

Multi-Protocol Label Switching (MPLS)

Tópicos Avançados de Redes
2023/2024

A note on the use of these ppt slides:

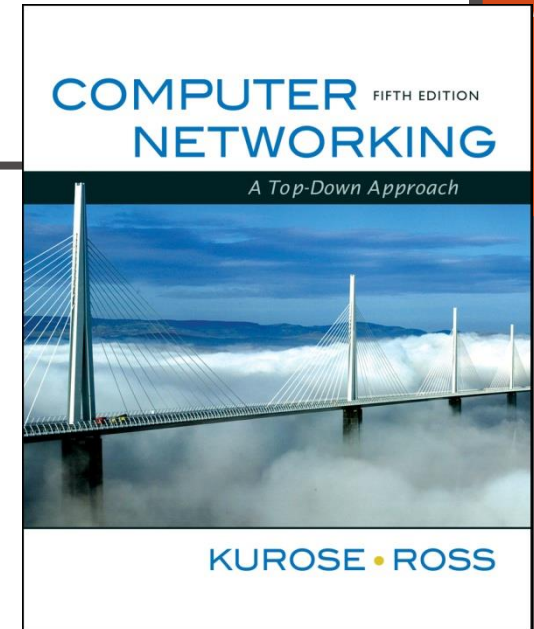
We're making these slides freely available to all (faculty, students, readers). They're in PowerPoint form so you can add, modify, and delete slides (including this one) and slide content to suit your needs. They obviously represent a *lot* of work on our part. In return for use, we only ask the following:

- ❑ If you use these slides (e.g., in a class) in substantially unaltered form, that you mention their source (after all, we'd like people to use our book!)
- ❑ If you post any slides in substantially unaltered form on a www site, that you note that they are adapted from (or perhaps identical to) our slides, and note our copyright of this material.

Thanks and enjoy! JFK / KWR

All material copyright 1996-2009
J.F Kurose and K.W. Ross, All Rights Reserved

With changes by pbrandao & rprior



*Computer Networking: A Top
Down Approach
5th edition.*

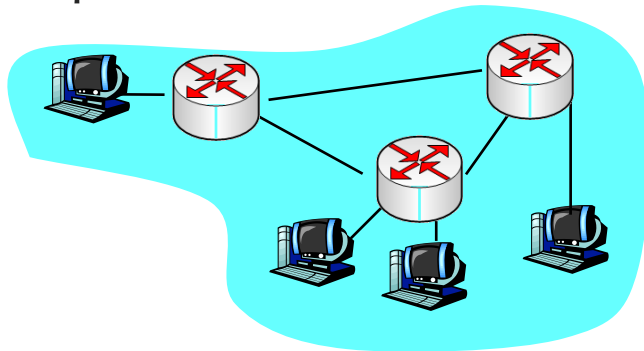
Jim Kurose, Keith Ross
Addison-Wesley, April 2009.

Virtualization of networks

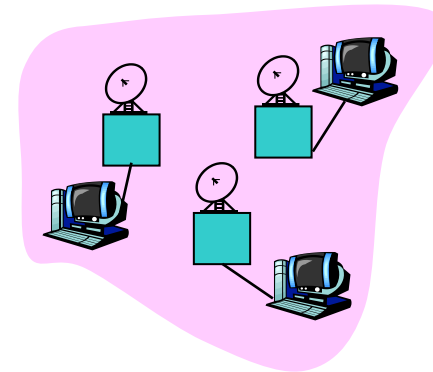
- Virtualization of resources is a powerful abstraction in systems engineering
- Computing examples: virtual memory, virtual devices
 - Virtual machines: hardware, abstract (e.g., JVM)
 - IBM VM OS from 1960's/70's (now z/VM)
 - SDN
- Layering of abstractions
 - Don't sweat the details of the lower layer, only deal with lower layers abstractly

The Internet: virtualizing networks

- 1974: multiple unconnected nets
 - ARPAnet
 - data-over-cable networks
 - packet satellite network (Aloha)
 - packet radio network
- ... differing in:
 - addressing conventions
 - packet formats
 - error recovery mechanisms
 - routing



ARPAnet



satellite net

"A Protocol for Packet Network Intercommunication",
V. Cerf, R. Kahn, IEEE Transactions on Communications,
May, 1974, pp. 637-648.

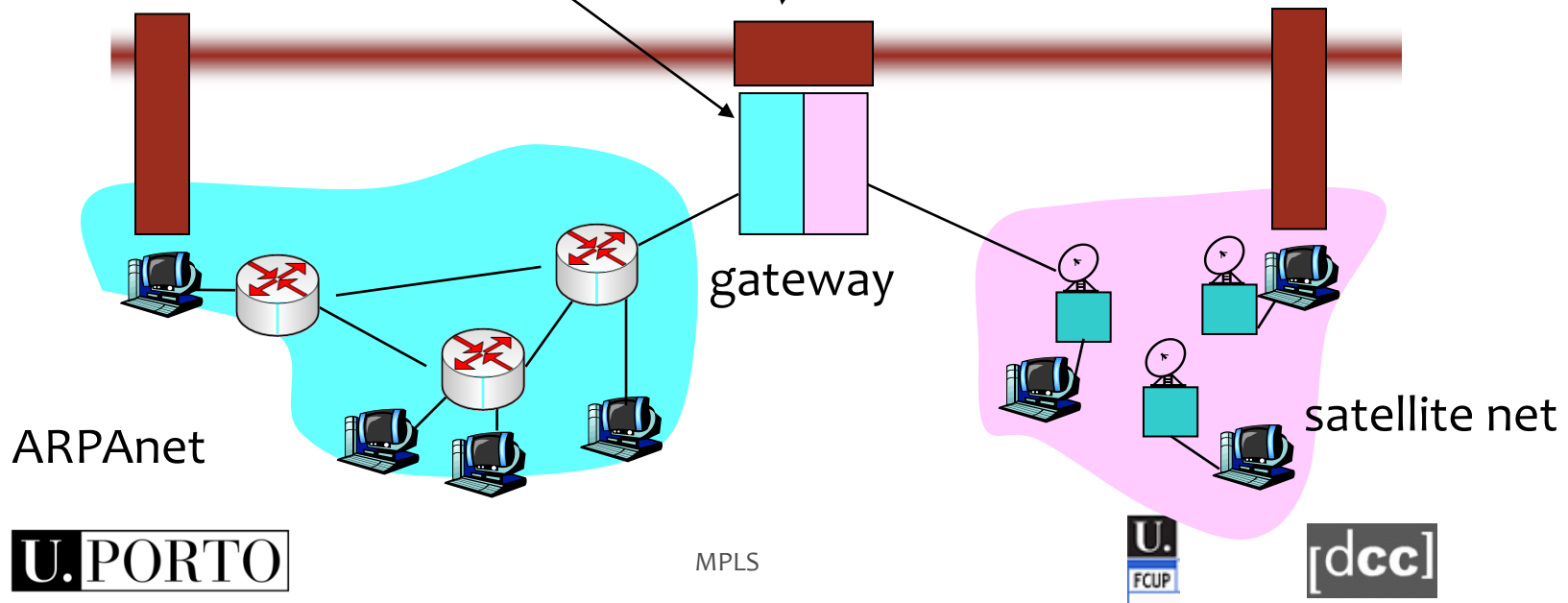
The Internet: virtualizing networks

Gateway:

- “embed internetwork packets in local packet format or extract them”
- route (at internetwork level) to next gateway

Internetwork layer (IP):

- addressing: internetwork appears as single, uniform entity, despite underlying local network heterogeneity
- network of networks



Cerf & Kahn's Internetwork Architecture

- Two layers of addressing: local network and internetwork
- New layer (IP) makes everything homogeneous at internetwork layer
- Underlying local network technology
 - ethernet
 - satellite
 - ATM
 - **MPLS**
- ... “invisible” at internetwork layer. Looks like a link layer technology to IP!

Virtual Circuit vs. Datagram Networks

Virtual Circuit Networks

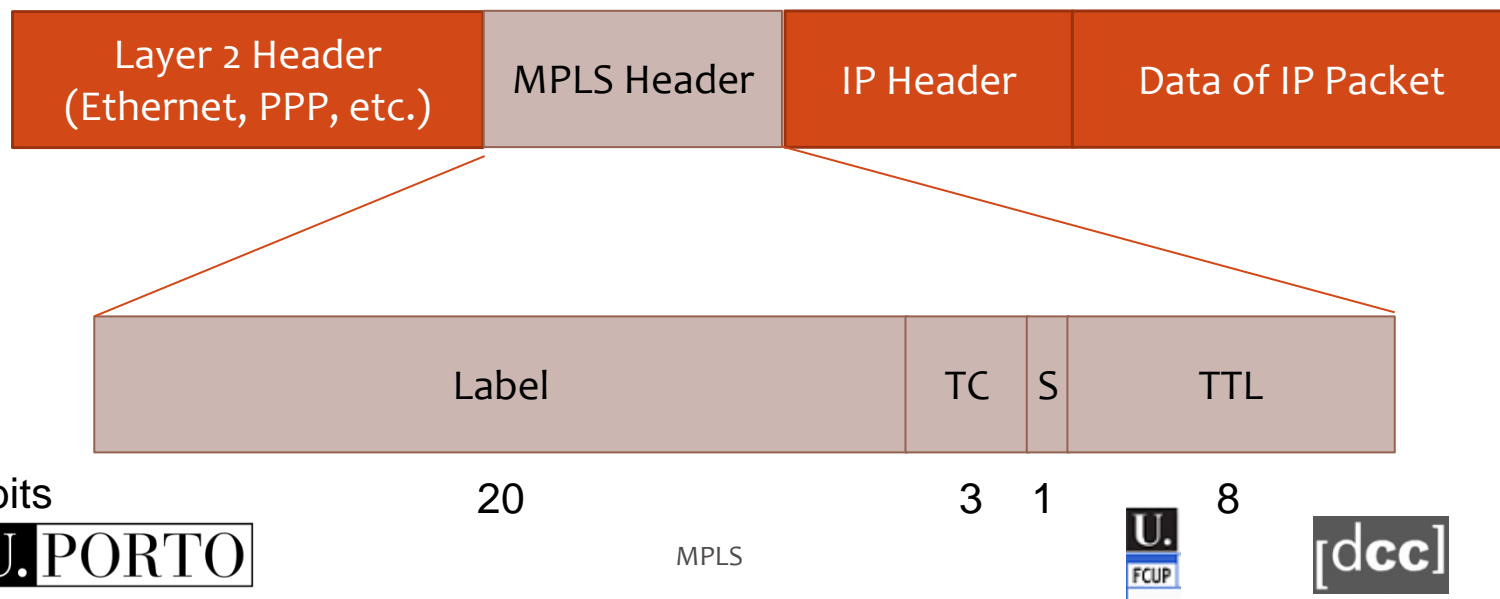
- VC establishment prior to data transmission, first packets delayed
- All packets follow the same path
- In-order delivery
- Failures must be explicitly handled
- Exact matching of VC identifier
- Packets contain VC identifier
- Routers maintain per-VC info
- Easy to combine with resource reservation
- Traffic engineering easy

Datagram Networks

- No VC establishment, data may be sent immediately
- Packets forwarded independently
- Packets may be reordered in transit
- Robust to link or node failures
- Longest prefix matching of addrs
- Packets contain src & dst addrs
- Routers maintain only aggregate destination info
- Resource reservation hard, requires additional protocols
- Traffic engineering harder

Multiprotocol label switching (MPLS)

- [RFC 3031](#) – Multiprotocol Label Switching Architecture
- Initial goal: speed up IP forwarding by forwarding based on a fixed length label (instead of IP address)
 - Borrowing ideas from Virtual Circuit (VC) approach
 - IP datagram still keeps IP address!



Multi-Protocol Label Switching



- Objective 1: flow detection and routing based on labels
 - Simpler and faster decision process
- Advantages:
 - Greater scalability
 - Better performance
 - Main reason at the beginning, not significant nowadays...
 - Separation of Routing and Forwarding
 - Routing
 - How to send packets from source to destination – global action
 - Forwarding
 - Transfer a packet from an entry port to an exit port – local action

Multi-Protocol Label Switching



13

- Objective 2: enable establishing VPNs across telecom operator's network
- Advantages:
 - Interconnection simplicity for clients that want to use different sites as if they were a single network

Multi-Protocol Label Switching



14

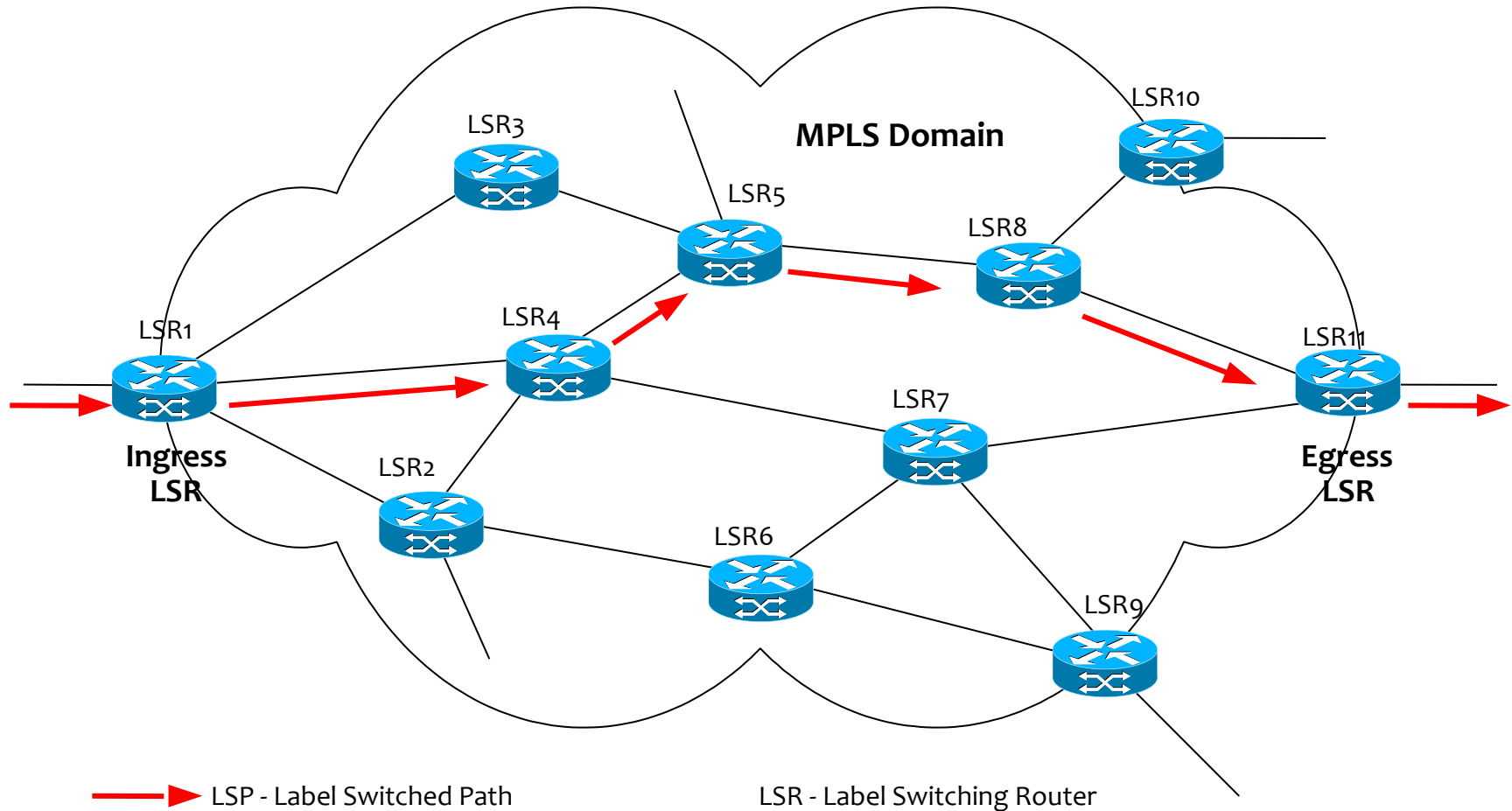
- Objective 3: enable traffic engineering
- Advantages:
 - Allows going beyond the routing protocols when deciding the path for a packet

MPLS

- MPLS supports
 - Integration with routing protocols (BGP, OSPF, etc.)
 - Unicast, multicast
 - Source routing
 - Route pinning
 - QoS



MPLS – Global view



MPLS – Architecture I

- Switching based on labels
 - Labels change at every node (label-swapping)
- Labels transported in:
 - Frames of a technology that directly supports it (e.g. ATM, Frame Relay)
 - Shim header
- **MPLS node:**
 - Node that supports MPLS
 - Capable of forwarding based on labels
 - Supports one or more L3 routing protocols
 - Does not need to search at L3

MPLS – Architecture II

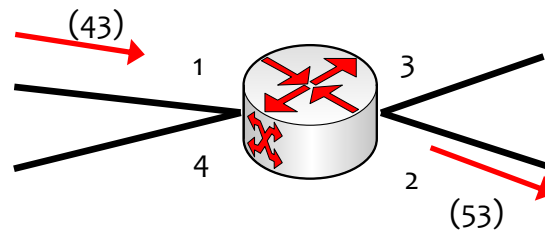
- **Label Switching Routers (LSR)**
 - MPLS Nodes capable of forwarding native L3 packets
- **Edge routers**
 - MPLS nodes at the border of MPLS domains
 - Ingress
 - Decides Forwarding Equivalence Class (FEC)
 - Transmits packet with label corresponding to FEC
 - Egress
 - Removes label*

MPLS – Architecture III

- MPLS Routers
 - Search the label on the Label Information Base (LIB)
 - New label – Next-Hop Label Forwarding Entry (NHLFE)
 - Transmit packet in out interface with new label
- Requires a mechanism for label distribution

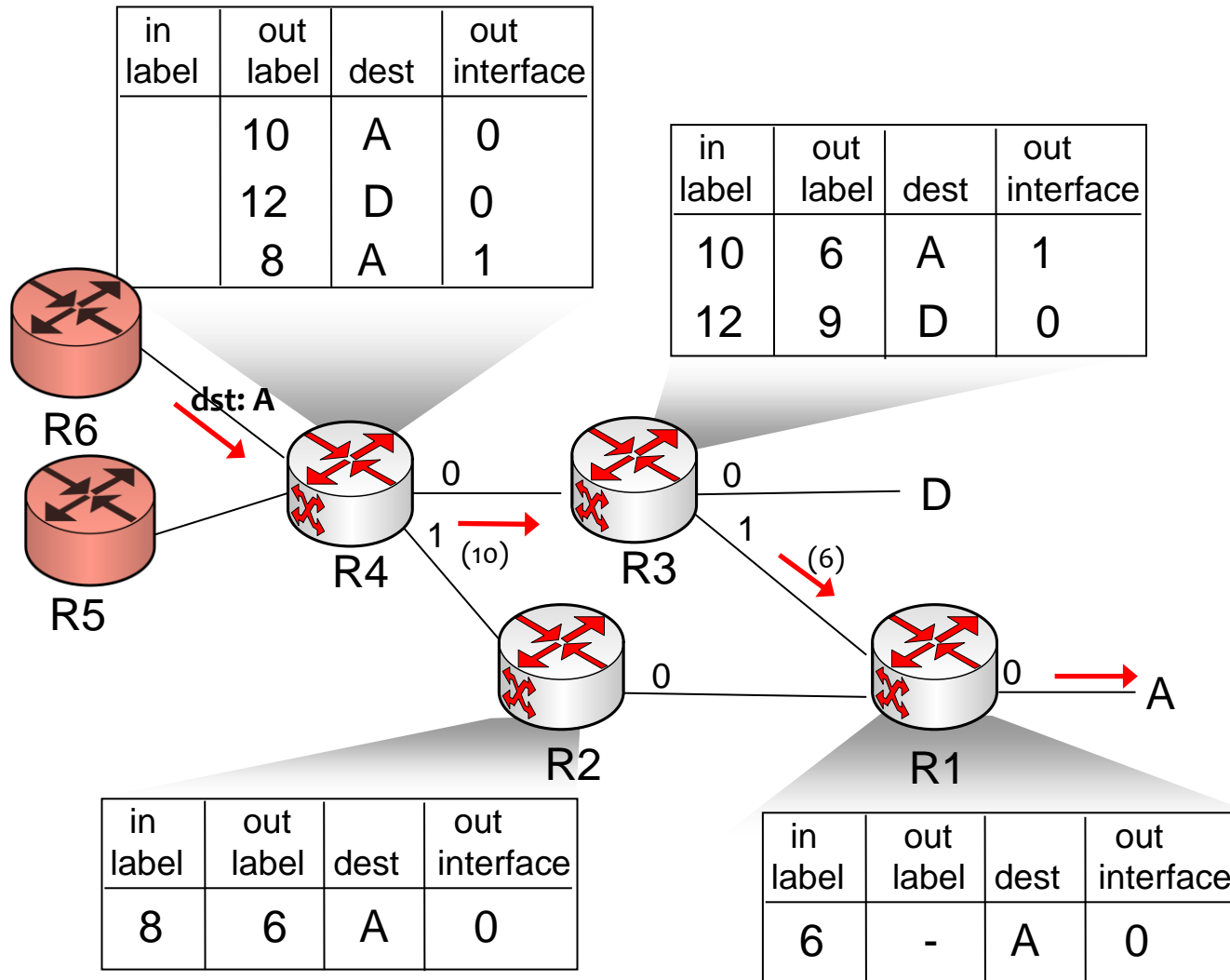
Itf In	Label In	Itf Out	Label Out
1	43	2	53
4	4	1	16

LIB



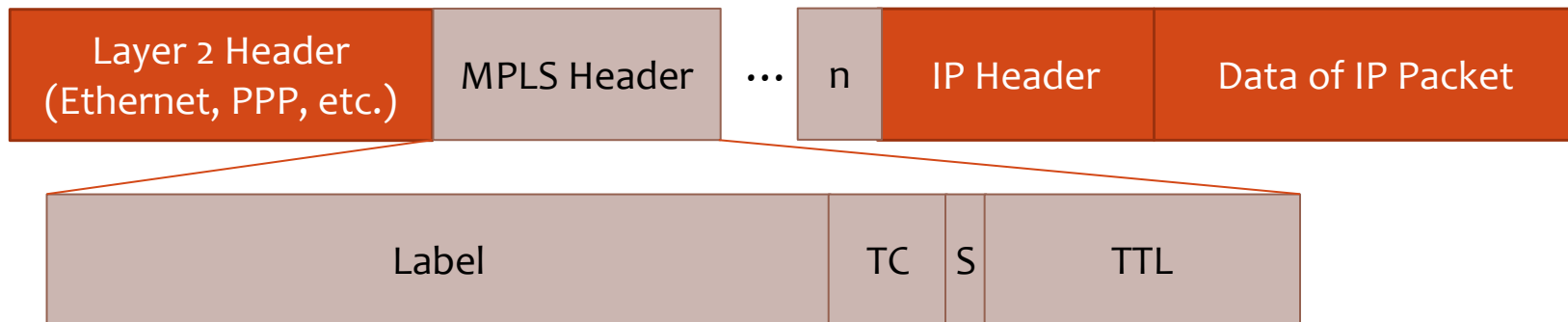
MPLS

MPLS forwarding tables



MPLS – Labels

- Shim header
 - Generic
 - “Layer 2.5”
 - Stackable
- Label
 - Small
 - Fixed size
 - Local meaning



Label: 20 bits

TC: 3 bits (traffic class information)

S: 1 bit (1 means bottom of stack)

TTL: 8 bits

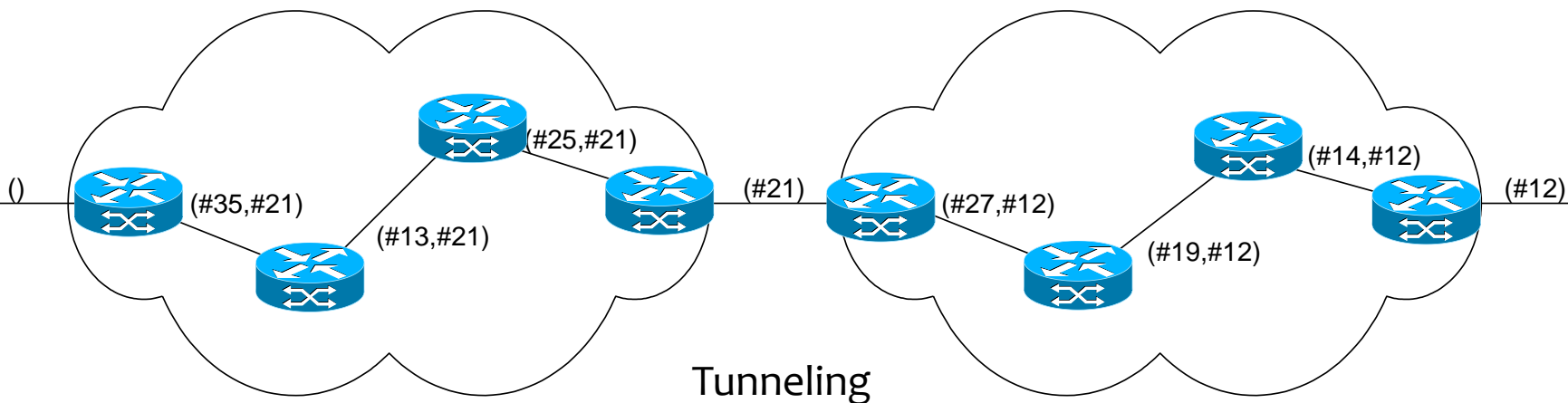
No need for shim header when the layer 2 technology supports labels (e.g., with ATM maps to VPI/VCI)

MPLS – Need for a TTL field

- Nodes after the MPLS domain must see the same TTL as if MPLS were not used
- TTL in shim header is set from the IP header
- TTL in the shim header is decremented in each MPLS node the packet goes through
- When removing the label, TTL in the IP header should be set to the value in the shim header
- More in [RFC3031#3.23](#) and [RFC3032#2.4](#)

Label stacking

- Non-hierarchical
 - Different labels added at the ingress LSR
 - Each router removes a label from the stack
 - More overhead, but even faster forwarding performance
- Hierarchical (tunneling)
 - Intra-domain and inter-domain

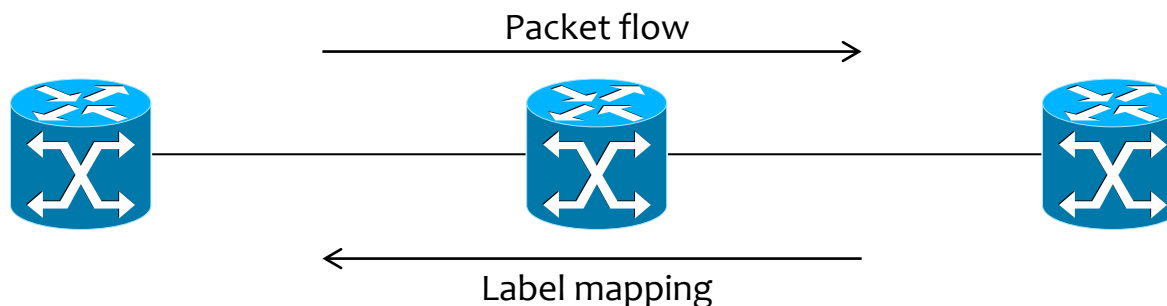


Forwarding Equivalence Class (FEC)

- Subset of packets handled similarly by the router, i.e.
 - Forwarded to the same Next Hop
 - Through the same interface
 - With the same treatment (e.g., queuing)
- The FEC
 - Is determined only at the ingress LSR
 - Determines the output label at that router
- Criteria for setting the FEC
 - IP prefix, aggregating
 - Egress edge router of domain
 - By flow, end-to-end
 - QoS / Traffic Engineering
 - Other criteria

Label distribution

- Routing information used to distribute labels
 - Piggyback on routing protocols
 - Multiprotocol-BGP
- MPLS nodes
 - Receive mapping from nodes “down” the path
 - Allocate and distribute labels for nodes “up” the path

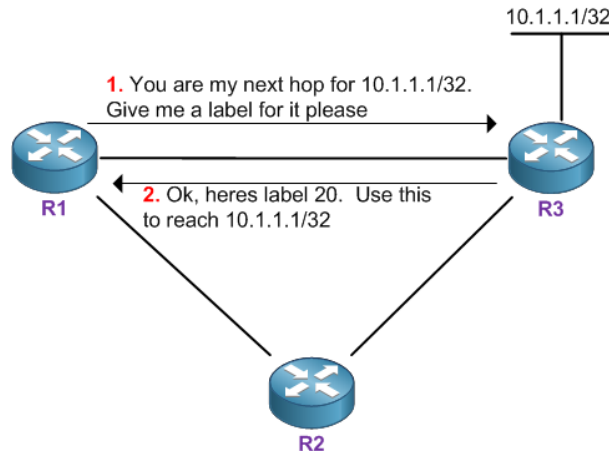


Label Distribution Protocol (LDP)

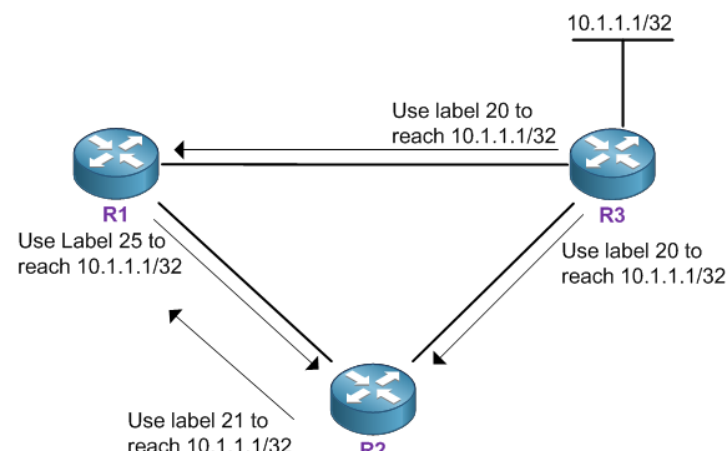
RFC5036

- Assign labels to routing table entries
- Define LDP adjacencies
 - Hello, etc.
- Two modes for label distribution:

Downstream On-Demand



Downstream Unsolicited



Label Distribution Protocol (LDP)

- Two modes for label control
 - Ordered Control
 - LSR advertises FEC if it is the egress LSR for the FEC or has received an advertisement from the downstream peer
 - Non-egress LSRs must wait for their downstream peers before advertising the FEC
 - Independent Control
 - LSRs advertise independently
 - May lead to (temporary) blackholing of some traffic
- Applicable when FECs are associated with destination address

Label Distribution Protocol (LDP)

- Alternatives to LDP
 - CR-LDP ([RFC3212](#))
 - Constraint-based Routing LDP
 - RSVP-TE ([RFC3209](#))
 - Extensions to RSVP for LSP Tunnels

Explicit Routing in MPLS

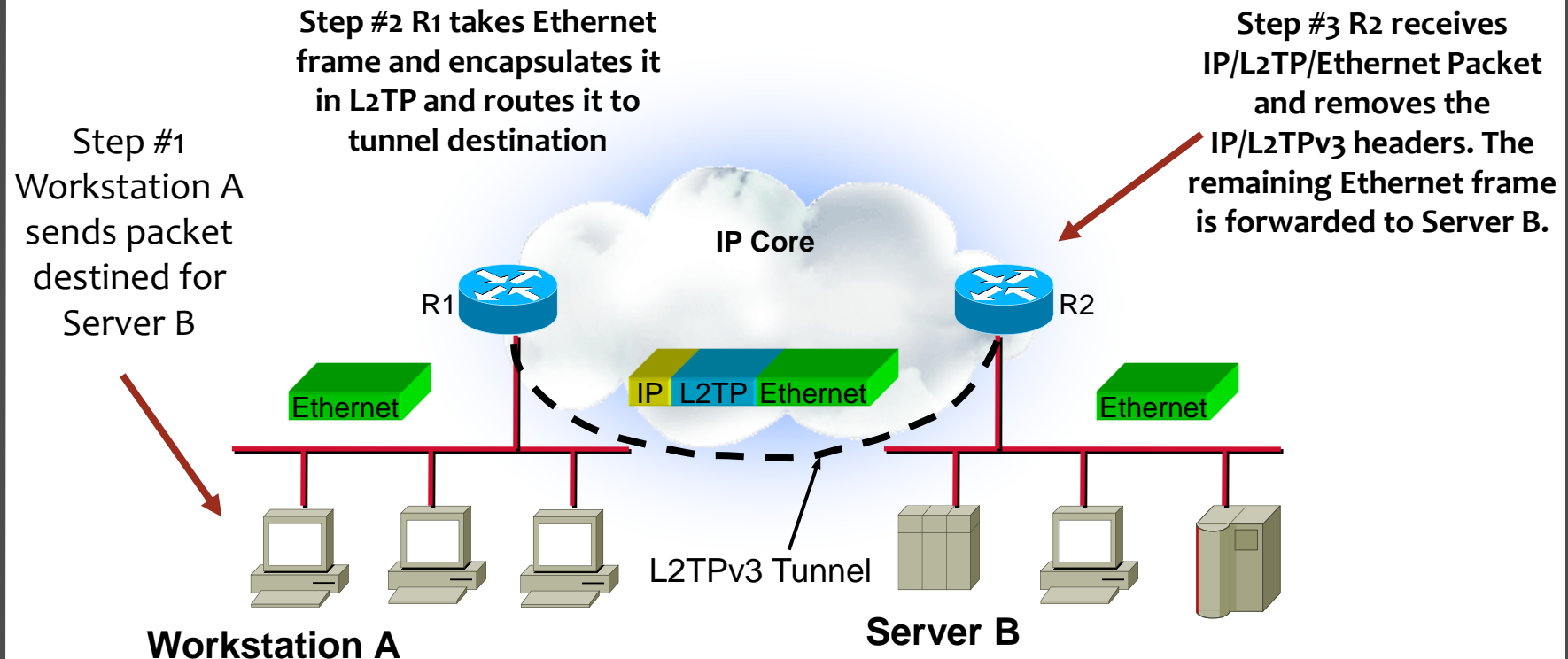
- Two options for route selection:
 - Hop by hop routing
 - Explicit routing
- Explicit Routing (Source Routing) is a very powerful technique
 - With pure datagram routing, overhead of carrying complete explicit route is prohibitive
 - MPLS allows explicit route to be carried only at the time the LSP is setup, not with each packet
 - MPLS makes explicit routing practical 😊

Explicit Routing (Cont'd)

- In an explicitly routed LSP
 - LSP next hop is not chosen by the local node
 - It is selected by a single node, usually the ingress
- The sequence of LSRs may be chosen by
 - Configuration
 - Administrator or centralized server
 - An algorithm
 - E.g., the ingress node may use topological information learned from a link state routing protocol

VPNs

Motivation – Layer2 Example



From [CSE8344]

Motivation – Overlay Model

- Service Provider provides PtP links to customer routers on other sites
 - Leased lines
 - Frame relay
 - ATM circuit
 - Emulated leased lines over MPLS
- Connectivity
 - Fully connected
 - Hub-and-spoke

From [CSE8344]

Motivation – Limitations of Overlay

- Customers need to manage the backbones
- Mapping between Layer 2 QoS and IP QoS
- Scaling problems
 - Cannot support large number of customers
 - $(n-1)$ peering requirement

From [CSE8344]

The Peer Model

- Service provider and customer exchange Layer 3 routing information
 - Provider relays data between customer sites using best path
- Goal: provide a large-scale VPN service
- Key technologies
 - Constrained distribution of routing info
 - Do not mix routes from different customers
 - Multiple forwarding tables (one per VPN)
 - VPN-IP addresses (combine VPN info and IP prefix)
 - MPLS switching
- [RFC 4364](#) BGP/MPLS IP VPNs

From [CSE8344]

Layer 2 vs Layer 3 VPNs

From [CSE8344]

Layer 2 VPNs




- Provider devices forward customer packets based on Layer 2 information
- Tunnels, circuits, LSPs, MAC address
- “pseudo-wire” concept ([RFC3985 - Architecture](#))
- VPLS - Virtual Private LAN Service Using (LDP) Signaling ([RFC 4762](#))

Layer 3 VPNs

- Provider devices forward customer packets based on Layer 3 information (e.g., IP)
- Service Provider involvement in routing
- MPLS/BGP VPNs ([RFC 4364](#)), GRE, virtual router approaches

The following discussion will concern Layer 3 VPNs

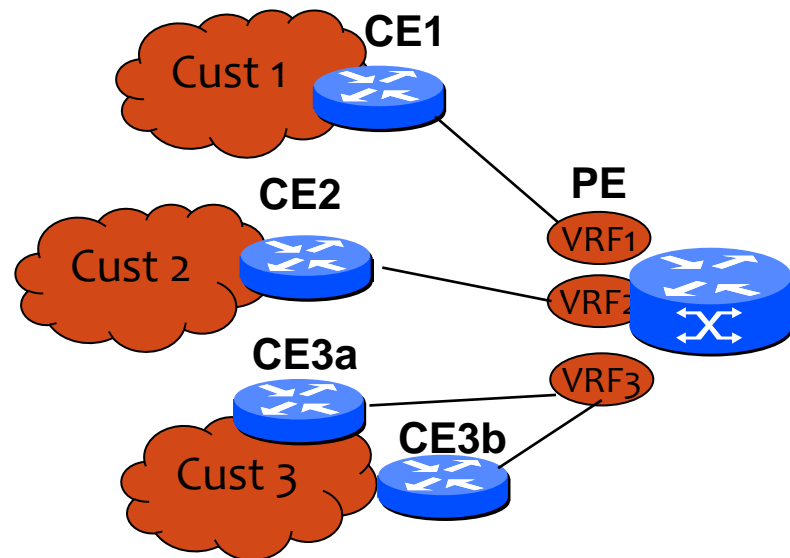
Terminology

- CE router 
 - Customer Edge router
- PE router 
 - Provider Edge router
 - Interfaces to CE routers
 - Is a Label Edge Router (LER)
- P router 
 - Provider (core) router, without knowledge of VPN or customer routes
 - Is a Label Switching Router (LSR)

From [CSE8344]

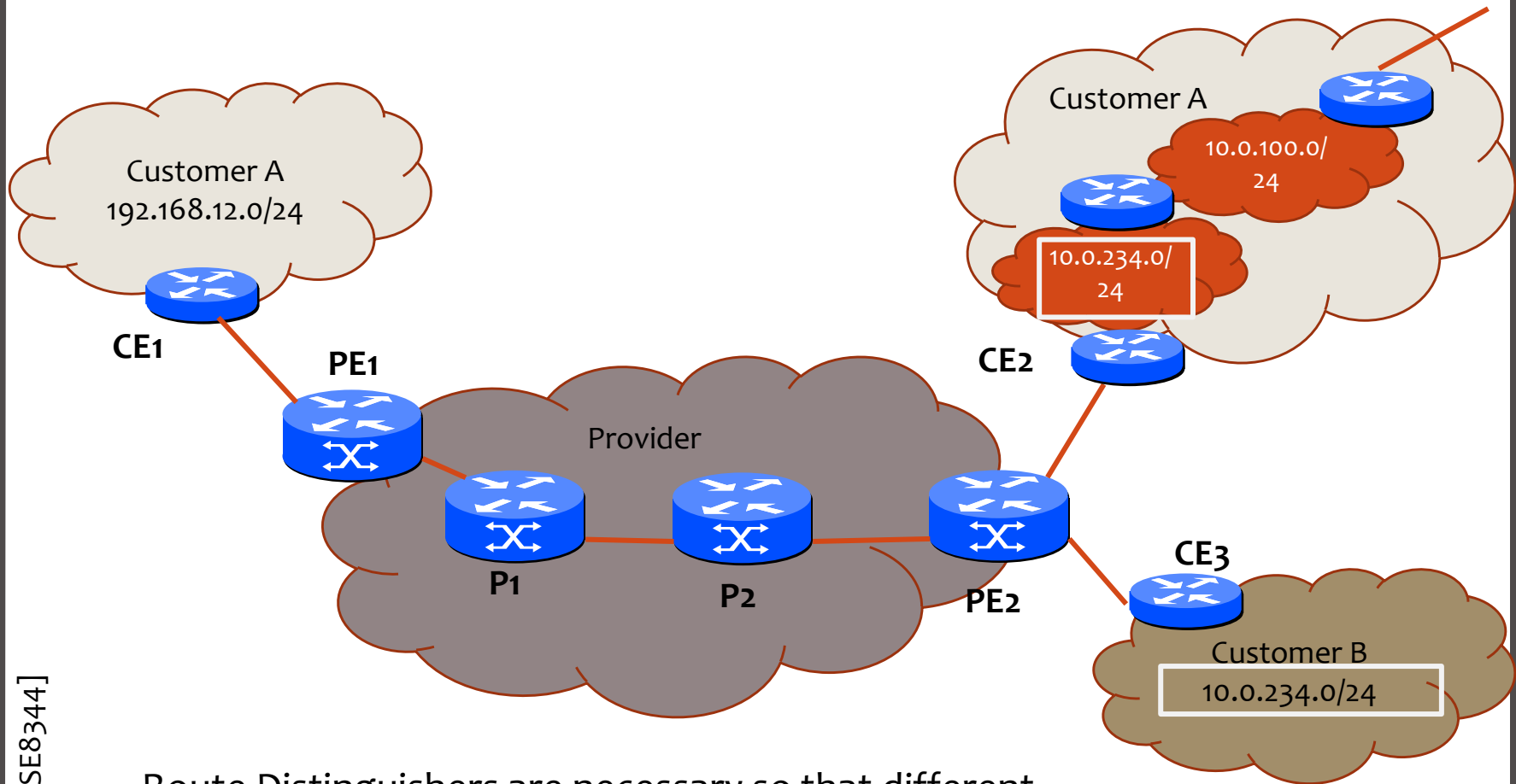
Terminology (cont'd)

- Route Distinguisher (aka route target)
 - Attribute of each route used to uniquely identify prefixes among VPNs (64 bits)
- VPN-IPv4 addresses ([RFC4364#4.1](#))
 - Address including the 64 bits Route Distinguisher and the 32 bits IP address
- VRF ([RFC4364#3](#))
 - VPN Routing and Forwarding Instance
 - Routing table and FIB table



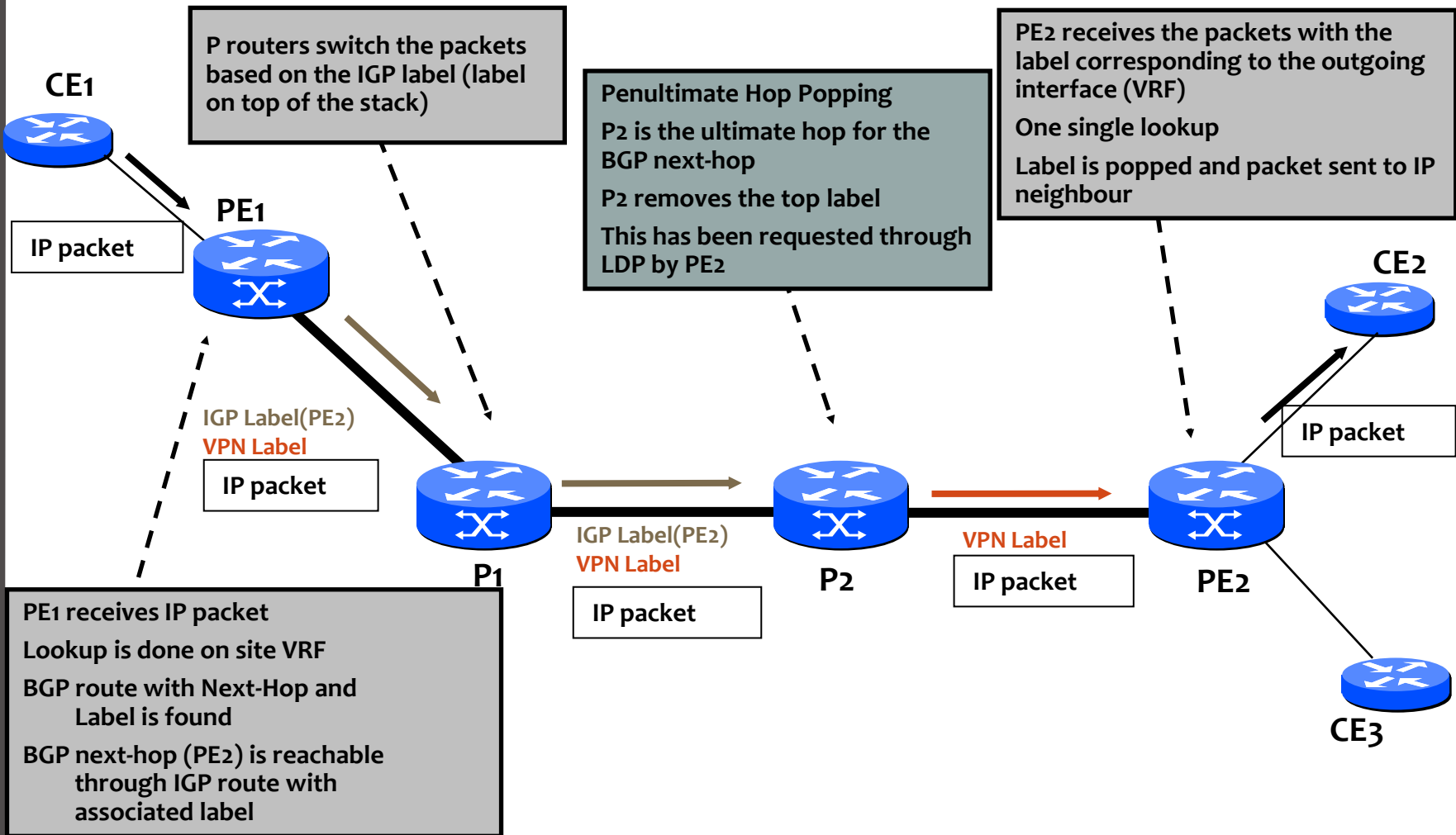
From [CSE8344]

Network example



Route Distinguishers are necessary so that different customers may use the same prefixes but they are not confused

Forwarding Example



From [CSE8344]

Connection Model

- The VPN backbone is composed by MPLS LSRs
 - PE routers (edge LSRs)
 - P routers (core LSRs)
- PE routers are faced to CE routers and distribute VPN information through MBGP to other PE routers
- P routers do not run MBGP and do not have any knowledge of VPNs
 - Complexity kept at the edges

From [CSE8344]

Model (cont'd)

- P and PE routers share a common IGP
 - Routing for destinations in the provider network
- PE and CE routers exchange routing information through some means
 - eBGP, OSPF, RIP, static routing
 - Protocol may be different in different sites of the same customer
- CE routers run standard routing software

From [CSE8344]

Routing

- Routes PE receives from CE are installed in the appropriate VRF
 - Assigned according to the incoming interface
 - VRF necessary to segregate customers
- By using separate VRFs, addresses need NOT be unique among VPNs
 - Useful with private addressing
- Routes PE receives through the backbone IGP are installed in the global routing table
 - Routes in the provider network

From [CSE8344]

Forwarding

- PE routers use MBGP to exchange reachability information and learn the BGP Next-Hop
- PE and P routers use IGP to establish the IGP Next-Hop towards the BGP Next-Hop
 - The BGP Next-Hop is the egress PE router
- Labels corresponding to BGP Next-Hops are distributed through LDP (hop-by-hop)
- Label Stack is used for packet forwarding
 - Top (outer) label indicates IGP Next-Hop
 - Bottom (inner) label indicates outgoing interface or VRF

From [CSE8344]

Forwarding (cont'd)

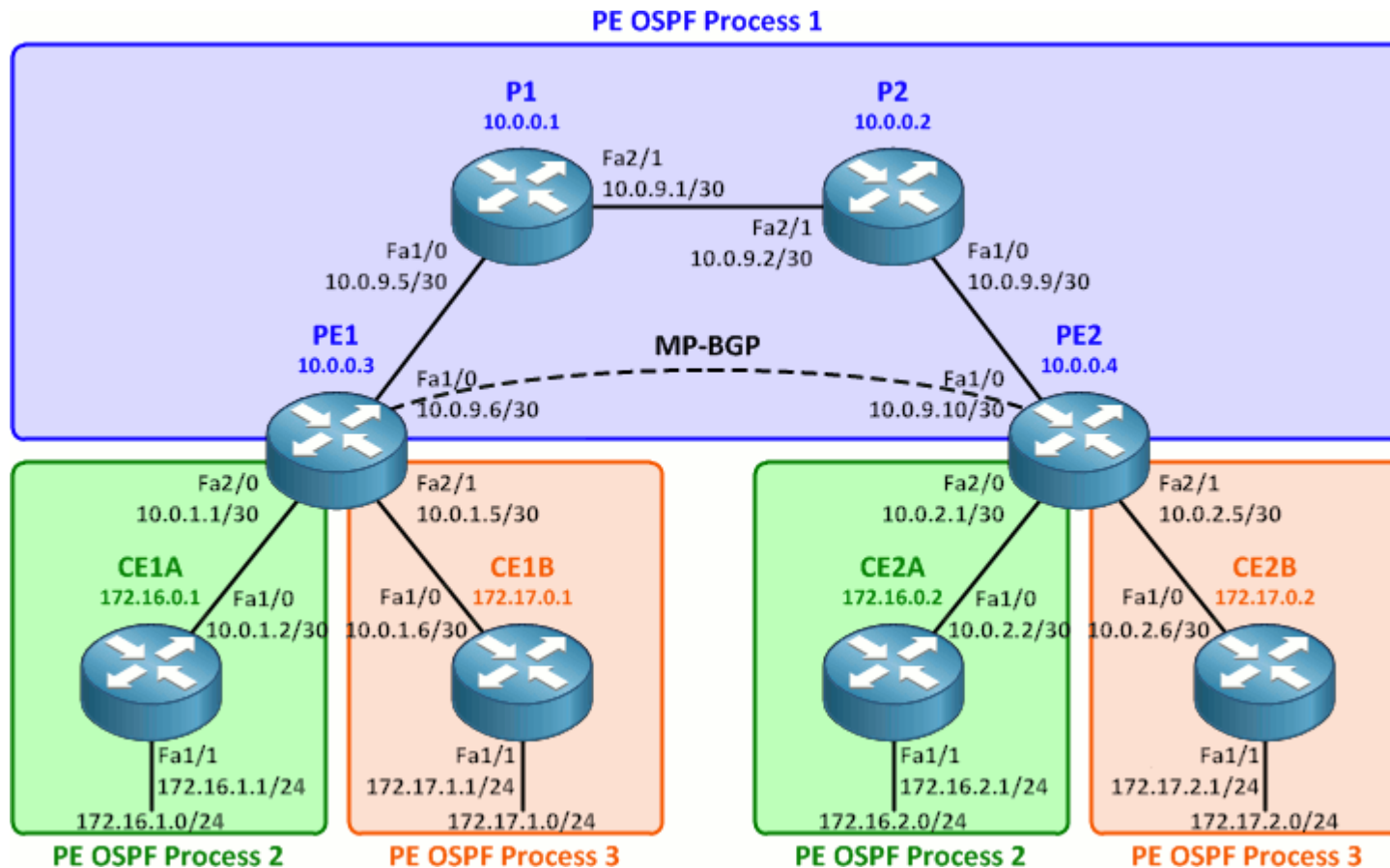
- The upstream LDP peer of the BGP next-hop (PE router) will pop the first level label
 - Penultimate Hop Popping
 - Avoid double processing at the egress PE Router
- The egress PE router will forward the packet based on the bottom label
 - The only one it receives
 - Determines the outgoing VPN and interface

From [CSE8344]

Scalability

- Existing BGP techniques can be used to scale the route distribution
 - E.g, use of route reflectors
- Each edge router needs only the information for the VPNs it supports
 - Directly connected VPNs
- Easy to add new sites
 - Configure the site on the PE connected to it, the network automatically does the rest ☺

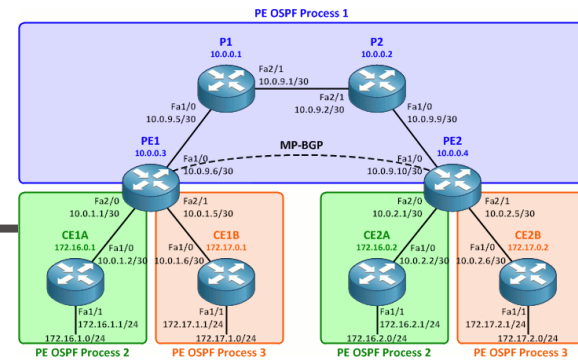
Demo



Adapted from <https://packetlife.net/blog/2011/may/16/creating-mpls-vpn/> replacing interfaces Fo/x with F2/x

Demo – Notes

- In this demo, OSPF is used in
 - The provider network
 - The customer networks
- In general, different protocols can be used
 - In the provider and customer networks
 - In the networks of different customers
 - In different sites of the same customer



Routes on PE1

- Global routing table

10.0.0.0/8 is variably subnetted, 7 subnets, 2 masks

```

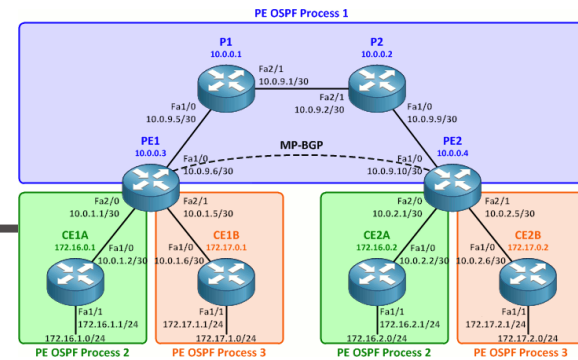
O      10.0.9.0/30 [110/2] via 10.0.9.5, 00:23:17, FastEthernet1/0
C      10.0.9.4/30 is directly connected, FastEthernet1/0
O      10.0.0.2/32 [110/3] via 10.0.9.5, 00:23:17, FastEthernet1/0
C      10.0.0.3/32 is directly connected, Loopback0
O      10.0.9.8/30 [110/3] via 10.0.9.5, 00:23:17, FastEthernet1/0
O      10.0.0.1/32 [110/2] via 10.0.9.5, 00:23:17, FastEthernet1/0
O      10.0.0.4/32 [110/4] via 10.0.9.5, 00:23:17, FastEthernet1/0
  
```

- Routing table for vrf Customer_A

172.16.0.0/16 is variably subnetted, 4 subnets, 2 masks

```

O      172.16.1.0/24 [110/2] via 10.0.1.2, 00:23:35, FastEthernet2/0
O      172.16.0.1/32 [110/2] via 10.0.1.2, 00:23:35, FastEthernet2/0
B      172.16.2.0/24 [200/2] via 10.0.0.4, 00:22:48
B      172.16.0.2/32 [200/2] via 10.0.0.4, 00:22:48
10.0.0.0/30 is subnetted, 2 subnets
B      10.0.2.0 [200/0] via 10.0.0.4, 00:22:48
C      10.0.1.0 is directly connected, FastEthernet2/0
  
```



Routes on CE1A

- Global routing table

172.16.0.0/16 is variably subnetted, 4 subnets, 2 masks

C 172.16.1.0/24 is directly connected, FastEthernet1/1

C 172.16.0.1/32 is directly connected, Loopback0

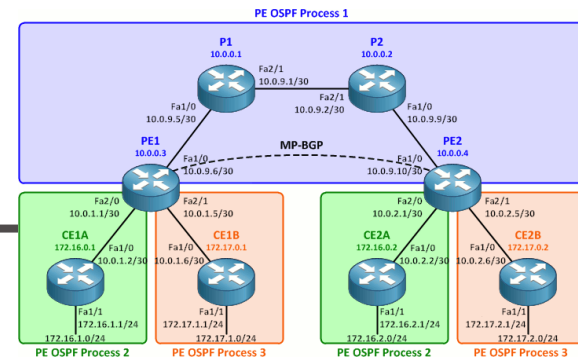
O IA 172.16.2.0/24 [110/3] via 10.0.1.1, 00:23:39, FastEthernet1/0

O IA 172.16.0.2/32 [110/3] via 10.0.1.1, 00:23:39, FastEthernet1/0

10.0.0.0/30 is subnetted, 2 subnets

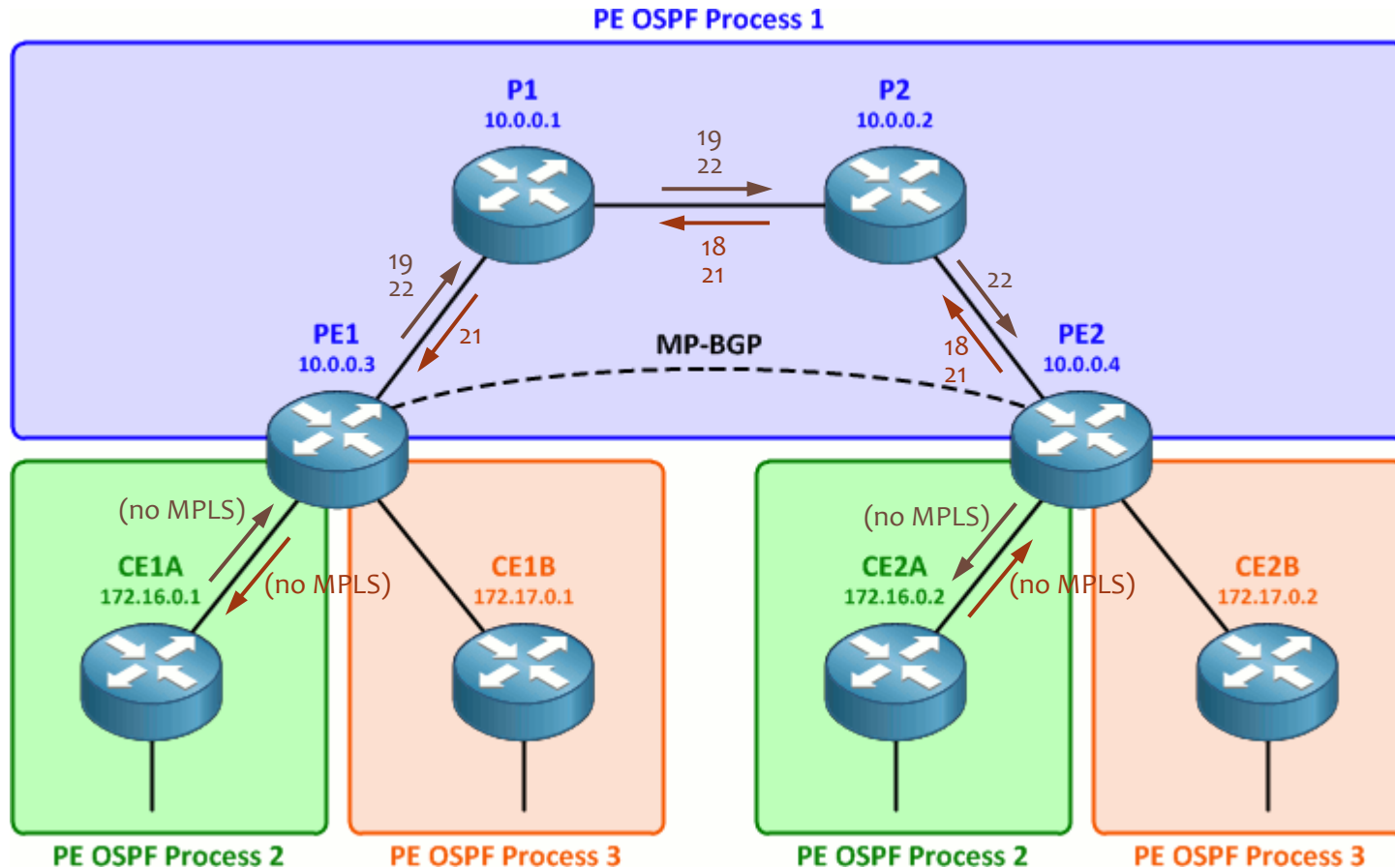
O IA 10.0.2.0 [110/2] via 10.0.1.1, 00:23:39, FastEthernet1/0

C 10.0.1.0 is directly connected, FastEthernet1/0



MPLS Labels

Ping from CE1A (10.0.1.2) to 172.16.0.2 (CE2A)

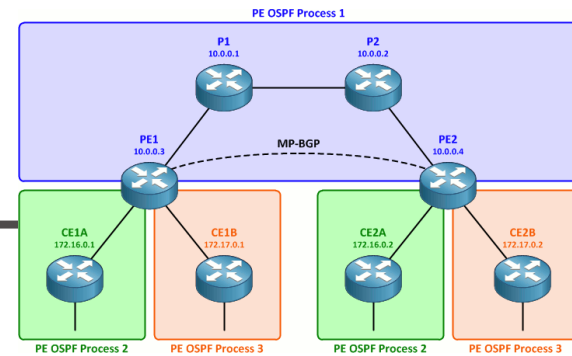


NOTE: P1 and P2 assigned the same label to 10.0.0.3/32 (19) and to 10.0.0.4/32 (18) due to the symmetry of the topology / by coincidence; usually, they will all be different.

BGP VPNV4 Labels

- PE1

Network	Next Hop	In label/Out label
Route Distinguisher: 65000:1 (Customer_A)		
10.0.1.0/30	0.0.0.0	21/aggregate(Customer_A)
10.0.2.0/30	10.0.0.4	nolabel/21
172.16.0.1/32	10.0.1.2	22/nolabel
172.16.0.2/32	10.0.0.4	nolabel/22
172.16.1.0/24	10.0.1.2	23/nolabel
172.16.2.0/24	10.0.0.4	nolabel/23
Route Distinguisher: 65000:2 (Customer_B)		
10.0.1.4/30	0.0.0.0	24/aggregate(Customer_B)
10.0.2.4/30	10.0.0.4	nolabel/24
172.17.0.1/32	10.0.1.6	25/nolabel
172.17.0.2/32	10.0.0.4	nolabel/25
172.17.1.0/24	10.0.1.6	26/nolabel
172.17.2.0/24	10.0.0.4	nolabel/26



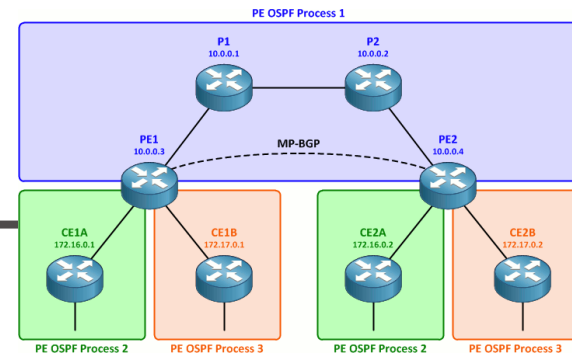
MPLS Fwd Table

• PE1

Local tag	Outgoing tag or VC	Prefix or Tunnel Id	Bytes tag switched	Outgoing interface	Next Hop
16	Pop tag	10.0.9.0/30	0	Fa1/0	10.0.9.5
17	16	10.0.9.8/30	0	Fa1/0	10.0.9.5
18	Pop tag	10.0.0.1/32	0	Fa1/0	10.0.9.5
19	17	10.0.0.2/32	0	Fa1/0	10.0.9.5
20	19	10.0.0.4/32	0	Fa1/0	10.0.9.5
21	Aggregate	10.0.1.0/30[V]	0		
22	Untagged	172.16.0.1/32[V]	1140	Fa2/0	10.0.1.2
23	Untagged	172.16.1.0/24[V]	0	Fa2/0	10.0.1.2
24	Aggregate	10.0.1.4/30[V]	0		
25	Untagged	172.17.0.1/32[V]	570	Fa2/1	10.0.1.6
26	Untagged	172.17.1.0/24[V]	0	Fa2/1	10.0.1.6

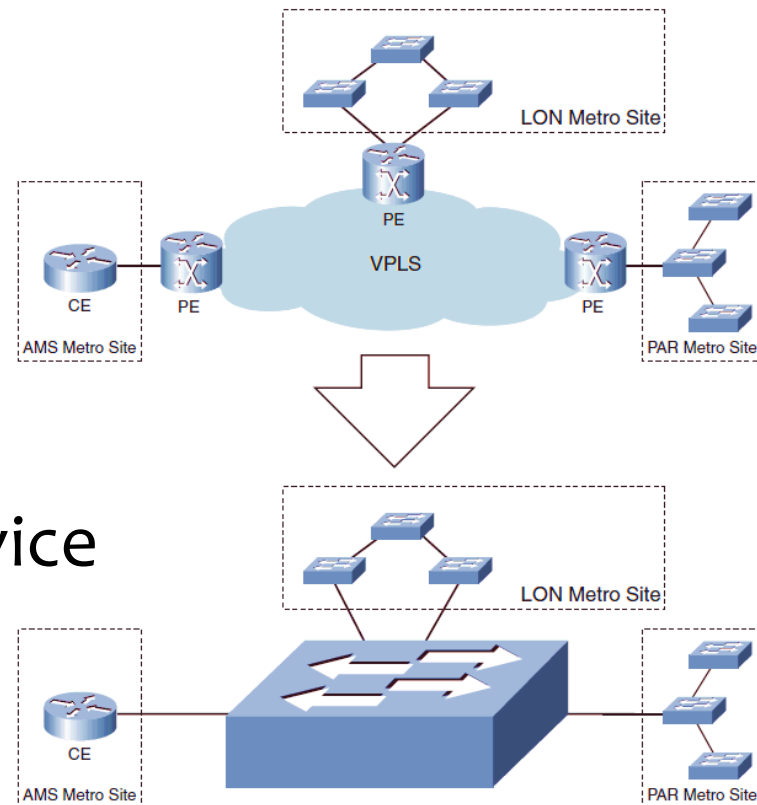
• P1

Local tag	Outgoing tag or VC	Prefix or Tunnel Id	Bytes tag switched	Outgoing interface	Next Hop
16	Pop tag	10.0.9.8/30	0	Fa2/1	10.0.9.2
17	Pop tag	10.0.0.2/32	0	Fa2/1	10.0.9.2
18	Pop tag	10.0.0.3/32	29201	Fa1/0	10.0.9.6
19	19	10.0.0.4/32	20005	Fa2/1	10.0.9.2



Other uses of MPLS

- Traffic engineering
- IPv6 over MPLS
- QoS
- Pseudowire – [IETF WG](#)
- Virtual Private LAN Service (VPLS) [IETF WG](#)



From [MPLSFund]

The end

Acronyms I – MPLS

- ATM – Asynchronous Transfer Mode
- BGP – Border Gateway Protocol
- CoS – Class of Service
- CE – Customer Edge (router)
- CR-LDP – Constraint-based Routing LDP
- FEC – Forward Equivalence Class
- FIB – Forwarding Information Base
- L2TP – Layer 2 Tunnelling Protocol
- LDP – Label Distribution Protocol
- LER – Label Edge Router

Acronyms II – MPLS

- LIB – Label Information Base
- LSP – Label-Switched Path
- LSR – Label Switching Router
- MPLS – Multi-Protocol Label Switching
- OSPF – Open Shortest Path First
- PE – Provider Edge (router)
- QoS – Quality of Service
- RSVP-TE - Resource reSerVation Protocol Traffic Engineering
- VRF – VPN Routing and Forwarding

References

- [CSE8344] MPLS Architecture CSE 8344 presentation, SMU
- [MPLSFund] MPLS Fundamentals, Luc de Ghein, CiscoPress 2006
- Tutorial: PacketLife [MPLS-VPN](#) using Cisco Routers and OSPF