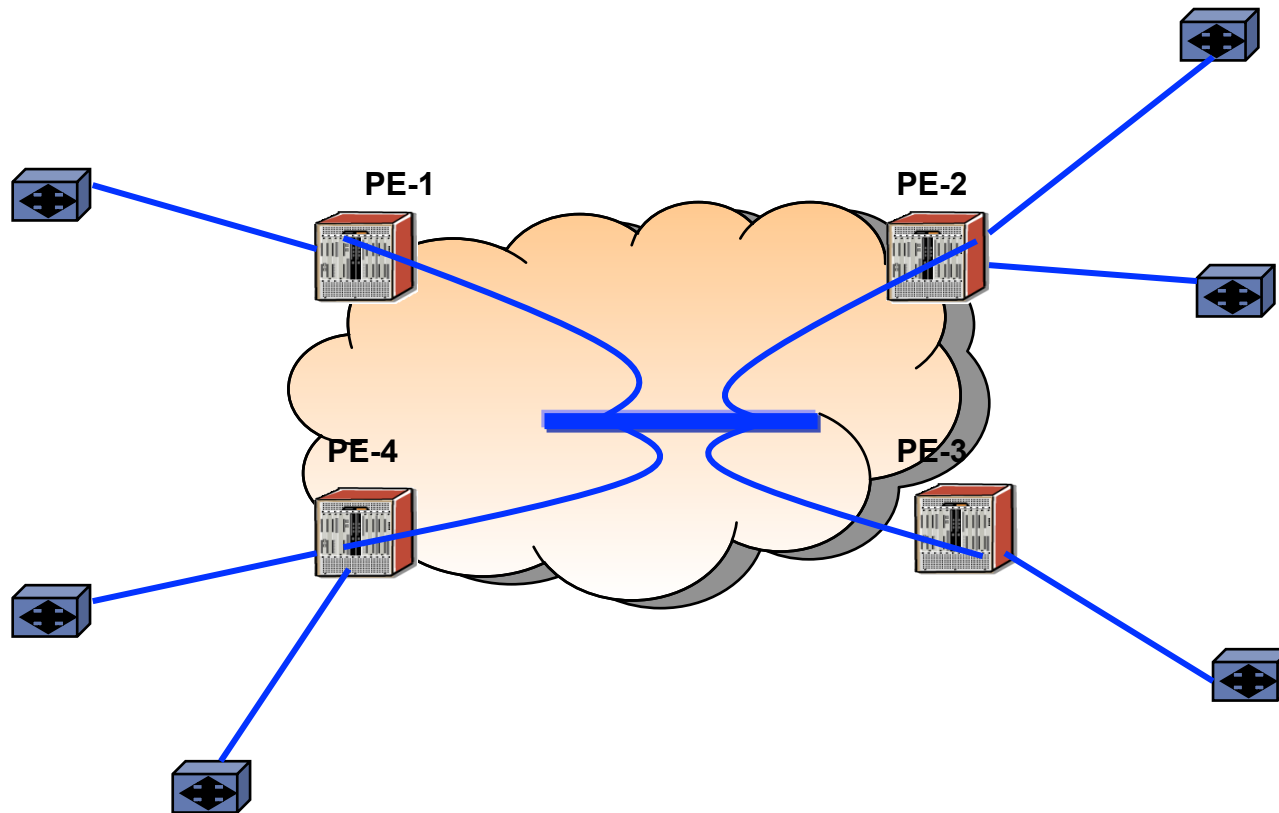
4.6 VPLS Topologies



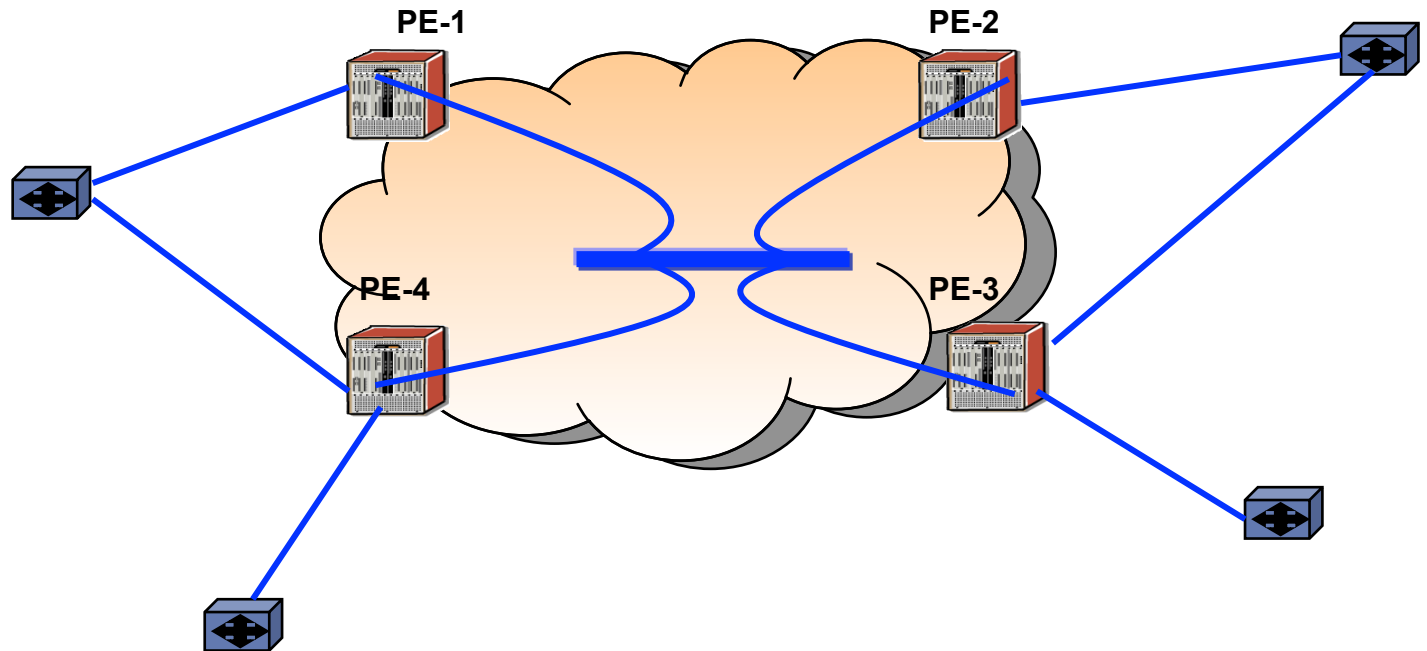
Topologies: Mesh



Topologies: Hierarchical



Topologies: Dual-homing



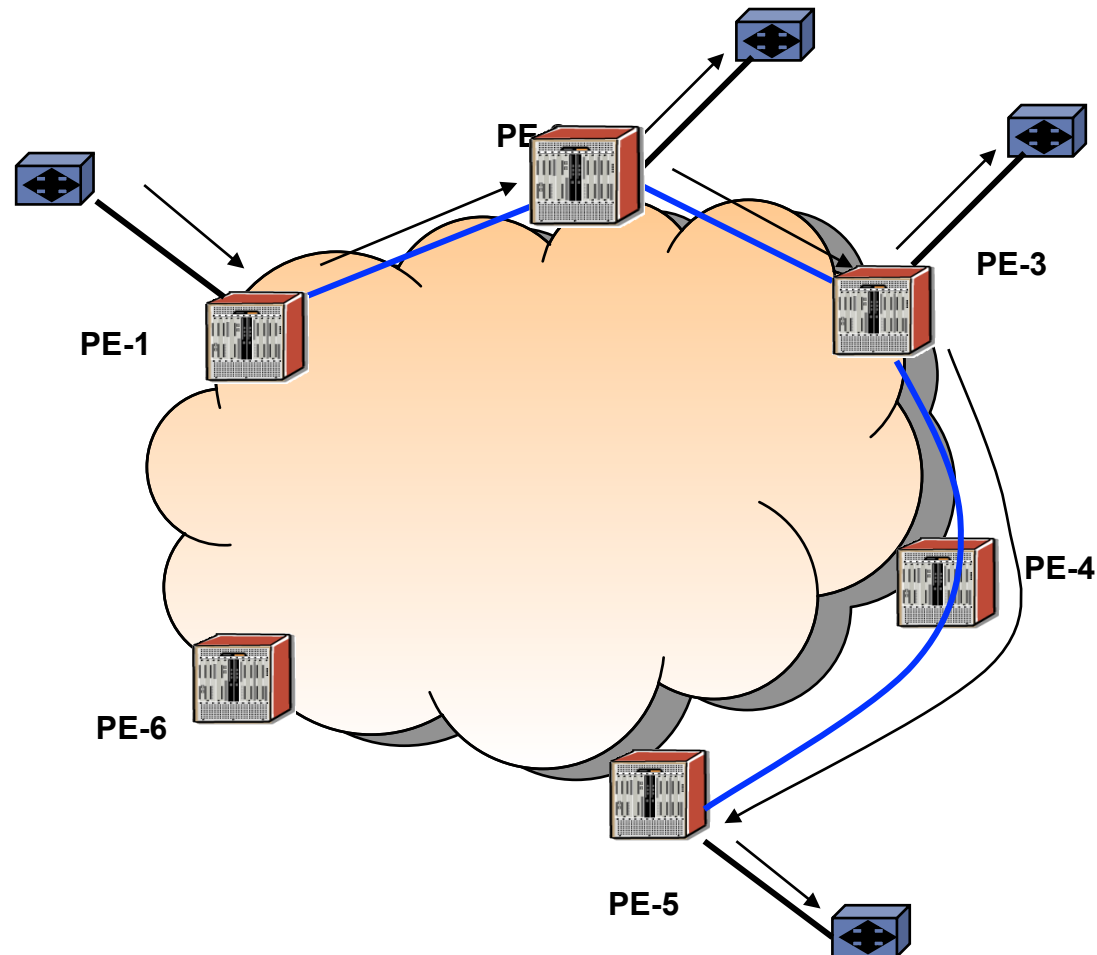
Topologies: Ring

A full mesh would have too many duplicate packets

Each PE has a spoke to the next PE in the VPLS

Packets are flooded into the adjacent spokes and to all VPLS ports

When MACs are learned, packets stop at the owning PE



5

Summary

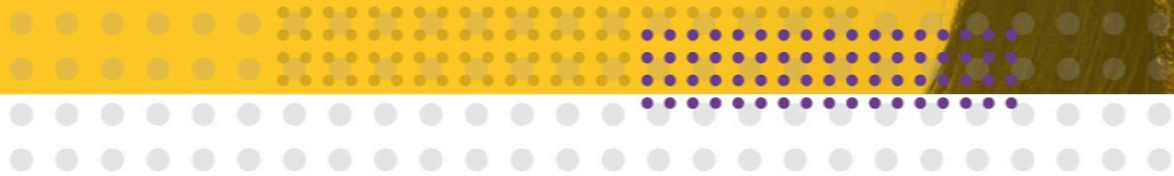


Summary

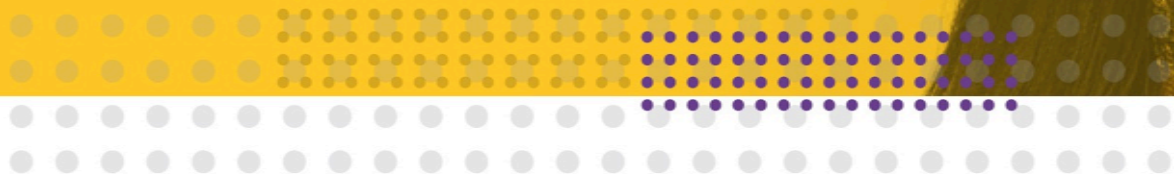
- Ethernet Services are in a period of tremendous growth with great revenue potential for service providers
- The Metro Ethernet Forum has standardised Ethernet services and continues to enhance specifications
- Traditional forms of Ethernet delivery are no longer suitable for the delivery of “carrier-grade” Ethernet services
- MPLS provides a proven platform for the delivery of scalable, flexible, feature-rich Ethernet services using the same infrastructure used to deliver other MPLS-based services

6

Questions ???



Thank You



www.alcatel-lucent.com

