Carrier Ethernet

Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides and audio/video recordings of this class lecture are at:

http://www.cse.wustl.edu/~jain/cse570-13/

Overview

1. Technologies for Data Center Interconnection:
   PDH, SDH, MPLS, PWE3
2. Metro Ethernet Services: E-Line, E-LAN, E-Tree, E-Access
3. Administration and Management Ethernet Services
4. Extensions to Ethernet for carriers:
   PB (Q-in-Q), PBB (MAC-in-MAC), PBB-TE

Note: Although these technologies were originally developed for carriers, they are now used inside multi-tenant data centers (clouds)

Options to Connect Two Data Centers?

Danforth Campus

Medical Campus

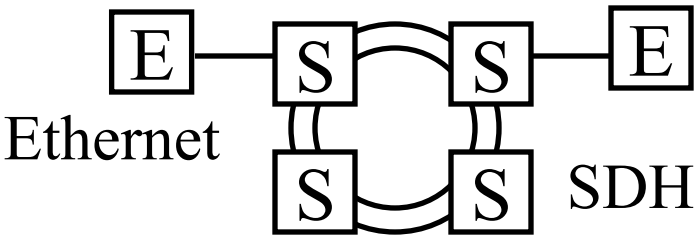
1. Dedicated Optical fiber (leased from the phone company)
2. Ethernet over Optical Transport Network (all-Optical Switches)
3. Ethernet over Wavelength Division Multiplexing (DWDM)
4. Ethernet over Synchronous Digital Hierarchy (SDH)
5. Ethernet over Plesiochronous Hierarchy (PDH)
6. Ethernet over Pseudo-wire over MPLS
7. Ethernet over Micro-wave
8. Single Pair High-Speed Digital Subscriber Line (SHDSL)
9. Ethernet with enhancements

Plesiochronous Digital Hierarchy (PDH)

- Plesios + Synchronous = Near synchronous
- Phone Line = 64 kbps = 1 User channel
- North America
  - T1 = 1.544 Mbps = 24 User channels
  - T2 = 6.312 Mbps = 96 Channels
  - T3 = 44.736 Mbps = 480 Channels
- Europe:
  - E1 = 2.048 Mbps = 32 Channels
  - E2 = 8.448 Mbps = 128 Channels
  - E3 = 139.264 Mbps = 2048 Channels

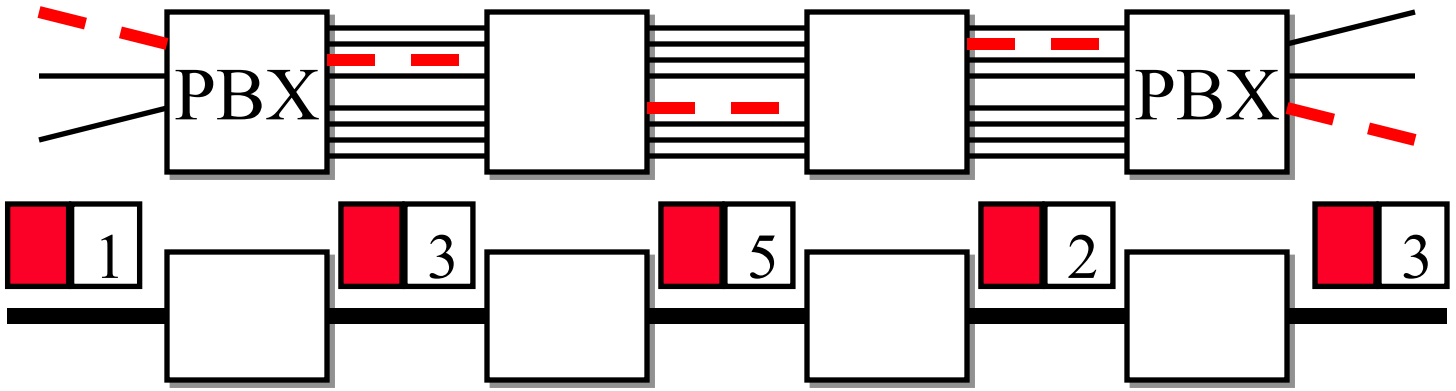
T1

SONET/SDH

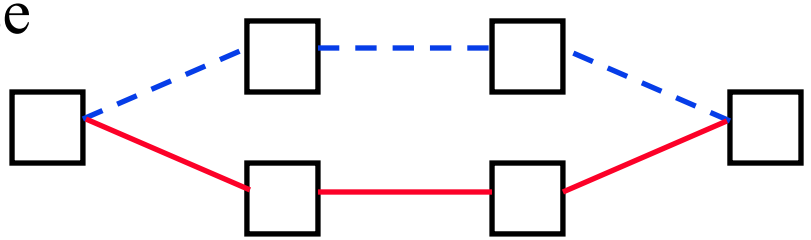


- SONET=Synchronous optical network
- Standard for digital optical transmission
- Standardized by ANSI and then by ITU
  ⇒ Synchronous Digital Hierarchy (SDH)
- Protection: Allows redundant Line or paths
- Fast Restoration: 50ms using rings
- Sophisticated management
- Ideal for Voice: No queues. Guaranteed delay
- Fixed Payload Rates: OC1=51.84 Mbps, OC3=155M, OC12=622M, OC48=2.4G, OC192=9.5G
  Rates do not match data rates of 10M, 100M, 1G, 10G
- Static rates not suitable for bursty traffic
- One Payload per Stream ⇒ High Cost

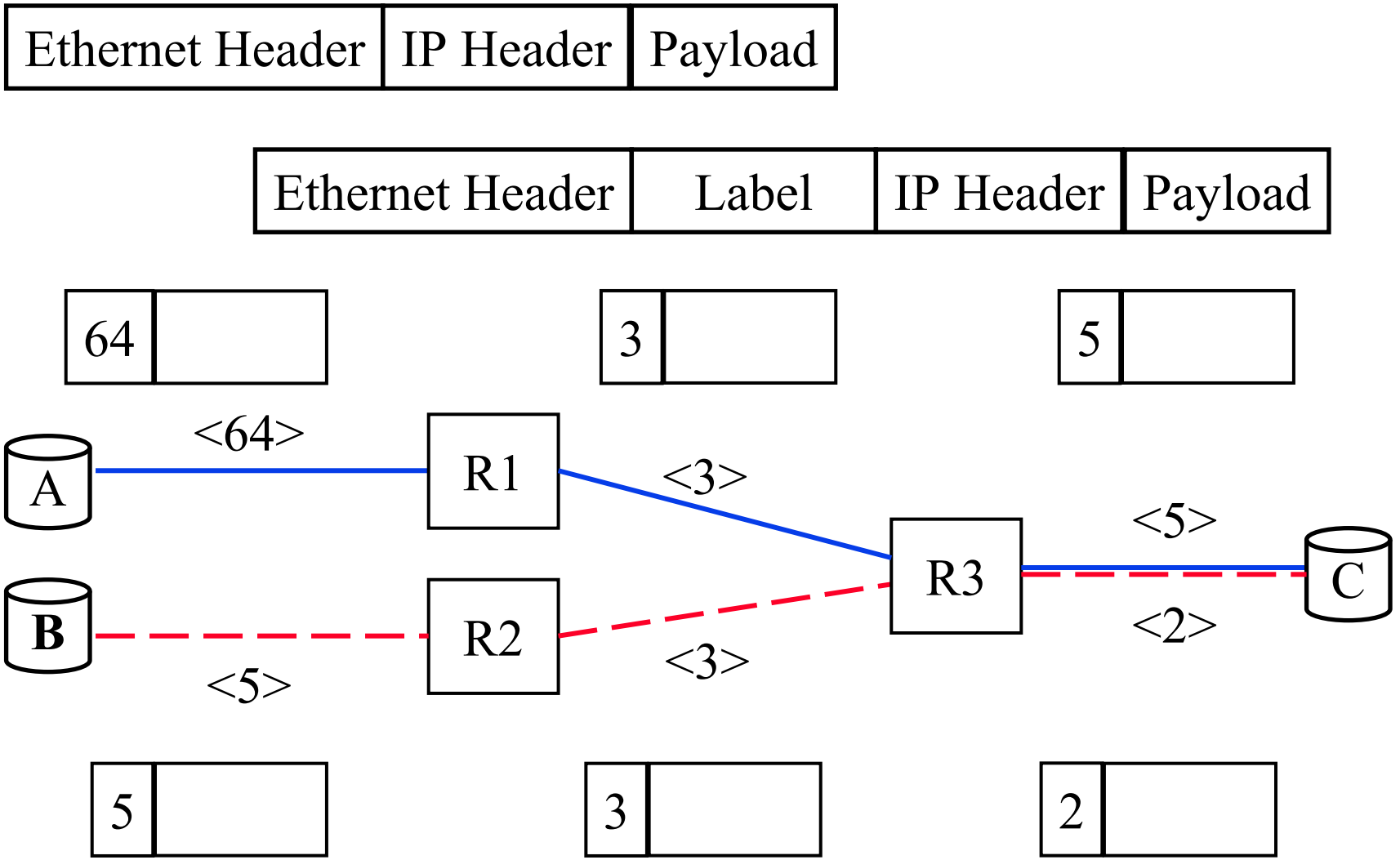
Multiprotocol Label Switching (MPLS)



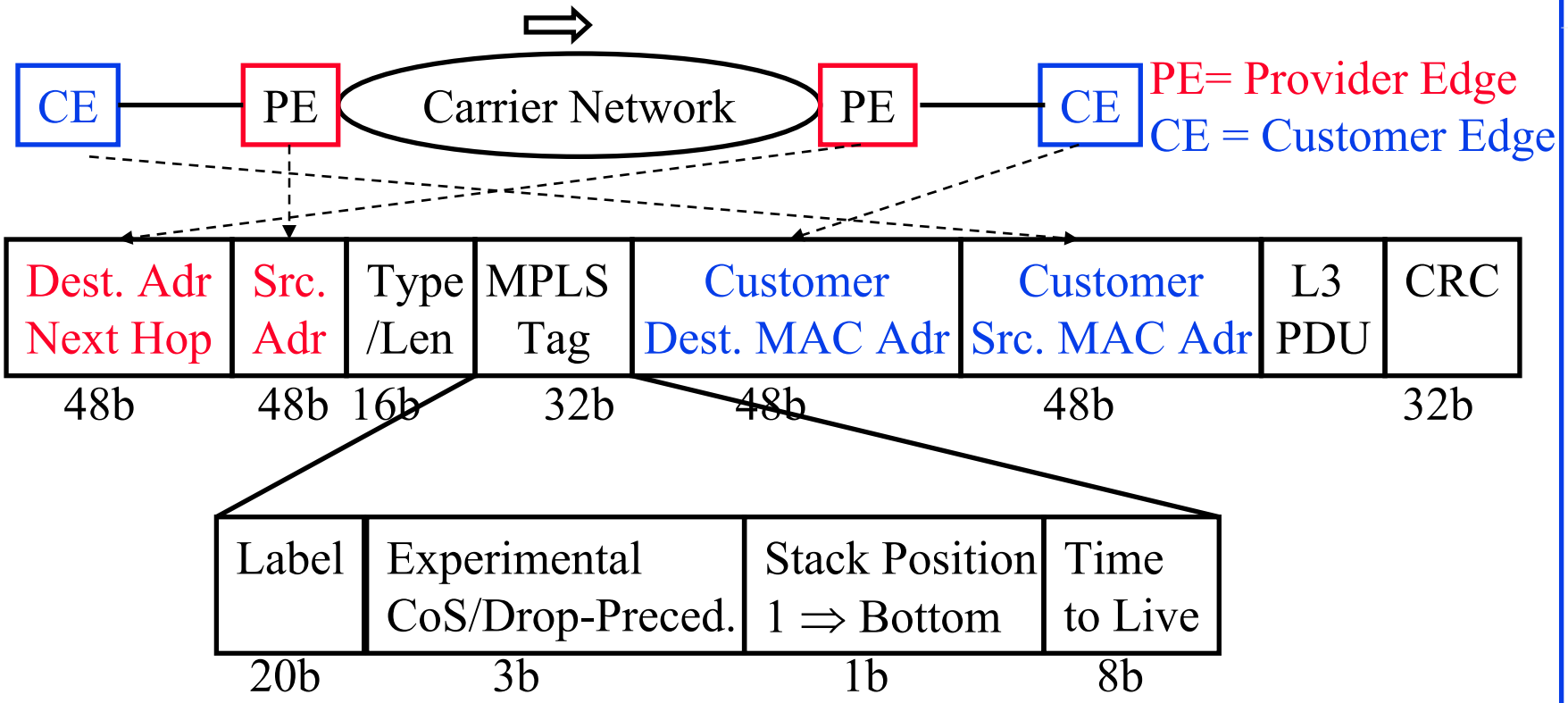
- Allows virtual circuits in IP Networks (May 1996)
- Each packet has a virtual circuit number called 'label'
- Label determines the packet's queuing and forwarding
- Circuits are called Label Switched Paths (LSPs)
- LSP's have to be set up before use
- Allows traffic engineering



Label Switching Example



IP over MPLS over Ethernet

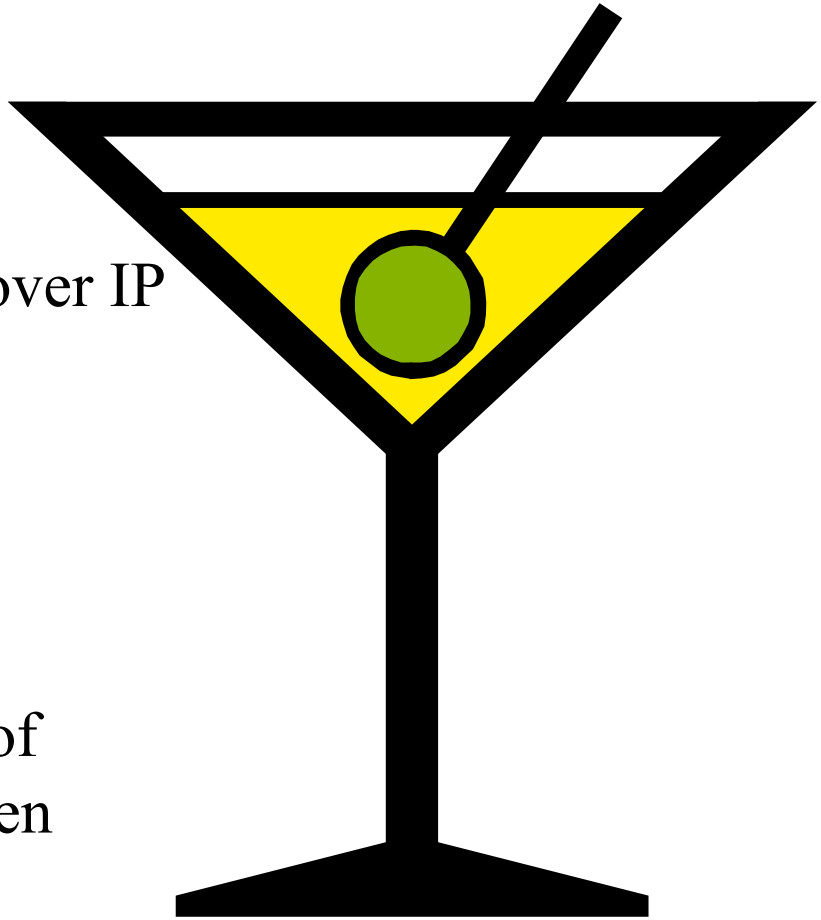


- Allows 2^{20} Label switched paths (LSP)
- Each path can have reserved capacity
  ⇒ MPLS became a very popular

Martini Draft

- 1995-1999: IP over ATM, IP over Ethernet

L3	IP		
L2	Ethernet	ATM	PPP

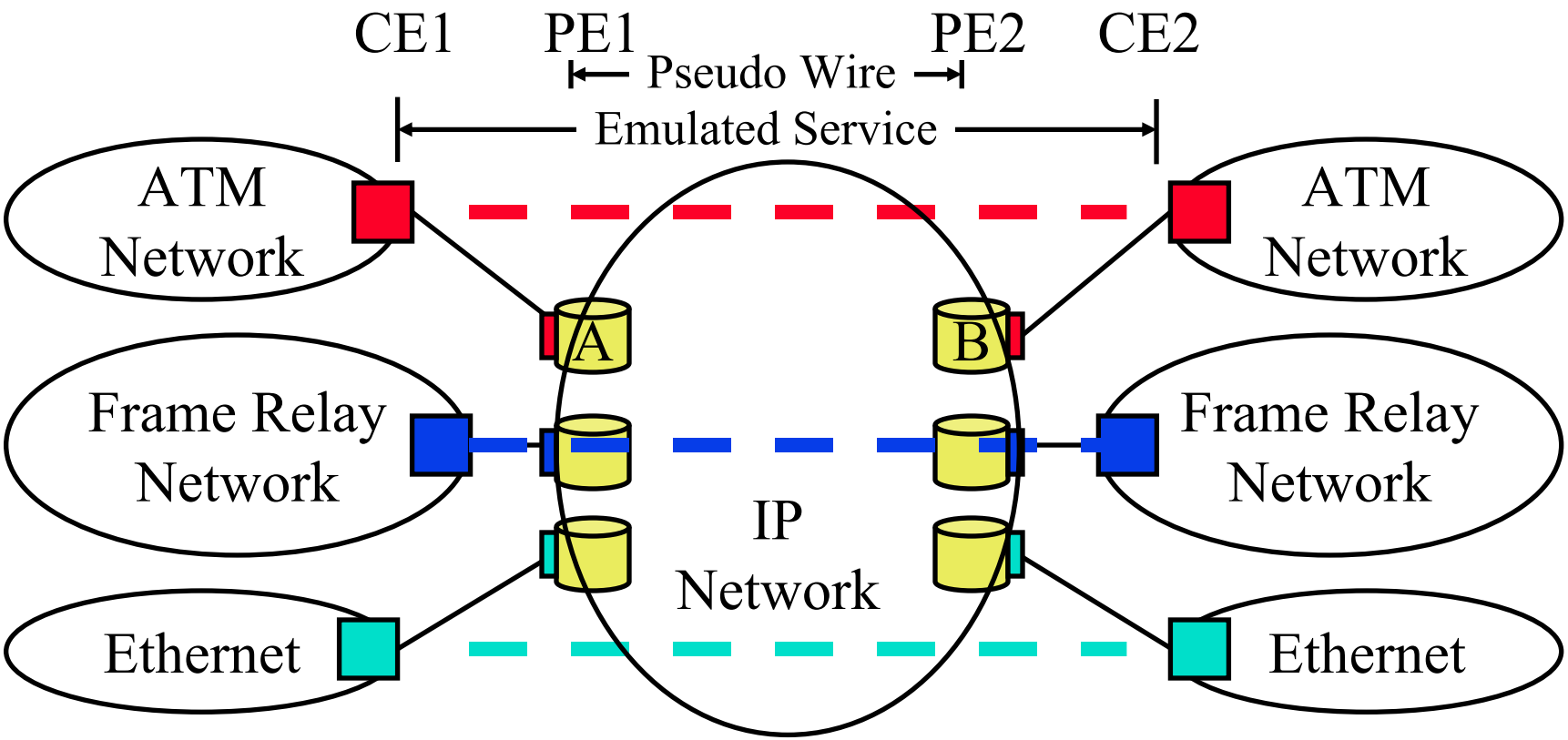


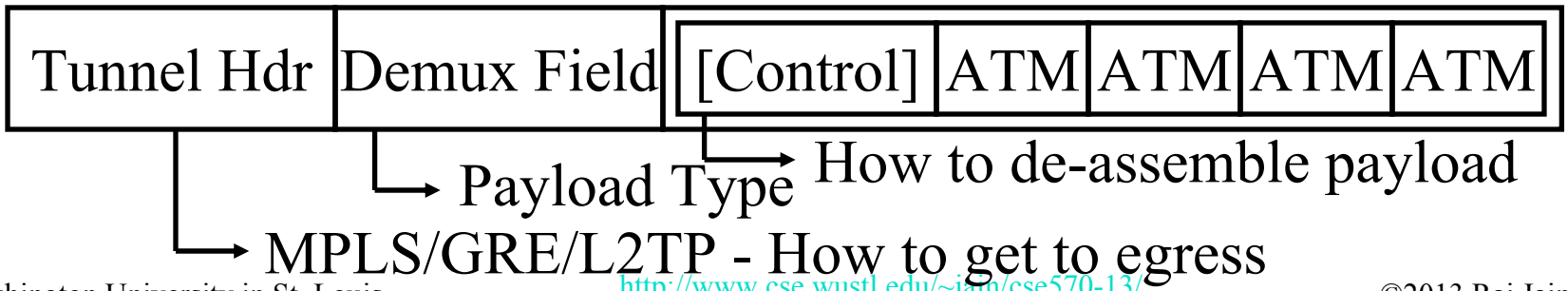
- 2000+: ATM over IP, Ethernet over IP
  SONET over IP

L2	Ethernet	ATM	PPP
L3	IP		

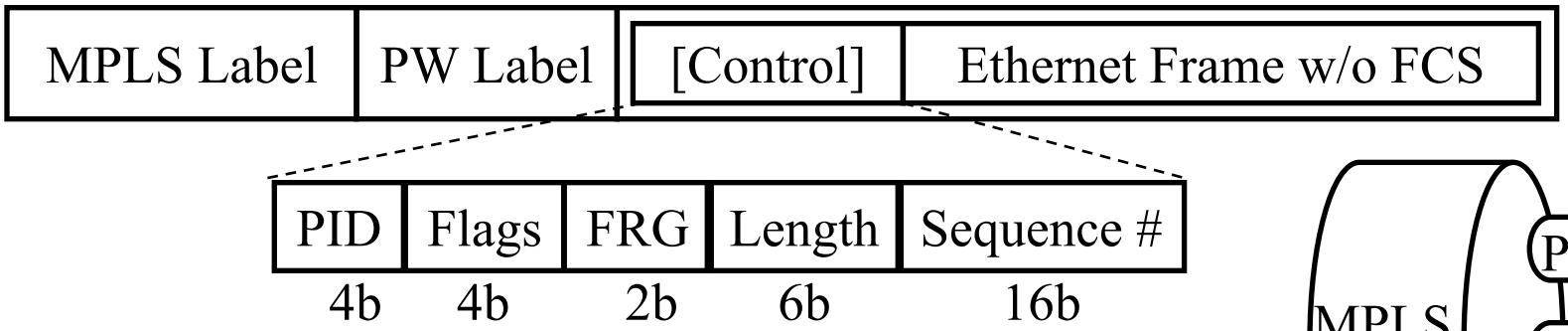
- Idea proposed by *Luca Martini* of Level 3 Communications and then Cisco

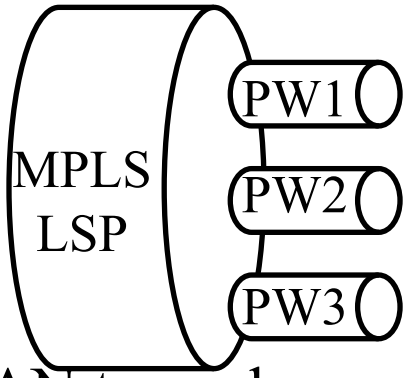
Pseudo Wire: L2 Circuits over IP





Ethernet over PWE3 over MPLS

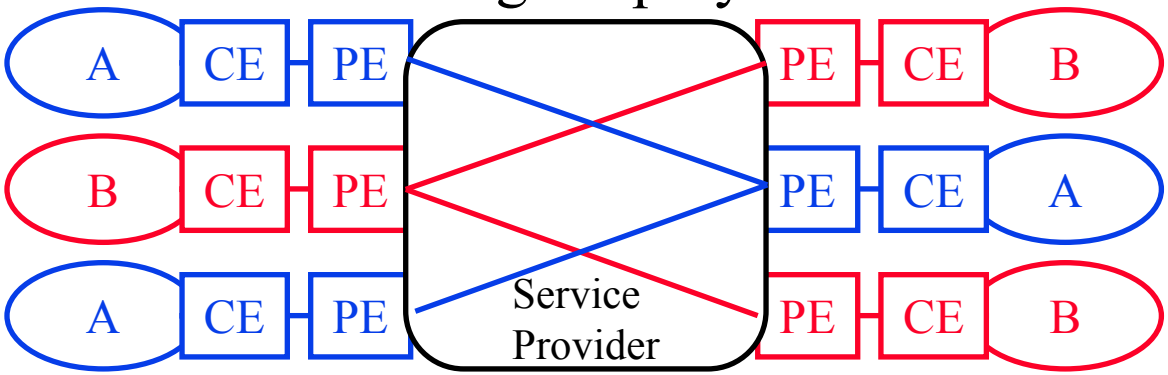




- Pseudo-Wire Emulation Edge-to-Edge (PWE3)
- Multiple pseudo-wires per MPLS LSP
- Payload ID (PID): 5=Untagged Ethernet, 4=VLAN tagged, …
- 4⇒VLAN tag put by carrier and customers may or may not be relevent for forwarding. Determined administratively by PE.
- Flags: Payload specific. FRG: Used for fragmentation
- Pause frames are obeyed locally. Not transported.
- May put 802.1p priority in exp field of MPLS label

Virtual Private LAN Service (VPLS)

- Allows *multi-point* Ethernet services over MPLS networks using *pseudo-wires*
- Non-Hierarchical VPLS: Edge routers do complete routing table, label distribution, multicast/broadcast replication
  ⇒ Suitable for small deployments
- Hierarchical VPLS (H-VPLS): Edge routers communicate with only other edge devices and do not have large routing tables
  ⇒ Suitable for large deployments

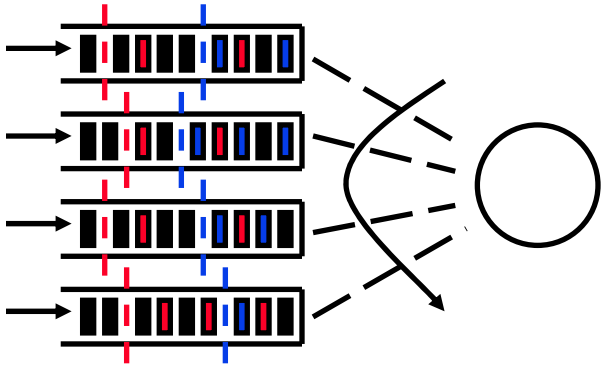


CE: Customer Edge
PE: Provider Edge

Ref: H. Saboowala, M. Abid, S. Modali, "Designing Networks and Services for the Cloud: Delivering business-grade cloud applications and services," Cisco Press 2013, ISBN:1587142945

Differentiated Services

- A way for IP routers to provide QoS
- Expedited Forwarding (EF): Also known as Premium Service
  - Virtual leased line ⇒ Guaranteed minimum service rate
  - Policed: Arrival rate < Minimum Service Rate
  - Not affected by other forwarding classes
- Assured Forwarding (AF):
  - Four Classes: No particular ordering
  - Three drop preference per class: Low, Medium, High

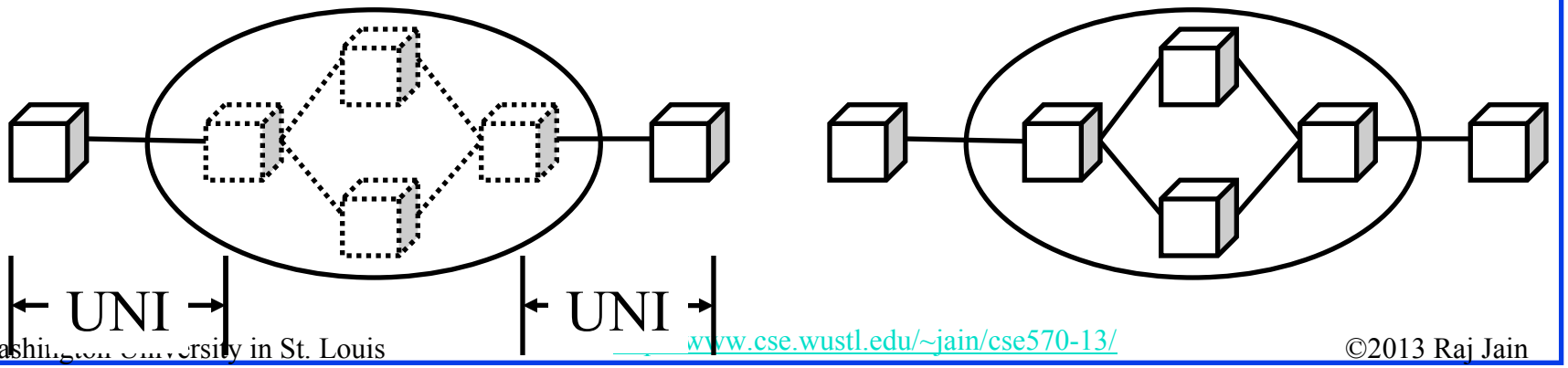


- Best Effort Service
- Differentiated Service Code Point (6 bits) encode the service, E.g., 101110 = EF

Carriers vs. Enterprise

Issue: UNI vs Peer-to-Peer Signaling

- Two Business Models:
  - Carrier: Overlay or cloud
    - Network is a black-box
    - User-to-network interface (UNI) to create/destroy light paths (in OIF)
  - Enterprise: Peer-to-Peer
    - Complete exchange of information



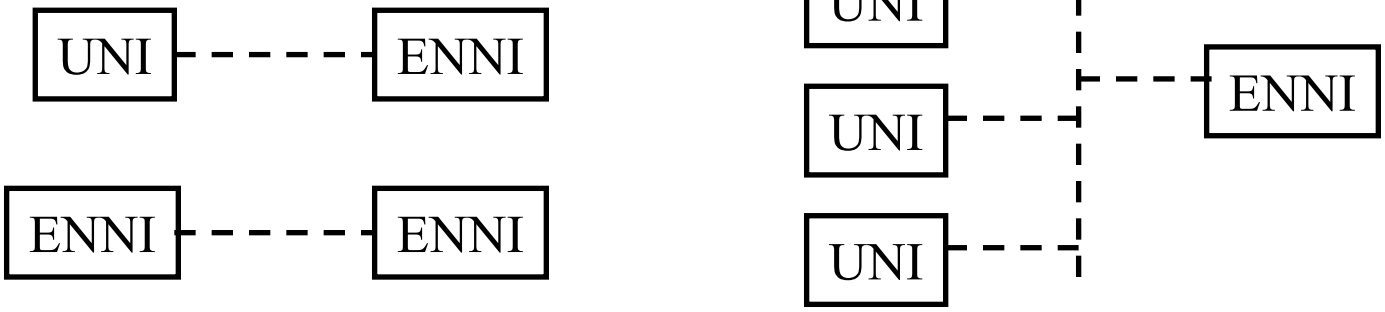
UNI vs. ENNI

- **User to Network Interface (UNI)**:
  - Separates responsibilities between the user and the provider. (Troubleshooting, failures etc).
  - Like the wired phone box outside your home.
  - Only one customer's traffic.
- **External Network to Network Interface (ENNI)**:
  - Separates responsibilities between two providers.
  - Many customer's traffic passes through an ENNI
  - Tier 2 *operators* sell services to Tier 3 service providers.

Ref: Fujitsu, "Carrier Ethernet Essentials," http://www.fujitsu.com/downloads/TEL/fnc/whitepapers/CarrierEthernetEssentials.pdf

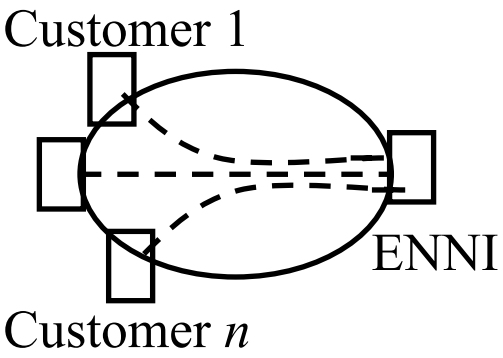
Operator Virtual Connection (OVC)

- Between UNI and ENNI or between two ENNIs.
- For wholesale service providers
- Two types: Point-to-Point and Multipoint-to-Multipoint
- Untagged or single tagged frames at NNI. Q-in-Q at ENNI
- UNIs may be 10 to 100 Mbps. ENNIs at 1 to 10 Gbps.



Metro Access Ethernet Private Line

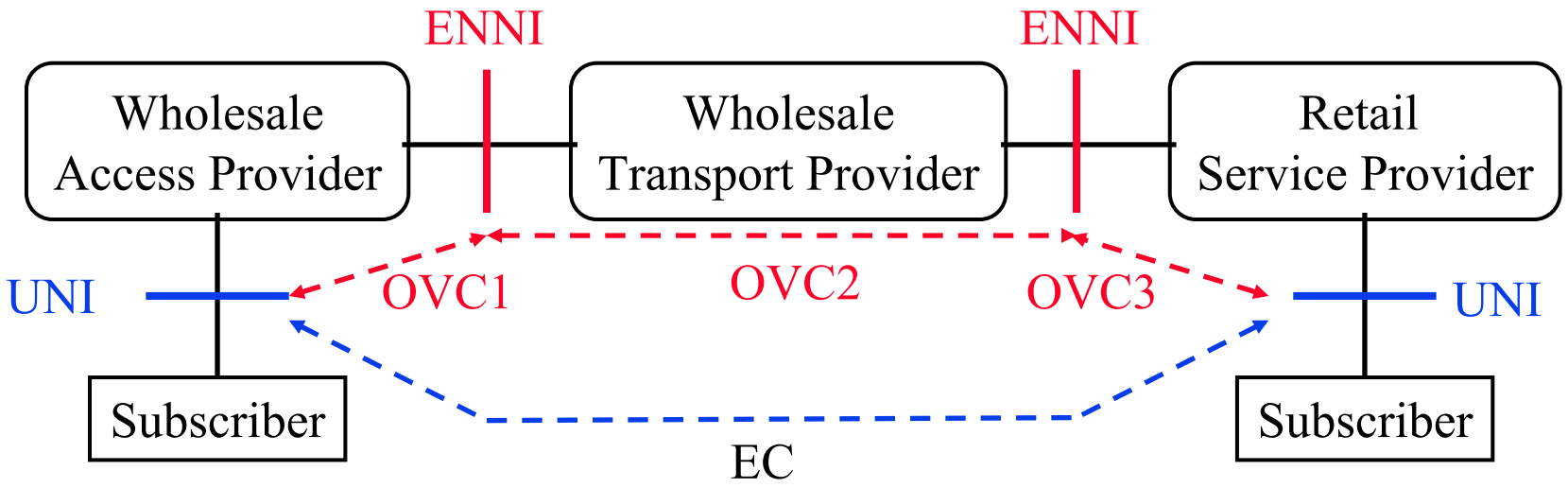
- **Access Ethernet Private Line (Access-EPL)**:
  - Port-based service for Internet access
    Like the service at your home.
  - Ends at your access provider, where many other Access-EPLs may end
  - Access provider has only one interface
    Shared by many Access-EPLs ⇒Different from p2p EPL.



E-Access

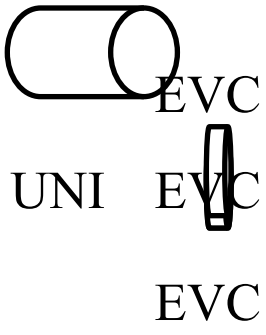
End-to-End Metro Ethernet Connection

- An EC may go through multiple service providers
  ⇒ Multiple OVCs can be concatenated to create an EC



Ethernet Virtual Connections (EVCs)

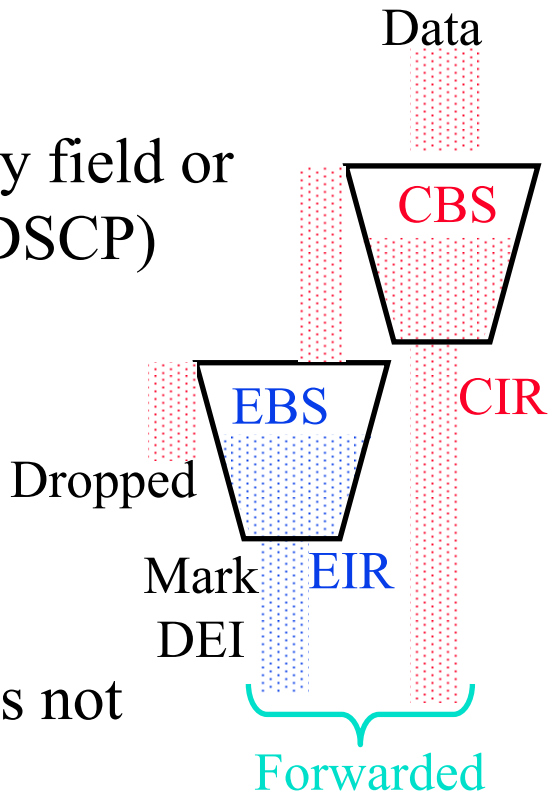
- **Port-based ECs**: Forwarding not based on VLANs.
  Frames delivered to remote UNI/ENNI for P2P or
  Based on destination address for P2MP
- **VLAN-based ECs**: Forwarding based on VLAN tag.
  ⇒ Multiple Virtual UNIs
  ⇒ Ethernet *Virtual* Connection (*EVC*)
  More cost-effective for Enterprise customers



- Types of EVCs:
  1. Ethernet Virtual Private Line (EVPL)
  2. Ethernet Virtual Private Tree (EVP-Tree)
  3. Ethernet Virtual Private LAN (EVPLAN)
  4. Access Ethernet Virtual Private Line (Access EVPL)
- Note: Service providers always share an ENNI for multiple connections ⇒ OVCs are always virtual ⇒ No OCs

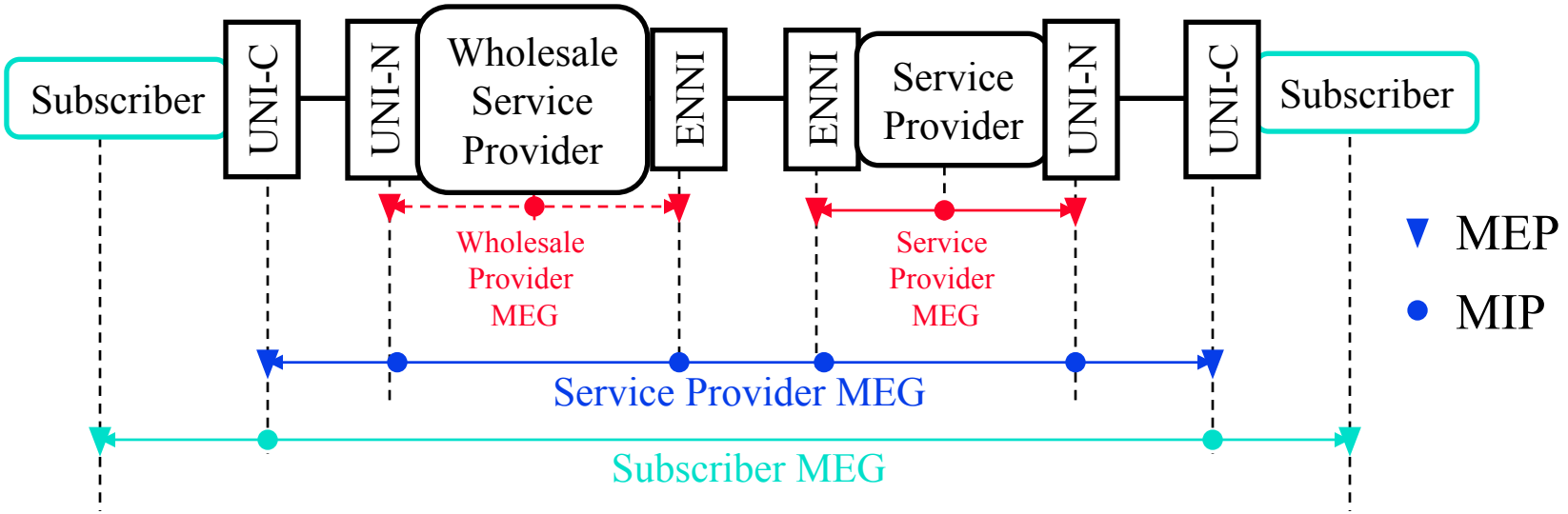
Metro Ethernet Service Attributes

- Bandwidth Profiles: Limits on data dates
  - Ingress Profile: Incoming data rate
  - Egress Profile: Outgoing data rate
- Per UNI, Per EVC or OVC, or Per EVC/OVC per Class of Service (CoS)
- CoS is indicated by the 3-bits in the priority field or 4-bit Differentiated Services Code Point (DSCP)
- Rate specified by 5 parameters
  1. Committed Information Rate (CIR)
  2. Committed Burst Size (CBS)
  3. *Excess* Information rate (EIR)
  4. Excess Burst Size (EBS)
  5. Color Mode (CM): Customer does/does not mark drop eligibility indicator (DEI)



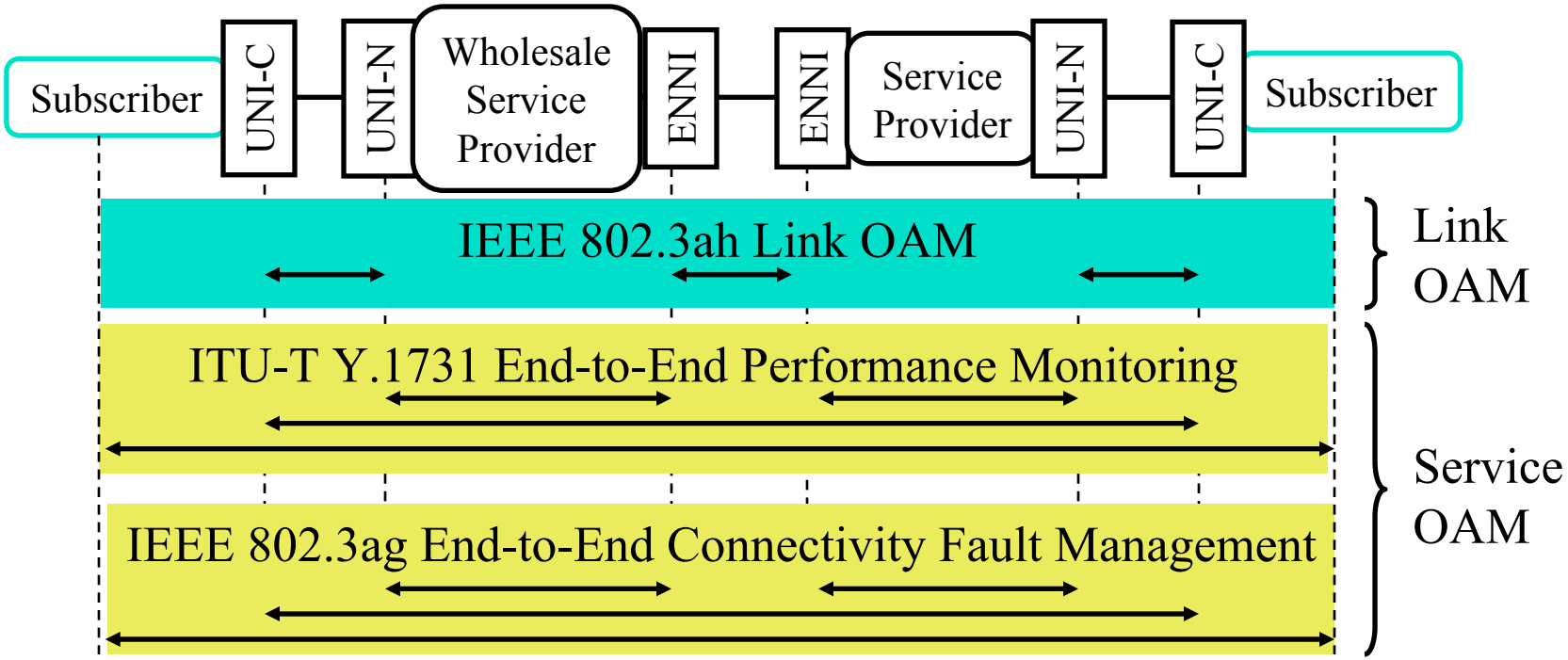
Metro Ethernet OAM

- Operation, Administration and Maintenance (OAM)
- Defined in IEEE 802.1ag, IEEE 802.3ah, and ITU Y.1731
- Maintenance end points (MEPs)
- Maintenance Intermediate Points (MIPs)
- Maintenance Entity Group (MEG): Level of Administration



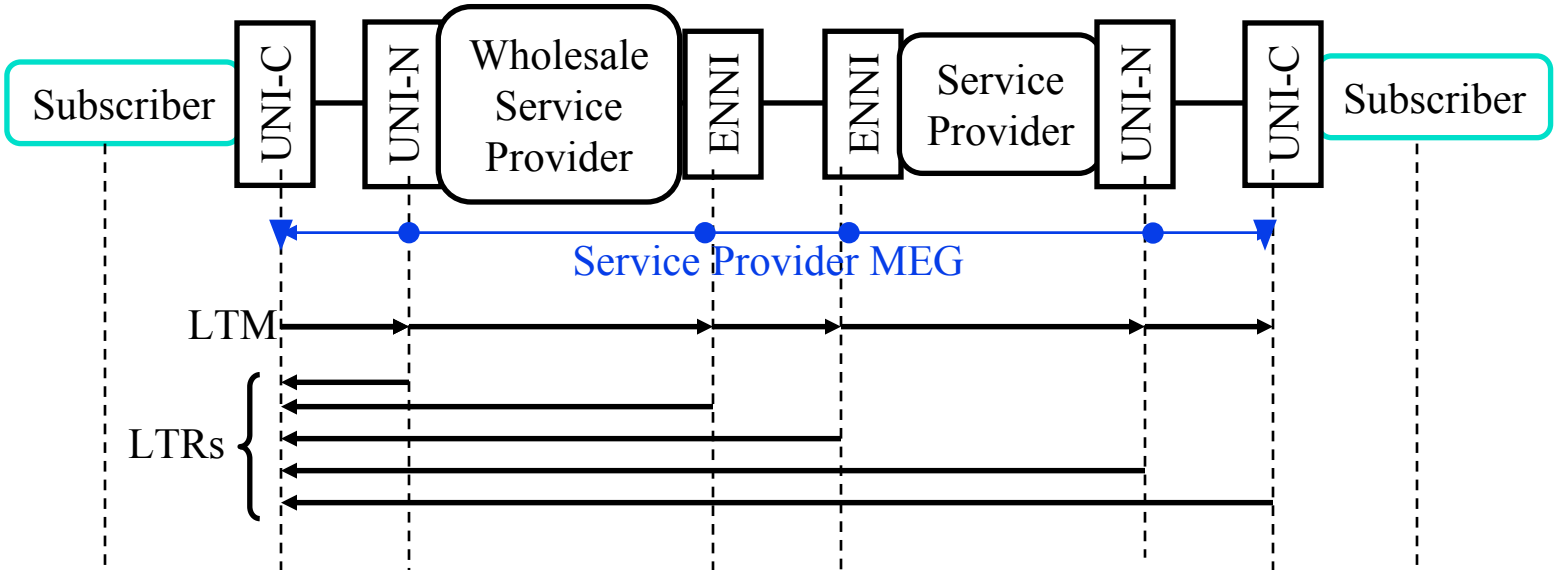
Metro Ethernet OAM (Cont)

- Performance Monitoring: Measure throughput and latency
- Connectivity Fault Management: Monitor downtime
  - Service Fault Management
  - Link Fault Management



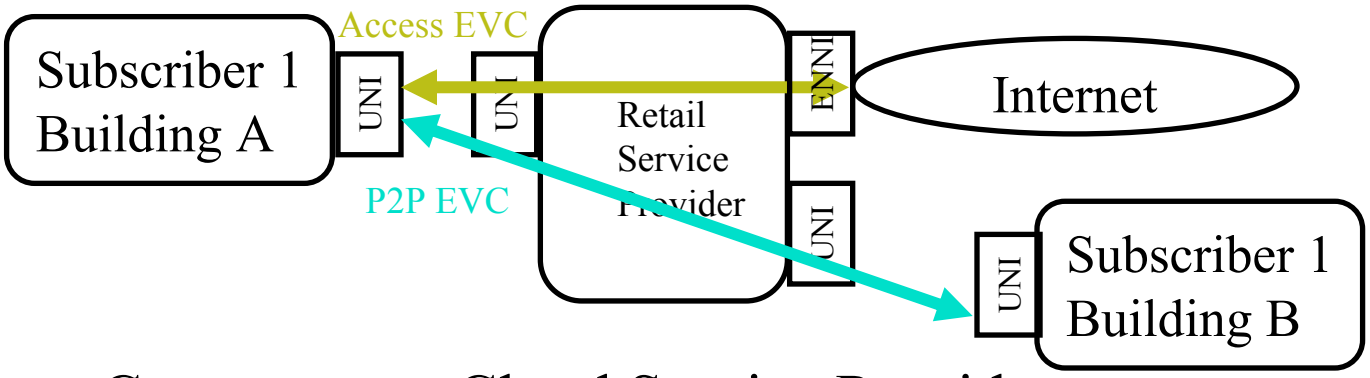
Metro Ethernet OAM Messages

- Continuity Check Message (CCM) in both directions (Similar to IP Ping)
- Link Trace Message (LTM): Locates fault. Link Trace Response (LTR) is returned by each end point and intermediate point (similar to IP trace route)

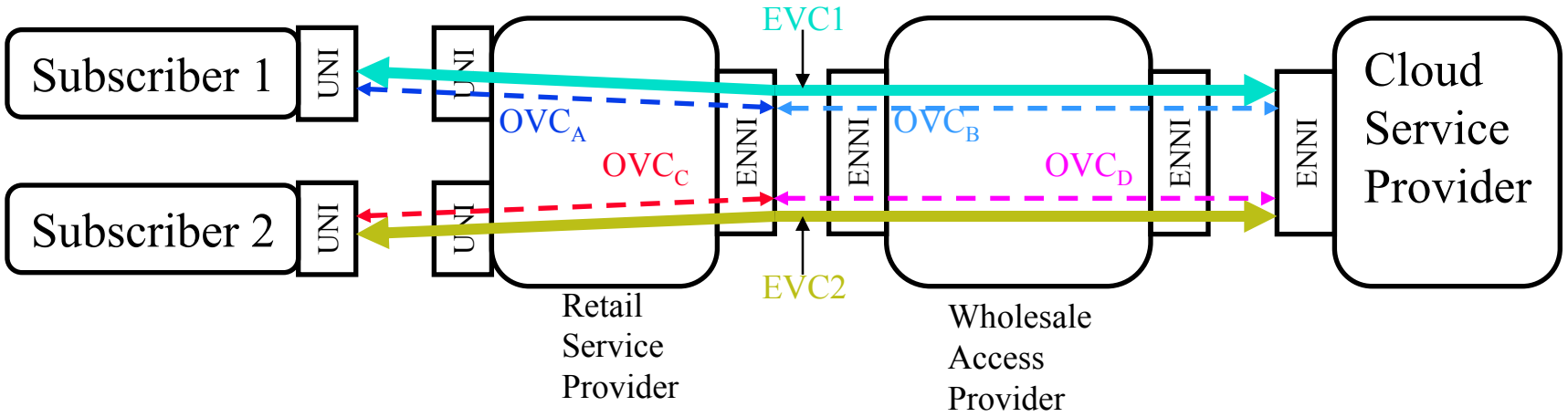


Metro Ethernet Use Cases

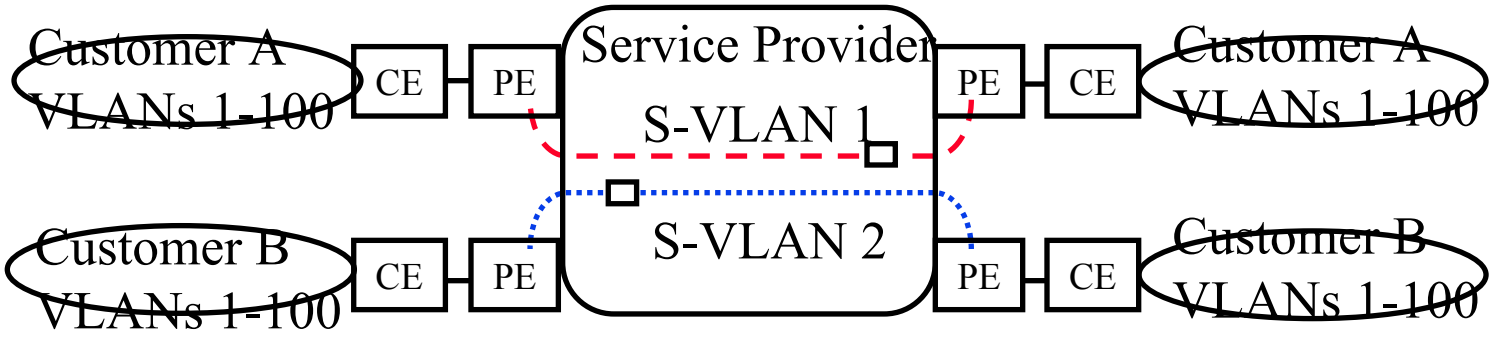
1. Head office to Satellite offices and/or Internet



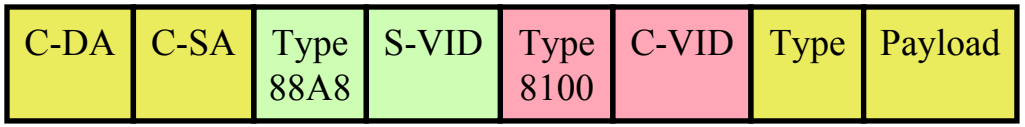
2. Customers to Cloud Service Provider



Ethernet Provider Bridge (PB)



- IEEE 802.1ad-2005 incorporated in IEEE 802.1Q-2011
- Problem: Multiple customers may have the same VLAN ID. How to keep them separate?
- Solution: Q-in-Q. Provider inserts a service VLAN tag
- Allows 4K customers to be serviced. Total 16M VLANs



Ref: D. Bonafede, "Metro Ethernet Network," http://www.cicomra.org.ar/cicomra2/asp/TUTORIAL-%20Bonafede.pdf
Ref: P. Thaler, et al., "IEEE 802.1Q," IETF tutorial, March 10 2013, http://www.ietf.org/meeting/86/tutorials/86-IEEE-8021-Thaler.pdf

Provider Bridge (Cont)

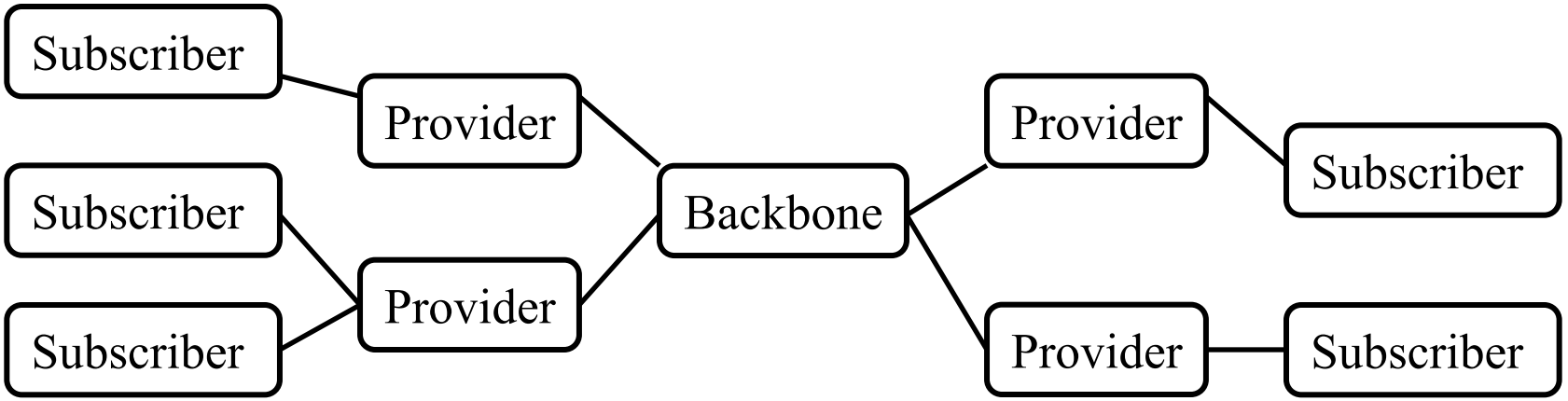
- 8 Traffic Classes using Differentiated Services Code Points (DSCP) for Assured Forwarding

S-Tag

Priority	CFI	S-VLAN ID	Type/Length
3b	1b	12b	16b

Priority	Class
7	Strict Priority
6	AF1
5	AF2
4	AF3
3	AF4
2	AF5
1	AF6
0	Best Effort

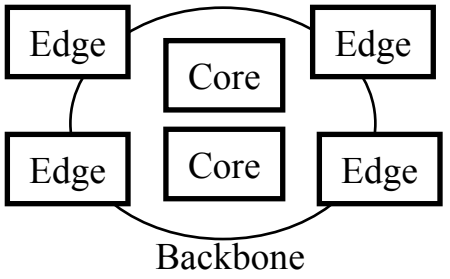
Provider Backbone Network (PBB)

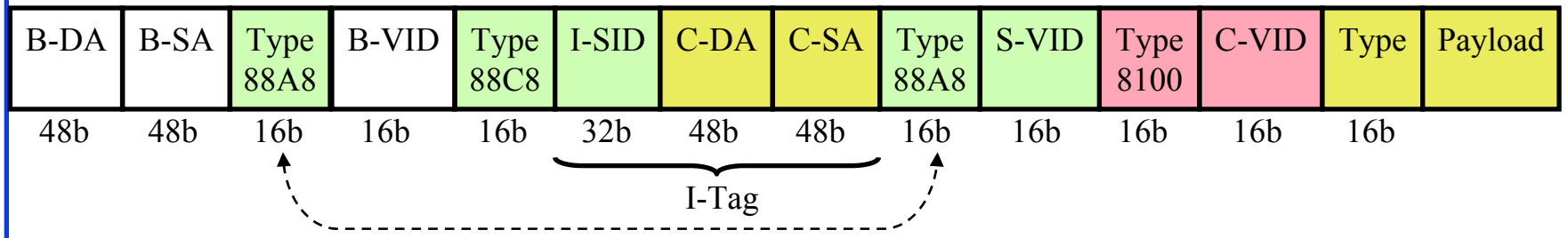


- Problem: Number of MAC addresses passing through backbone bridges is too large for all core bridge to remember
  Broadcast and flooded (unknown address) frames give unwanted traffic and security issues
- Solution: IEEE 802.1ah-2008 now in 802.1Q-2011
- Add new source/destination MAC addresses pointing to ingress backbone bridge and egress backbone bridge
  ⇒Core bridges only know edge bridge addresses

MAC-in-MAC Frame Format

- Provider backbone edge bridges (PBEB) forward to other PBEB's and learn customer MAC addresses
  ⇒ PB *core* bridges do not learn customer MACs
- B-DA = Destination backbone bridge address Determined by Customer Destination Address
- Backbone VLANs delimit the broadcast domains in the backbone



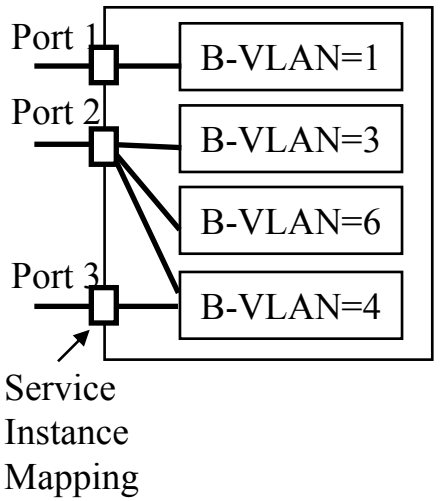


- PBB Core switches forward based on Backbone Destination Bridge Address and Backbone-VLAN ID (60 bits)
  Similar to 802.1ad Q-in-Q. Therefore, same EtherType.

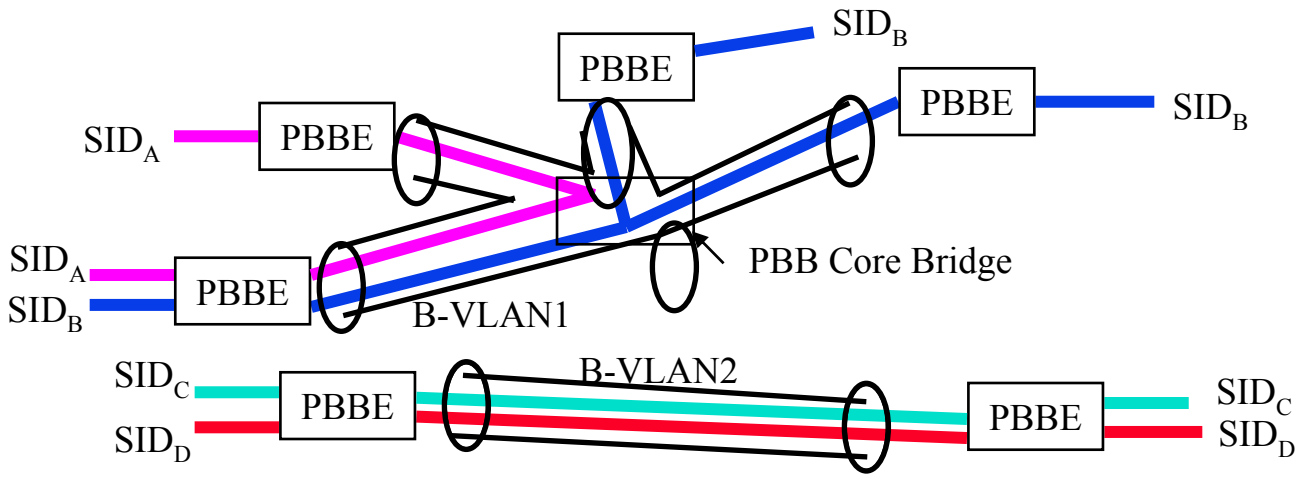
PBB Service Instance

- Service instance ID (I-SID) indicates a specific flow
  - All frames on a specific port, or
  - All frames on a specific port with a specific *service* VLAN, or
  - All frames on a specific port with a specific service VLAN and a specific *customer* VLAN

SID	Definition	B-VLAN
1	Port 1	1
20	Port 2, S-VLAN=10	3
33	Port 2, S-VLAN=20	6
401	Port 2, S-VLAN=30, C-VLAN=100	4
502	Port 3, S-VLAN=40, C-VLAN=200	4



MAC-in-MAC (Cont)

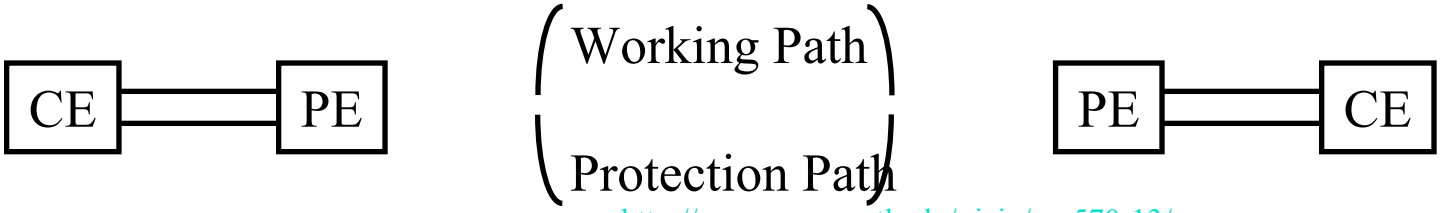


- Each Backbone VLANs (B-VLAN) can carry multiple services
- 24-bit SID $\Rightarrow 2^{24}$ Service Instances in the backbone
- I-Tag format: I-Tag not used in the core. Includes C-DA+C-SA
  UCA=1 $\Rightarrow$ Use customer addresses (used in CFM)

Priority Code Point (I-PCP)	Drop Eligibility Indicator (I-DEI)	Use Customer Address (UCA)	Reserved 1	Reserved 2	Service Instance ID (I-SID)	Customer Destination Address (C-DA)	Customer Source Address (C-SA)
3b	1b	1b	1b	2b	24b	48b	48b

Connection Oriented Ethernet

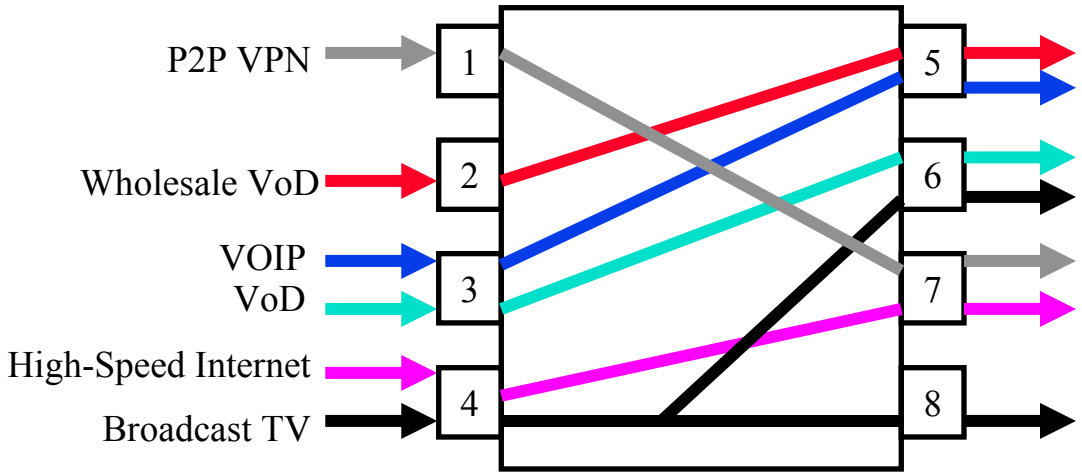
- Connectionless: Path determined at forwarding
  ⇒ Varying QoS
- Connection Oriented: Path determined at provisioning
  - Path provisioned by management ⇒ Deterministic QoS
    - No spanning tree, No MAC address learning,
    - Frames forwarded based on VLAN Ids and Backbone bridges addresses
    - Path not determined by customer MAC addresses and other customer fields ⇒ More Secure
  - Reserved bandwidth per EVC
  - Pre-provisioned Protection path ⇒ Better availability



VLAN Cross-Connect

- Cross-connect ⇒ Circuit oriented
- Connection Oriented Ethernet with Q-in-Q
- Forward frames based on VLAN ID and Input port
  ⇒ No MAC Learning

Input Port	VLAN ID	Output Port
1	200	7
2	201	5
3	20	5
3	21	6
4	100	7
4	101	8

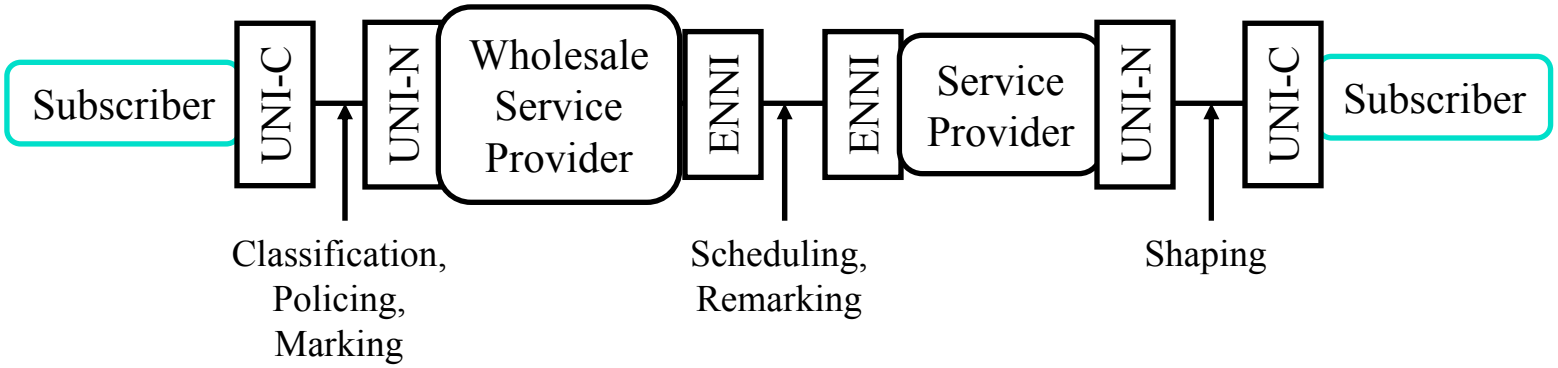


PBB-TE

- Provider Backbone Bridges with Traffic Engineering (PBB-TE)
- IEEE 802.1Qay-2009 now in 802.1Q-2011
- Provides connection oriented P2P (E-*Line*) Ethernet service
- For PBB-TE traffic VLANs:
  - Turn off MAC learning
  - Discard frames with unknown address and broadcasts.
    ⇒ No flooding
  - Disable Spanning Tree Protocol.
  - Add protection path switching for each direction of the trunk
- Switch forwarding tables are administratively populated using management
- Same frame format as with MAC-in-MAC. No change.

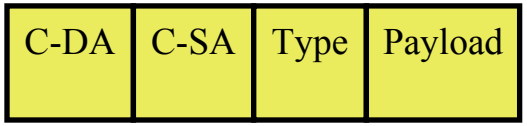
PBB-TE QoS

- Guarantees QoS ⇒ No need for MPLS or SONET/SDH
- UNI traffic is classified by Port, Service VLAN ID, Customer VLAN ID, priority, Unicast/Multicast
- UNI ports are *policed* ⇒ Excess traffic is dropped
  No policing at NNI ports. Only remarking, if necessary.
- Traffic may be marked and remarked at both UNI and NNI

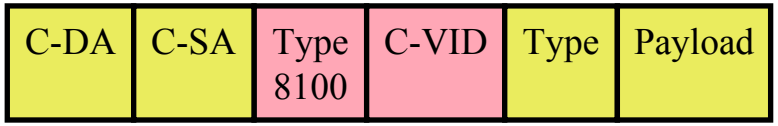


Ethernet Tagged Frame Format Evolution

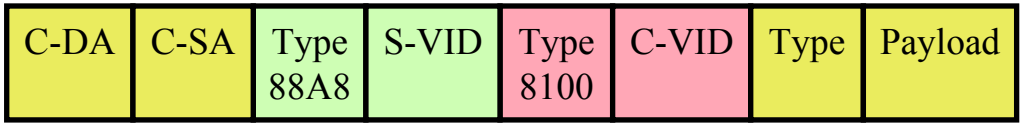
- Original Ethernet



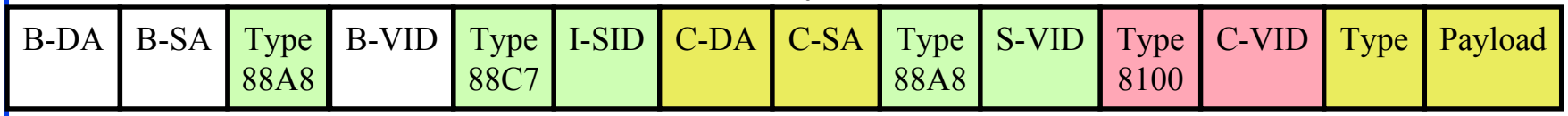
- IEEE 802.1Q VLAN



- IEEE 802.1ad PB



- IEEE 802.1ah PBB or 802.1Qay PBB-TE

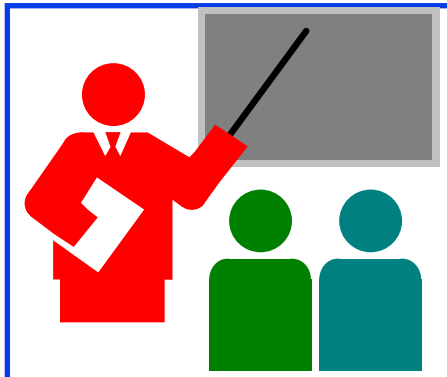


Tag Type	Value
Customer VLAN	8100
Service VLAN or Backbone VLAN	88A8
Backbone Service Instance	88C7

Comparison of Technologies

	Basic Ethernet	MPLS	PB	PBB-TE
Resilience	No	Protection Fast Reroute	SPB/LAG	Protection Fast Reroute
Security	No	Circuit Based	VLAN	Circuit Based
Multicast	Yes	Inefficient	Yes	No. P2P only
QoS	Priority	Diffserve	Diffserve+ Guaranteed	Diffserve+ Guaranteed
Legacy Services	No	Yes (PWE3)	No	No
Traffic Engineering	No	Yes	No	Yes
Scalability	Limited	Complex	Q-in-Q	Q-in-Q+ Mac-in-MAC
Cost	Low	High	Medium	Medium
OAM	No	Some	Yes	Yes

Ref: Bonafede
Washington University in St. Louis



Summary

1. Carrier networks are moving away from voice oriented networks (SDH, PDH) to data oriented networks (MPLS, Ethernet)
2. Ethernet over Pseudo wires over MPLS is used to interconnect Ethernet switches over long distances
3. Metro Ethernet Forum has defined: EVPL, EVP-LAN, EVP-Tree, and EVP-Access services
4. Q-in-Q and MAC-in-MAC extensions allow very large Ethernet networks spanning over several backbone carriers
5. PBB-TE extension allows connection oriented Ethernet with QoS guarantees and protection

Reading List

- Fujitsu, "Carrier Ethernet Essentials," http://www.fujitsu.com/downloads/TEL/fnc/whitepapers/CarrierEthernetEssentials.pdf (must read)
- D. Bonafede, "Metro Ethernet Network," http://www.cicomra.org.ar/cicomra2/asp/TUTORIAL-%20Bonafede.pdf
- P. Thaler, et al., "IEEE 802.1Q," IETF tutorial, March 10 2013, http://www.ietf.org/meeting/86/tutorials/86-IEEE-8021-Thaler.pdf
- H. Saboowala, M. Abid, S. Modali, "Designing Networks and Services for the Cloud: Delivering business-grade cloud applications and services," Cisco Press 2013, ISBN:1587142945 (Safari Book)

Wikipedia Links

- http://en.wikipedia.org/wiki/Carrier_Ethernet
- http://en.wikipedia.org/wiki/Connection-oriented_Ethernet
- http://en.wikipedia.org/wiki/Differentiated_services
- http://en.wikipedia.org/wiki/Ethernet_Private_Line
- http://en.wikipedia.org/wiki/Ethernet_Virtual_Private_Line
- http://en.wikipedia.org/wiki/IEEE_802.1ad
- http://en.wikipedia.org/wiki/IEEE_802.1ag
- http://en.wikipedia.org/wiki/IEEE_802.1ah-2008
- http://en.wikipedia.org/wiki/Label-switched_path
- http://en.wikipedia.org/wiki/Label_Distribution_Protocol
- http://en.wikipedia.org/wiki/Link_protection
- http://en.wikipedia.org/wiki/Metro_Ethernet
- http://en.wikipedia.org/wiki/Metro_Ethernet_Forum
- http://en.wikipedia.org/wiki/MPLS-TP
- http://en.wikipedia.org/wiki/MPLS_local_protection
- http://en.wikipedia.org/wiki/MPLS_VPN

Wikipedia Links (Cont)

- http://en.wikipedia.org/wiki/Multiprotocol_Label_Switching
- http://en.wikipedia.org/wiki/Network-to-network_interface
- http://en.wikipedia.org/wiki/Operations,_administration_and_management
- http://en.wikipedia.org/wiki/Optical_Carrier_transmission_rates
- http://en.wikipedia.org/wiki/Optical_Transport_Network
- http://en.wikipedia.org/wiki/Path_protection
- http://en.wikipedia.org/wiki/Plesiochronous_digital_hierarchy
- http://en.wikipedia.org/wiki/Provider_Backbone_Bridge_Traffic_Engineering
- http://en.wikipedia.org/wiki/Pseudo-wire
- http://en.wikipedia.org/wiki/Resilient_Packet_Ring
- http://en.wikipedia.org/wiki/Synchronous_optical_networking
- http://en.wikipedia.org/wiki/Traffic_policing
- http://en.wikipedia.org/wiki/Traffic_shaping
- http://en.wikipedia.org/wiki/User%E2%80%93network_interface
- http://en.wikipedia.org/wiki/Virtual_Private_LAN_Service
- http://en.wikipedia.org/wiki/Wavelength-division_multiplexing

Acronyms

- AF Assured Forwarding
- ATM Asynchronous Transfer Mode
- B-VLAN Backbone VLAN
- BGP Border Gateway Protocol
- CBS Committed Burst Size
- CCM Continuity Check Message
- CIR Committed Information Rate
- CM Color Mode
- CoS Class of Service
- DA Destination Address
- DEI Drop Eligibility Indicator
- DSCP Differentiated Services Code Points
- DWDM Dense Wavelength Division Multiplexing
- EBS Excess Burst Size

Acronyms (Cont)

- EC Ethernet Connection
- ECMP Equal-cost Multipathing
- EF Expedited Forwarding
- EIR Excess Information rate
- ENNI External Network to Network Interface
- EPL Ethernet Private Line
- EVC Ethernet Virtual Connection
- EVP-Access Ethernet Virtual Private Access
- EVP-LAN Ethernet Virtual Private Local Area Network
- EVP-Line Ethernet Virtual Private Line
- EVP-Tree Ethernet Virtual Private Tree
- EVPL Ethernet Virtual Private Line
- FRG Fragmentation
- ID Identifier

Acronyms (Cont)

- ITU International Telecommunications Union
- LAN Local Area Network
- LDP Label Distribution Protocol
- LSP Label Switched Paths
- LSR Label Switching Router
- LTM Link Trace Message
- LTR Link Trace Response
- MAC Media Access Control
- MEG Maintenance Entity Group
- MEP Maintenance End Points
- MIP Maintenance Intermediate Points
- MP Multi-Point
- MPLS Multi-Protocol Label Switching
- MTU Maximum Transmission Unit
- NNI Network-to-Network Interface

Acronyms (Cont)

- OAM Operation, Administration and Maintenance
- OC Optical Carrier
- OIF Optical Interoperability Forum
- OTN Optical Transmission Network
- OVC Operator Virtual Connection
- PB Provider Bridge
- PBB-TE Provider Backbone Bridge with Traffic Engineering
- PBB Provider Backbone Bridge
- PBEB Provider backbone edge bridges
- PDH Plesiochronous Digital Hierarchy
- PE Provider Edge
- PW Pseudo-Wire
- PWE3 Pseudo-Wire Emulation Edge-to-Edge
- QoS Quality of Service

Acronyms (Cont)

- SA Source Address
- SDH Synchronous Digital Hierarchy
- SHDSL Single Pair High-Speed Digital Subscriber Line
- SID Service Identifier
- SONET Synchronous optical network
- TE Traffic Engineering
- TTL Time to Live
- UNI User to Network Interface
- VC Virtual Circuit
- VID VLAN Identifier
- VLAN Virtual Local Area Network
- VPLS Virtual Private Line Service
- WDM Wavelength Division Multiplexing